

https://dspace.org/

The dependency index.

Rees, Gareth

2014-07-25
http://hdl.handle.net/10539/15009

Downloaded from a DSpace repository. https://dspace.org/

WITS BUSINESS SCHOOL

The dependency index

Research report MBA ft.

Gareth Rees 0204852N

11/2/2012

Supervisor: Gregory Lee

ABSTRACT

The economic exclusion of the majority of households in South Africa is highly systemic, and remains a fracture in the socio political economy of the country (Lichbach 1989, Thorbecke and Charumilind 2002). In order to promote human capital development and economic inclusion the state provides cash transfer social wages, derived from taxes, to citizens most in need of support (Oosthuizen 2007). This study defines a new index (the 'dependency index') that describes the economic relationship between households and the state, depending on the income decile (category) of the household. 21,144 households from the 2005/2006 Income and Expenditure survey, conducted by Statistics South Africa, were analysed in context of this new index. Results find that buying power of households per income decile is augmented significantly by the constitution of social grants as a function of tax adjusted earned income. Thus the index defines, per household, the consumption value of each rand earned through work and the net fiscal relationship with the state, on a single Cartesian scale. The study posed three hypotheses: a) the dependency index is significantly different per income decile, b) the index forms an inverse curvilinear trend across deciles, and c) the dependency index has a negative linear trend within income decile 10. ANOVA analysis and curve fitting provides support for all three hypotheses. The dependency index may be useful for policy formulation and evaluation, international benchmarking as well as household planning.

DECLARATION

I Gareth Rees declare that this research report is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Gareth David Rees

Signed at .....

On the ..... day of 2012

Dedicated to my lovely wife Petra.

Without you... none of this would be possible

Acknowledgements:

Gregory Lee: who worked tirelessly, dedicating many hours in assisting me reach the conclusion of my MBA and this research report

Contents

ABSTRACT 1

DECLARATION 2

1 INTRODUCTION 10

1.1 Purpose of the study 10

1.2 A story of three brothers 10

1.3 Context of the study 12

1.4 Problem statement 14

1.4.1 Main problem 14

1.5 Significance of the study 14

1.6 Delimitations of the study 15

1.7 Definition of terms 16

1.8 The following assumptions are implicit in this study 17

1.8.1 CSi is consumed in the same year 17

1.8.2 E less CSi is earned income for consumption 17

1.8.3 Free services subsidised by the state 18

1.8.4 Data used for this study is suitable 18

2 LITERATURE REVIEW 19

2.1 The household utility of income from wages 19

2.2 The social wage 20

2.2.1 Social grants in the form of cash 20

2.2.2 Social grants in the form of services 22

2.2.3 Summation of the social wage: 24

2.3 Household tax 25

2.3.1 Summation of household tax: 26

2.4 Concluding hypotheses from the literature review 27

2.4.1 Hypothesis 1 27

2.4.2 Hypothesis 2 27

2.4.3 Hypothesis 3 28
3 RESEARCH METHODOLOGY 29
3.1 Research methodology 29
3.2 Research Design 30
3.3 Population and sample 31
3.3.1 Population 31
3.3.2 Sample and sampling method 31
3.4 Measures 32
3.4.1 Average total expenditure per decile (E) 33
3.4.2 Income earned from wages per economically active individual in the household (Calculation of Ti): 33
3.4.3 Average consumption tax paid per decile (Tc) 34
3.4.4 Average cash transfer social grants received per household per decile (CSi) 34
3.4.5 Average subsidised social services in cash equivalents (SSi) 34
3.4.6 Mathematical model 35
3.5 Mathematical and statistical procedures 35
3.5.1 Step 1: Creating the Dependency Index 36
3.5.2 Step 2: Comparing income deciles using ANOVA procedures 37
3.5.3 Steps 3 and 4: Establishing the shape of the Index using curve-fitting 37
3.6 Interpretation 38
3.7 Limitations of the study 38
3.8 Validity and reliability 39
3.8.1 External validity 39
3.8.2 Internal validity 39
3.9 Reliability 40
4 CHAPTER 4: presentation of results 40
4.1 Introduction 40

4.2 Creating the dependency index (Main problem 2.4.1) 40
4.2.1 Consumption (E) 42
4.2.2 Income Tax (Ti) 43
4.2.3 Income from cash social transfers (CSi) 44
4.2.4 Subsidized services (SSi) 44
4.2.5 Validity of earned income assumption 45
4.2.6 Summary problem statement 46
4.3 Results showing differences in the index per income decile (Hypothesis 1) 46
4.3.1 Index manipulations 46
4.3.1 Non-parametric Kruskal-Wallis analysis of the raw index 47
4.3.2 GLM ANOVA of the winsorised index 49
4.3.3 GLM ANOVA of the Index less outliers: 50
4.3.4 Visual comparison of the factors 52
4.3.5 Summary of hypothesis 1 53
4.4 Results pertaining to Hypothesis 2 53
4.4.1 Curve fitting of the winsorized Index 54
4.4.2 Curve fitting of Index less outliers 55
4.4.3 Summary of per-decile curve fitting 56
4.5 Results pertaining to Hypothesis 3 57
4.5.1 Summary of hypothesis 3 59
4.6 Summary of the results 60
5 DISCUSIION OF THE RESULTS 60
5.1 Introduction 60
5.2 The dependency index 60
5.3 Discussion pertaining to Hypothesis 1 62
5.3.1 Non-parametric Kruskal-Wallis analysis of raw index 62
5.3.2 GLM ANOVA of the winsorized index 62

5.3.3 GLM ANOVA of index with removed outliers 63
5.3.4 Discussion on hypothesis 1 63
5.3.5 Why the indices differ in magnitude 64
5.3.6 Conclusion of Hypothesis 1 65
5.4 Discussion pertaining to Hypothesis 2 65
5.4.1 Conclusion of hypothesis 2 67
5.5 Discussion pertaining to Hypothesis 3 68
5.5.1 Conclusion for hypothesis 3 69
5.6 Conclusion of discussion 69
6 CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS 70
6.1 Introduction 70
6.2 International comparison and benchmarking 70
6.3 Provision of state services 72
6.4 The tax burden 72
6.5 Contributors or consumers of state capital 72
6.6 Bottle necks in the economy 73
6.7 Sensitivity testing of proposed tax and social policies 74
6.8 Recommendations 74
6.9 Suggestions for further research 75
7 References 78
ANNEXURE A 83

List of figures

Figure 1: Social welfare in South Africa (Woolard and Harttgen 2010) 13
Figure 2: Components of cash income by income decile in RSA (McEwen, Kannemeyer and Woolard 2009) 19
Figure 3: Disaggregated RSA government social protection expenditure 2010/11 (National treasury 2010; Ngandu and Altman 2010). 22

Figure 4: Government spending on social services 2010/11 (National treasury 2010; Ngandu and Altman 2010) 24
Figure 5: Composition of main sources of tax revenue 1995-2008 (National Treasury and SARS 2011) 25
Figure 6: Process of data integration, manipulation and transformation 42
Figure 7: Means of the three index iterations per income decile 52
Figure 8: Standardised dependency index iterations for trend analysis 53
Figure 9: PolyRatio(3,3) graph fitted to winsorized index per decile 55
Figure 10: PolyRatio(3,3) graph fitted to per decile mean index less outliers 56
Figure 11: PolyRatio(3,3) graph fitted to the mean index of the top ten percentile less outliers 59
Figure 12: Dependency index across three different index iterations, showing similar groups 63
Figure 13: expected curvilinear relationships of the mean factor per decile 66
Figure 14: Decision tree for obtaining household dependency index 67

List of tables

Table 1: Consumption (E) / Decile: 42
Table 2: Wages / Decile 43
Table 3: Social Grants (CSi) / Decile 44
Table 4: Total Subsidized Services (SSi) / Decile 44
Table 5: (Wages+ CSi) / Decile 45
Table 6: Raw Index / Decile : Kruskal-Wallis non-parametric ranked medians 47
Table 7:The Kruskal Wallis one-way ANOVA statistical fit 48
Table 8: Similar groups identified by the Kruskal Wallis ANOVA (Dunn's test) 48
Table 9: Winsorized Index/ Decile : Summary of Means 49
Table 10: Winsorized Index/ Decile: GLM ANOVA regression 49
Table 11: Similar groups identified in the winsorized index (Tukey Kramer pairwise comparison) 50
Table 12: Index Less Outliers / Decile : Summary of means 50
Table 13: Index less outliers/ Decile: GLM ANOVA regression 51
Table 14: Similar groups identified in the index less outliers (Tukey Kramer multiple comparison) 51
Table 15: Equation parameters for PolyRatio(3,3) winsorized index/ income decile 54
Table 16: Equation parameters for PolyRatio(3,3) Index less outliers / income decile 55
Table 17: Top 10 Percentile of index less outliers: GLM ANOVA regression 57

Table 18: Index Less Outliers / Top 10 Percentile: Summary of means 57
Table 19: Equation parameters for PolyRatio(3,3) Index less_outliers / top 10 percentile 58
Table 20:Income deciles 83
Table 21:Tax Brackets 83
Table 22:Free services per household 83
Table 23:VAT paid per income decile 84
Table 24: Representivity of sample data compared to the NiDS 2009 dataset 84

1 INTRODUCTION

1.1 Purpose of the study

The purpose of this study is to construct a model that predicts average household utility (total consumption plus free subsidised services) relative to each rand of income earned by income decile in South Africa. The index produced could constitute a useful metric for international macroeconomic comparisons, as well as microeconomic monitoring and recommendations. The model would be able to demonstrate a cross-section of the fiscal relationship between government and society per income decile. And in turn be translated into the relative dependency of specific population groups on the state, and indeed the relative dependency of the state on specific population groups.

1.2 A story of three brothers

All income is not created equal. Income and consumption tax affects the purchasing power of each rand of income earned, and these taxes are levied at differing rates depending on the economic status of the income earner or the products and services that are purchased by the earner. Similarly, due to the welfare state in South Africa the source of income per household differs greatly, further affecting the purchasing power of earned income. This can best be illustrated through a brief story about 3 brothers, Bob, Fred and Mike Skosana.

All three brothers grew up with their mother in a single parent family in rural Transkei in the Eastern Cape. Bob is the eldest son and as such had the responsibility of assisting his mother with the running of the household. He spent his youth farming the small field and looking after the two family cows to eke out a subsistence living for the family. As a result Bob never finished school and took over the household, when their mother died, where he resides with his family to this day. Fred was allowed to finish school due to his older brother running the household and taking care of their ailing mother. Fred was an entrepreneur and bought various consumer goods and liquor in the city and sold them with high margins in the rural areas. Inspired by the benevolence of his elder brother Fred used the funds from his bootlegging to put their youngest brother Mike through one of the better schools in the area that had higher school fees. Mike made the best of the opportunity and finished school with 6 distinctions and landed a scholarship at one of the best universities in the country to study engineering. The Skosana brothers are still very close, but their different experiences have resulted in them each occupying a different economic status, which can be demonstrated effectively through the use of the dependency index.

Bob is the least well off of the brothers due to his lack of opportunities and education. Although he earns a very small amount of income from selling excess produce, his primary source of income is from social grants. His brothers have assisted him in applying for child care grants for his three children as well a disability grant for his wife who is blind in one eye. Without their assistance it is likely that Bob would receive none of these grants due to his complete illiteracy. Thus despite Bob earning very little income through work his total annual household income is R18,000, placing him in income decile 3. Fred on the other hand became a taxi driver in Durban and currently has a fleet of two vehicles that he operates with a partner. Fred is growing in wealth but currently has a total income of R43,000 per annum, placing him in income decile 6. Mike on the other hand received a good offer from Group 5 construction after his studies, and currently has a total annual household income of R450,000 per annum, placing him in income decile 10. Without knowing the constitution of income of each brother, their income tax brackets, or nature of their consumption basket, it is difficult to determine their relative utility of earned income. The dependency index can generalise these figures per income decile however, and has this to say about the circumstances that each brother finds himself in.

Bob is in income decile 3. If he is representative of the average household in this decile he receives 41% of his income from social grants, pays no direct income tax, has an average consumption tax of 7.6%, and receives 6,000l of water and 50Kwh of electricity for free per annum. Thus for every one rand that Bob earns in income he can spend two rand and twenty five cents. His household is therefore a net burden to the state of one rand and twenty five cents for every rand that he spends of earned income from his farming. Fred on the other hand is in income decile 6. If he is representative of the average household in this decile he receives 28% of his income through social grants, pays no direct income tax, has an average consumption tax of 10%, and receives 6,000l of water and 50Kwh of electricity for free per annum. Thus for every one rand that Fred earns he can spend one rand and ninety two cents. His household thus is a net burden to the state of ninety two cents for every one rand of earned income that he spends from operating his taxis. Mike on the other hand is near the top end of income decile 10. If he is representative of the average household at this spectrum of the decile, he receives 1% of his income from social grants, pays 40% income tax, has an average consumption tax of 13% and receives no free services from the state. Thus for every one rand that Mike earns in income he can spend sixty seven cents, meaning that the state gains income of thirty three cents for every one rand that Mike spends.

The story of the Skosana brothers indicates the relative dependency on the state of each of their households, and goes some way in unpacking their economic story. The dependency index therefore generalises the nature of a household's economic interaction with the state depending on the income

decile of that household, based on historical national income and expenditure statistics. Importantly the index also shows that the gap between the brothers is even larger than the difference in their total household incomes, as both Bob and Fred have a significant portion of their total income arising through state support. These topics as well as additional uses for the index will be unpacked in this study.

1.3 Context of the study

South Africa has one of the widest income distributions and unemployment rates in the world. Depending on whom you ask, the national Gini co-efficient – a measure of social inequality - ranges between 5.9 and 7.2 (IES analysis 2008; Woolard, Harttgen, and Klasen 2010). There are several factors that have resulted in current income disparity, social structure and institutional bias, the causality and persistence of which is highly intertwined (Fedderke and Luiz 2008, Fedderke, de Kadt and Luiz 2001). Income disparity is a deeply rooted and contentious apartheid legacy issue, with the wealthy class still largely constituted by the ethnic minority (Oosthuizen 2007). Thus income disparity remains a fracture in the political economy of the country (Sen 1995, Lichbach 1989, Thorbecke and Charumilind 2002) as the incumbent government needs to satisfy its constituency whilst creating an environment that promotes the growth of the private sector (Ngandu, Altman, Cross, Jacobs, Hart and Matshe 2010) and its affluent benefactors. Two critical levers have been used to address these legacy issues, the social wage which reduces poverty and distributes income (Samson, Babson, Haarmann, Haarmann, Khathi, Mac Quene, and Van Niekerk 2002), and differential income tax which pays for the former. South Africa unlike many welfare states, pays for its social wage from tax and not aid or sovereign debt (Woolard and Klasen 2010).

South Africa had a notably well-developed social security structure even before the end of Apartheid (Lund 1993; Case & Deaton 1998) although it was only accessible to those of the white ethnic group. The Social Assistance Act of 1992 entitled all citizens to be eligible for social assistance, which increased the spending on this function considerably (Woolard and Harttgen 2010, Oosthuizen 2007). The nature of this assistance had two broad goals, to reduce poverty among vulnerable groups and to ensure investment in human capital through education and basic services (Woolard and Harttgen 2010). Thus there tends to be a cash component to the social wage and a subsidised or free service component. Figure 1 below indicates the various spending categories on social welfare, and its source of revenue in South Africa.

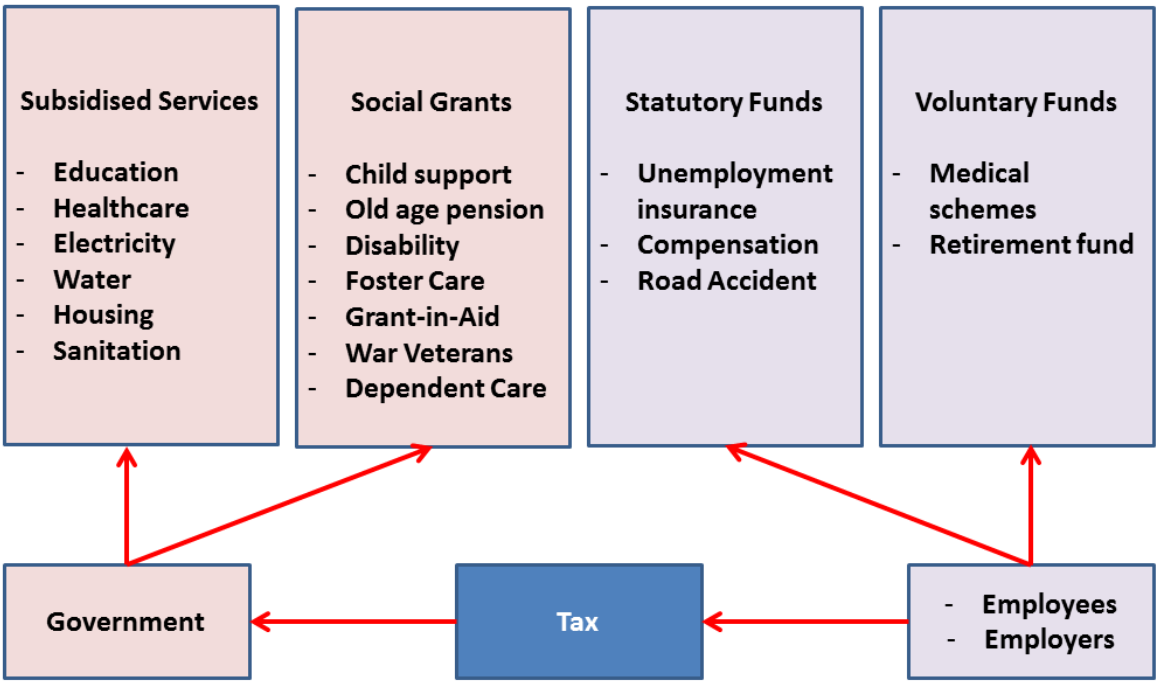


Figure 1: Social welfare in South Africa (Woolard and Harttgen 2010)

Thus, in its simplest form, increasing taxes on wealthy households and businesses are in part directed through government to poor households to reduce national income disparity and poverty. In the analysis of the 2006 income and expenditure report by Statistics South Africa, it is stated that the GINI co-efficient for South Africa would rise from 7.2 to 8.0 if it were not for the progressive social policies in place (IES 2008). Similarly the department of social development states that these policies are reducing poverty and promoting labour force inclusion which is increasing household earnings (Samson and Babson 2002). Literature (Woolard and Harttgen 2010) also points to the importance and success of social wage policies in reducing poverty and increasing economic inclusion.

However, not all literature is in agreement with the success of these initiatives in reducing poverty and increasing household employment. Whiteford and Van Seventer (2000) stated the structure of the economy would deepen income disparity almost inevitably irrespective of the social wage, due to the rising employment of skilled labour and the decreasing demand for unskilled labour. These workers are then forced into the informal sector which is usually for necessity and not opportunity. Employment creation has been a problem in post-apartheid South Africa, with 72% of households (68% of African households) containing no unemployed people in the mid 90's to 58% containing no unemployed people (52% of African households) by 2002 (Meth and Dias 2004). This suggests that poverty has almost certainly worsened as between 2 - 4 million people were added to the two bottom income deciles between 1999 and 2002 (Meth and Dias 2004). The findings of increasing poverty and

inequality since Apartheid were corroborated in a World Bank study in 2004 (Hoogeveen and Ozler 2004).

No matter which position is taken, South Africa has to devise mechanisms to address inequality in the long run. Inequality is a political destabilizer (Moaddel 1994, Haggard and Kaufman 1997) and it is likely that if progress is not made in narrowing this gap the stability of the economy in the long run will be compromised. In this study the appropriateness of policy around social wages and taxes will not be discussed, only the consequences of such policies and their implications on income distributions per income decile. This research should produce a model that provides a unique lens with which to view income disparity in South Africa.

1.4 Problem statement

1.4.1 *Main problem*

To construct and analyse a model (the dependency index) that predicts average household utility relative to income earned per income decile in South Africa.

1.5 Significance of the study

In South Africa, depending on your income, social grants can constitute a significant portion of that monthly income (Oosthuizen 2007, Whiteford and Van Seventer 2000 *Woolard and Harttgen 2010).* Similarly, but in inverse proportion, affluent households receive low levels of social assistance and have the utility of their income eroded by income tax (Oosthuizen 2007, Statistics SA 2008). Thus the utility of each household relative to pre-tax wages earned is different depending on the percentage of income tax paid and the number of equivalent rand's received through social grants. This is imperative to household and family planning as the utility gained through the acquisition of each additional rand earned through wages could be different per income decile. For example, if a member of the household gets a better job the household would assume the ability to consume more goods and thus increase their utility. The increase in household income however, could place the household in a different income bracket, and social grants could be removed and income tax could be incurred resulting in a lower than expected utility to income ratio. This is a simplistic example but it establishes the need to understand the utility of income as wealth increases.

Of further interest is the analysis of the dependency index as an international comparative, such as that of the GINI co-efficient (Morgan 1962). The GINI co-efficient shows the income disparity within a country but it does not demonstrate the constitution of such income. Also income is generally poorly

reported in national studies, and informal trade and subsistence living is excluded all together. Thus there are advantages in measuring the consumption function as a measure of utility. The sources of the income for the said consumption can then be partitioned as a ratio of dependency on the state (or vice versa) per income decile. The metrics are thus readily available in most countries and direct comparisons can be made without adjustments for purchasing power parity, as it is a ratio. The comparison of the data will demonstrate the structure of the economy, in each case the degree and depth of penetration of social services as well as the tax burden. Ultimately the state dependency can be compared directly per income decile in a single index.

The study will be insightful to those who are concerned with the fundamental structure of the national economy, and how its citizenry perceive income and consumption relationships at each step of the economic staircase (income deciles). The dependency index could:

- Be useful in international comparison and benchmarking
- Identify bottle necks in the progression of citizens up the economic staircase
- Identify weaknesses and successes of the current implementation of state services
- View the impact of tax on spending power
- View which groups are contributing to the state fiscus and to what extent
- View which groups are consumers of the state fiscus and to what extent

1.6 Delimitations of the study

- The study is not longitudinal, and shows a snap-shot of the relative value of income per decile based on statistics by Statistics SA in the 2005/2006 Income & Expenditure survey (IES). The reason for this is that differences between the methodology and accuracy of the last three IES reports renders the information difficult to interpret over time without complicated statistical alterations.
- The data is taken from 3,000 primary sampling units (PSU) to represent the spread of household income and expenditure, from each PSU 8 samples are taken to give a total of 21,144 dwelling units within the survey. Thus 21,144 households are measured as a representation of the broader population of South Africa, not every household in the country was factored in individually however. Rather weightings were allocated to the households to ensure representivity (IES 2008).
- The data was cleaned by DataFirst's "data quality project" to ensure the integrity of the data files. Despite this there are reported internal discrepancies within the data-set itself (Statistics SA, 2008), but these will not be considered in the analysis.

- The data is for South Africa only and will not consider the racial or demographic makeup of the income deciles. There could be analysis conducted on a by province, race or even gender basis, all of which could have very interesting results and implications. This is beyond the scope of this study however, and could obscure the focus of the study from its current purpose. Further research in these areas could be of interest.

1.7 Definition of terms

$$Dependency_Index = \frac{E*(1-Tc)+SSi}{(E-CSi)/(1-Ti)}$$

Dependency Index : The dependent variable in this study as defined by the equation shown above is calculated for each household in the 2005/2006 RSA IES, and then divided into means by income decile. Occasionally where appropriate the index has been referred to as the factor. The reasoning behind the dependency index will be discussed through the literature review, and the construction of the equation will be further elucidated in the research methodology.

E: total household consumption per year

CSi: total cash social wages (transfers) per household per year

SSi: total cash equivalent of free water and electricity per household per year

Ti: Average household income tax

Tc: Average consumption tax paid

Income Source: the source of the income to the household (wages, salaries, child grant, dividends, alimony etc.)

IES 2005/2006: This is the income and expenditure survey published by Statistics South Africa every 5 years (1995, 2000, 2005, 2010 pending). It contains survey results of several thousand households across the country to try and gain a representative sample of metrics per income decile.

Income Decile: breakdown of the 21,144 observation household sample in 10 roughly equally sized discrete categories by total household income. The parameters used to define income deciles are those published by Stats SA in the 2005/2006 IES. The bottom ten economic % are known as income decile 1, and the top ten economic % are known as income decile 10. This is the primary sorting function within this study.

Utility: the utility of each household is defined as the households total consumption less paid consumption tax (VAT), plus the economic equivalents of free services provided to the household by the state as water and electricity.

Wages: refer to income earned through salaries and wages for services rendered

Earned Income: Total income less social wages, which will be elaboratedon in the study.

Purchasing power parity: (PPPs) are the rates of currency conversion that equalise the purchasing power of different currencies across different countries.

Winsorizing: replacing observations exceedingthreshold values with a predetermined value to offset the effect of outliers

1.8 The following assumptions are implicit in this study

1.8.1 *CSi is consumed in the same year*

This assumption is intuitively correct, as grants are provided to those in need and these funds transferred as cash would be consumed and not saved in the year in which they are given. This is fundamental to the dependency index as these social cash transfers are deducted directly from consumption to gain consumption from earned income. There were 1,681 households of the 21,144 that had a negative dependency index, which indicates that cash social transfers exceeded consumption in these cases. It was necessary to adjust these observations or remove them in the interpretation of the results, so this assumption has fundamental implications on the index and will be further elaborated on in the study. Suffice to say that the vast majority of cash transfer social wages are consumed in the year in which they are distributed, and this study will not consider a savings rate on this income.

1.8.2 *E less CSi is earned income for consumption*

There are several elements of income that are not captured in the simplification above (dividends, rent, alimony, gifts, interest etc.). These additional elements are either not significant enough in magnitude to affect the results, or form the basis for capital formation and savings and not necessarily consumption. This can be illustrated in that the cash social transfer wages (CSi) plus earned wages equals more than 100% of consumption from income decile 5 -10 (table 6). Interestingly income decile 1 – 4 consume significantly more than they earn in total income (table 2), to a maximum of 230% of total income consumed in income decile 1. This is because consumption captures other factors such as subsistence farming and bartering of goods and services. These elements can be considered earned income as they require work to generate. There is also no tax implication in this substitution as income

tax is only paid after income decile 6. Therefore, this is a fundamental yet acceptable assumption. Due to the importance of this assumption it will be discussed several times throughout the report.

1.8.3 Free services subsidised by the state

It was assumed that if a household has access to an electrical connection and are below income decile 6, that household is assumed to be receiving 50kWh of electricity and 6,000l of water for free per annum. It was necessary to assume a value for this function, as it was not captured by the IES. These are the minimum service entitlements of poor citizens since 2007 in South Africa. The results show that this element is a very small part of consumption, and does not affect the results materially. However, the price of services is increasing and although this measure is not significant in this study, it could well prove to be in future studies, or in other countries.

1.8.4 *Data used for this study is suitable*

It is assumed that the 2005/2006 Income and Expenditure data from Statistics South Africa, after it was cleaned by DataFirst, is a viable source of national household income and expenditure data. The integrity of the Statistics South Africa IES has been questioned before but is still considered reliable for research (Woolard and Wilson 2004). Despite this there are shortfalls and acknowledged differences between the Stats SA findings and those of National treasury (Statistics SA IES 2008). This sample is the third of its kind with 1995 being the first, 2000 being the second and 2005/2006 being the third IES iteration (Statistics SA IES 2008, www.statssa.gov.za). The 2005/2006 report is said to be the most reliable set of IES data, due to extensive geographic reach as well as the more reliable diary method as opposed to a once off questionnaire (recall method). As the reliability of the report is ultimately a function of respondents sharing sensitive information that might not be known exactly by the households, Statistics SA acknowledged the following biases (Statistics SA 2008, www.statssa.gov.za):

- Under-reporting of income
- Poor understanding of non-wage related income, thus often income from capital and subsistence produce is not captured as such
- The three IES iterations have been progressive, and thus each successive IES has differed methodologically and procedurally to improve the integrity of the study, but in the process making strict comparison between surveys very difficult

But the implications of these data irregularities will not be the focus of this study, the data will be assumed correct for analysis. It is recommended that a similar study, with the soon to be released 2010/2011 IES data, be conducted for comparative purposes.

2 LITERATURE REVIEW

This section contains a literature review of the key elements of the study, and the fundamental assumptions required for the mathematical model defined in the research problem statement. The definitions of household utility used in this study are presented, as well as literature around critical assumptions made for the construction of the dependency index. This will include literature on the constitution of social grants per income category and their direct effect on household expenditure and utility. The nature of the progressive income tax system in South Africa will then be discussed, as well as varying levels of consumption tax per income category. This literature review will construct a theoretical framework for the effective input of metrics into our mathematical model such that it retains internal validity. Finally the essence of the literature will be consolidated into three hypotheses to statistically test the internal validity of the dependency index.

2.1 The household utility of income from wages

Depending on the economic bracket of the household in South Arica the likely constitution of income streams is very different (McEwen, Kannemeyer and Woolard 2009). This can be seen in figure 2 below representing cash income streams of South African households per income decile in 2008.

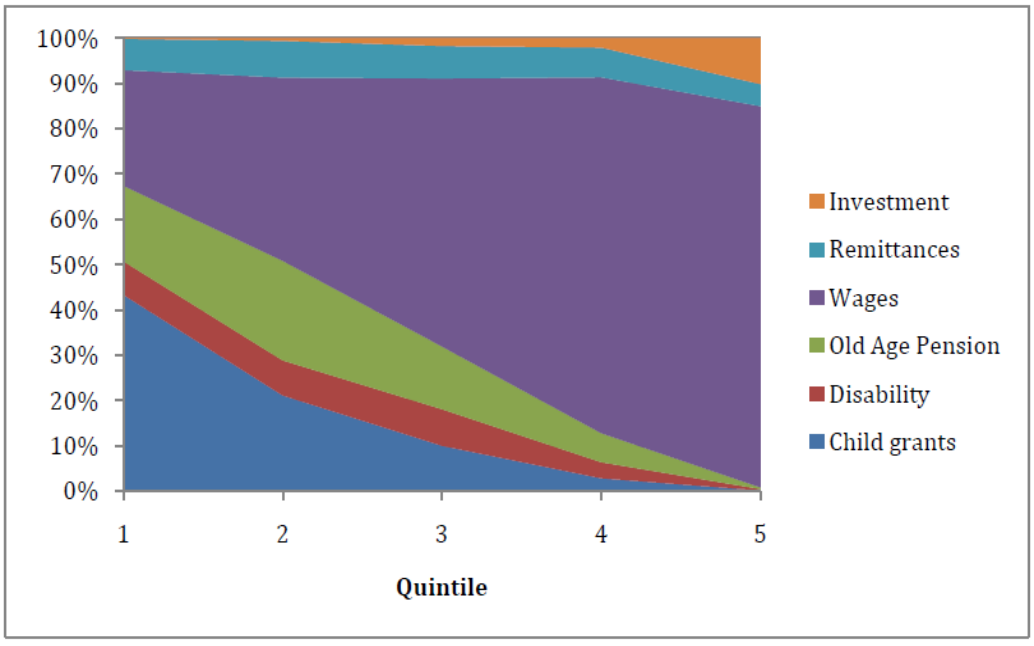


Figure 2: Components of cash income by income decile in RSA (McEwen, Kannemeyer and Woolard 2009)

As can be seen above, the percentage contribution of cash income streams is highly polarised, with quintile 5 receiving 83% of its income from wages whilst quintile 1 receives only 25% of income from

wages. Furthermore total social grants (child grants, disability and pension) constitute 67% of total income in income quintile 1 and only 2% in income quintile 5. Thus the total household utility (spending power and free services) relative to each rand of income earned through wages is very different. Furthermore income tax erodes the purchasing power of income derived from wages differentially depending on total yearly earnings (National Treasury and SARS 2011). The result is the utility of income from wages is increasingly reduced the more income is earned. These observations are the core of this study, and the review will add context to this overview.

2.2 The social wage

As shown in figure 1, the social wage in RSA is provided by two key mechanisms, cash contributions and subsidised products and services provided by the state from tax revenue. These different forms of social wage have differing effects on households, and this section will seek to position them in terms of the problem statement.

2.2.1 *Social grants in the form of cash*

Social cash transfer has a number of benefits as well as a number of negative artefacts. The social transfer of cash benefits from the economic multiplier effect. If these funds are spent and not saved they are multiplied through the value chain to promote broader employment and economic growth (Samson 2009, Heideman 2011). Conversely these cash transfers tend to be unconditional in southern Africa and promote dependency of households (Sabates-Wheeler and Devereux 2010), and are often used irresponsibly and not to the benefit of beneficiaries (Devereux 2007).

Other ways of allocating a cash social wage without these negative artefacts have been considered. For example the National Rural Employment Guarantee Act (NREGA) passed into law in India in 2005, where 100 hours of minimum wage labour is the right of every household through expanded public works programmes (Ravallion, Datt and Chaudhuri 1993). This encourages adults within the household to work and provide for the household through earned income and not unconditional cash grants. This of course is not appropriate for substitution of child and disability grants, which constitute the vast majority of cash transfers in RSA, but dependency and low levels of employment and productivity are a significant long term problem in RSA (Edwards 2001, Klasen and Woolard 2009).

It was estimated in 2009 that one quarter of RSA citizens live on less than $1.25 US per day (Woolard and Harttgen 2010). This is inclusive of all cash income, from wages to child and disability grants, but does not include any service components. In the aftermath of the 2008 recession household income decreased, particularly at the low levels of the income spectrum, whilst a sharp increase in

social spending was noticed (Woolard and Harttgen 2010). This has outlined the importance of the social security system in keeping citizens above the poverty line in RSA (Meth and Dias 2004; Ngandu and Altman 2010).

But the allocation of unconditional cash grants to individuals for specific purposes such as child and disability grants have been reported to have widespread dilution effects (Waddell 2002; Devereux 2007; Devereux 2010; Lindholm and Rosen 2000). The expectation is that this money will be spent on the beneficiaries, but often through the dilution effect this becomes part of the meagre household income or in some cases squandered by irresponsible members of the household (Devereux 2007). Thus the use of unconditional grants can either be for positive choices (healthcare, food etc.) or negative choices (use of grants not benefitting household). In 2008 the child grant was R240 per child per month (McEwen and Kannemeyer 2009), and constituted 42% of total household income in quintile 1 (figure 2). Roughly 60.5% of children under 14 in RSA receive child support grants (94% of total child grants), the value of which constitutes the largest source of income in income deciles 1 and 2. In total, social grants increased by R12.2bn in 2010, which includes the extension of child grants to 18 years of age (Ngandu and Altman 2010). Disability grants are also cash transfers, and are paid to roughly 5% of the population, and constitute between 7-9% of income up until income decile 3 (McEwen and Kannemeyer 2009).

Therefore due to the magnitude of these cash transfers as a percentage of household income, it would be naïve to believe that the use of such funds would be exclusively for the beneficiaries. As a result the cash transfer social wages have been considered in this study simply as an income input into the household irrespective of its designation.

Below in figure 3 is the breakdown of government spending (in billions) across the different functions of social assistance in South Africa. This is important in understanding the fiscal trend in social service provision that will underlie the construction of our mathematical model.

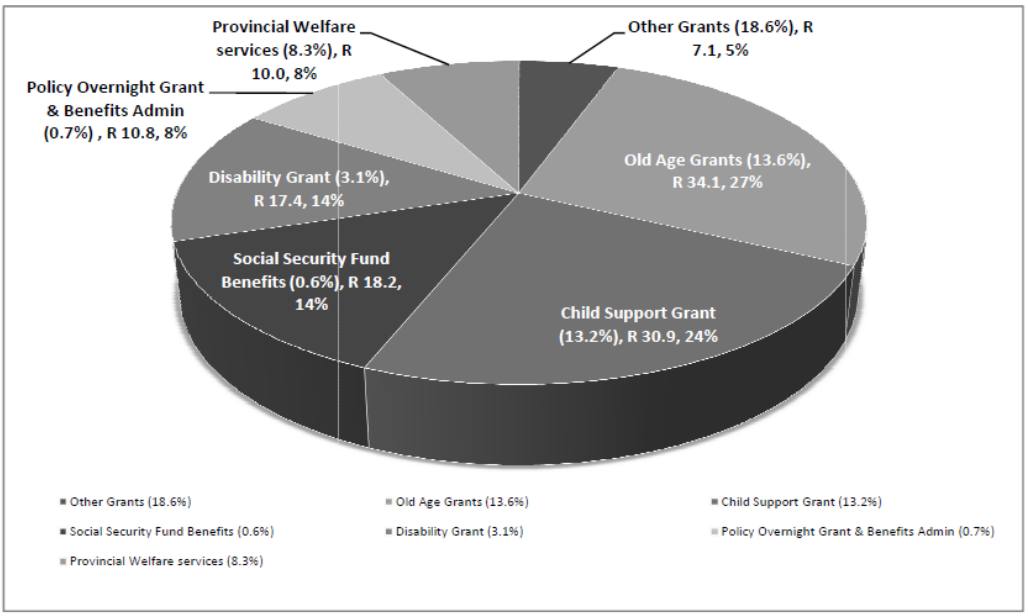


Figure 3: Disaggregated RSA government social protection expenditure 2010/11 (National treasury 2010; Ngandu and Altman 2010).

This figure represents the expenditure of the RSA government in billions per social protection function. The percentages reflect the increase in expenditure relative to the previous year 2009. Social services now constitute 54% of total government expenditure, which is 9.3% up on 2009 (Ngandu and Altman 2010). As can be seen the increased expenditure is above inflation even when considering that roughly 6.8% more citizens gained access to social grants in this period.

2.2.2 *Social grants in the form of services*

There is strong political and social support for the delivery of fundamental services to all South African citizens as a basic human right. The rationale behind this is solid, as expanding services to rural areas increases access of these markets to trade, which increases economic inclusion and ultimately tax revenue (Rogerson 2009). And more importantly the provision of basic services (water, sanitation, housing and electricity) improves general living conditions, which contributes positively to household health and improves the efficiency of education (Samson and Babson 2002). Increasing the human capital of a country is arguably the single greatest macroeconomic requirement for long run economic growth and development, and a healthy educated population is one of the key development indicators (Glaeser, La Porta, Lopez-De-Silanes, and Shleifer 2004; Hall and Jones 1999).

The prevalence of violent demonstrations associated with service delivery, which is a fundamental tenet of the ANC freedom charter, is increasing in underdeveloped areas (www.africa.fnst-freiheit.org/news/south-africa-will-there-be-an-arab-spring, www.polity.org.za/article/the-reasons-behind-service-delivery-protests-in-south-africa-2009-08-05). It would seem that South African citizens in rural

and underdeveloped towns are becoming increasingly dissatisfied with the lack and degree of state subsidised housing, water, sanitation and electricity that is being provided. But despite this dissatisfaction and call for increased delivery, large sums of state revenue have been spent and continue to be spent on improving service delivery (Ngandu and Altman 2010). It is estimated that 93.8% of households have access to clean water in RSA, according to the reconstruction and development programme standards (National Treasury 2010). Thus it is likely that RSA will achieve its 2014 millennium goal of access to potable water for all citizens. Each year nearly half a million citizens gain access to water, and in 2011 the RSA government (through municipalities) will invest a further R2.6bn into the development of water infrastructure in support of its universal access to water and sanitation goal (Ngandu and Altman 2010). The component of interest is the value of subsidised water provided to households from the state per decile, as infrastructure and connections are a public good but water has a market value per unit. The state allocated R1.2bn for the subsidisation of water in poor communities in 2010 (Ngandu and Altman 2010).

As with all service grants, it is not possible to determine whether these households would indeed allocate this value of money for the provision of this service, and thus its value is implied. Literature exists (Aaron and McGuire 1970) stating that the value of a provided good or service is equivalent to the value attributed to it by the recipient. This is a very difficult exercise and highly interpretive, however, other literature (de Wulf 1975; Cornes 1995) uses the approach of valuing public goods based on their marginal cost. This would require the profit structure of the service provider to be known. So the average per unit price will be charged to consumers multiplied by the number of free units (6000 litres of water per household per annum) as the value and utility of the service. This is a safe assumption as long as the provision of services is in line with requirements for human capital development and is a relevant allocation of household income.

Similarly there has been a large increase in access to electricity, the number of households that had access to electricity increased by nearly five million between 1994 and 2010 (National Treasury 2010). Currently 74.9% of households have access to electricity. In 2010 R6.7bn was allocated to municipalities for the subsidisation of electricity to the poor (Ngandu and Altman 2010). The Integrated National Electrification Programme currently allocates 50Kwh of energy free to households in lower income brackets. As with water service provision the measure will not incorporate the investment in infrastructure, but the value of the units of electricity provided for free by the state at their average chargeable rate. Remembering this study is focussed on the utility (what value is consumed) of household income relative to earned income, and not the transfer of income between deciles.

The only other appropriate consideration for public services that have a "saleable" market value is free housing provision, as education and healthcare are not resalable and are fundamental social considerations available to the entire population. Also all food aid is assumed to be consumed.

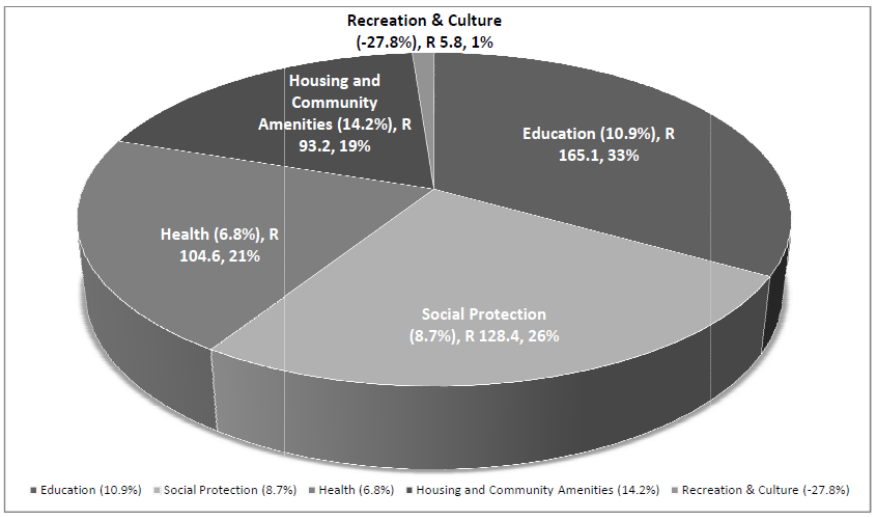


Figure 4: Government spending on social services 2010/11 (National treasury 2010; Ngandu and Altman 2010)

As can be seen housing and community amenities makes up a considerable amount of the national expenditure. The total value and number of houses built can be determined, as well as the entire amount of expenditure on this function (national treasury 2010). The challenge will be defining what the value of this subsidisation is in terms of an annual income market value to the household, as the asset (house) will most likely constitute the vast majority of that households net value in the lower income deciles. It is postulated that this will skew the data considerably, and housing would not be appropriately translated into income equivalents as the households were likely not paying rent before. Therefore, free housing although a large cost to the state will not be considered in the context of social grants.

2.2.3 *Summation of the social wage:*

The study proposes child grants, disability grants and pension grants are included as direct cash transfers that will be considered diluted household income enjoyed not just by the beneficiary. The inclusion of subsidised electricity and water to households as direct utility equivalents is also proposed.

2.3 Household tax

Broadly speaking there are two types of tax that affect households, income tax and consumption tax. Income tax in RSA is applied (2010) from a minimum annual income of R60,000 - R122,000 at a marginal rate of 18%, and increases to a maximum rate of 40% once annual income exceeds R490,000 (see Annexure A; Tax table). This is known as a progressive tax system (Samson 2009), as the tax rate increases based on personal income, assisting with wealth distribution and providing universal services. Several countries have a top tax bracket that exceeds that in the RSA (examples are England with 50% and Brazil with 65%). If employees earn less than R60,000 per annum their employers are required to pay a standard income tax on employees (SITE) on their behalf. The state receives the majority of its revenue from personal income tax, over 95% of which is collected through PAYE, as can be seen from figure 5 below (SARS 2010).

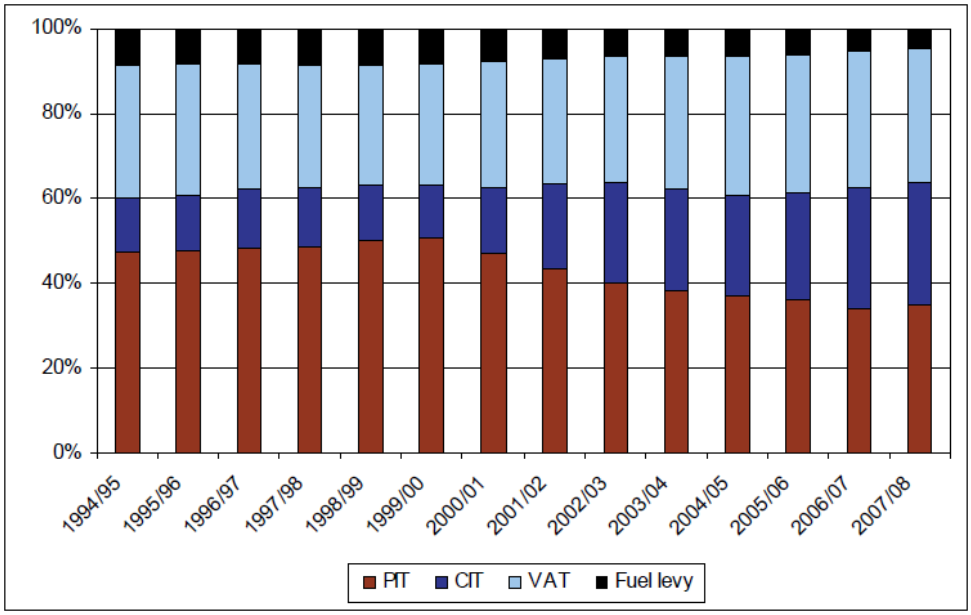


Figure 5: Composition of main sources of tax revenue 1995-2008 (National Treasury and SARS 2011)

Personal income tax constituted 34.3% of revenue collection, marginally larger than the contribution of VAT. The percentage of revenue generated by corporate income tax has steadily increased since 1994 and now constitutes 21% of overall tax revenue. It should be noted that over 98% of government generated revenue is from tax revenue, which is handled by the South African Revenue Services, SARS (National Treasury and SARS). South Africa has a population of just over 50 million people of which 5.9 million (11.8%) paid income tax in 2010, up 74% from 3.4 million people in 2002 (National Treasury and SARS). The significance of capital gains tax is beyond the scope of this study and will not be discussed.

VAT is a consumption tax levied at 14% (Value-Added Tax Act 1991) by registered vendors on the majority of goods and services provided. There are 19 basic food-stuffs and paraffin that are exempt from VAT, due to the high percentage of income spent on these products by the poor (National Treasury and SARS). Despite the targeting of vulnerable households, on the whole the middle class benefit from these subsidies the most, as they have greater purchasing power (National Treasury and SARS).

Thus VAT is only levied when purchasing goods or services with post tax income. This has a couple of interesting implications. Firstly households in the bottom three deciles have cash grants constituting the majority of their income, which then is eroded in terms of purchasing power by VAT at point of sale. So the state is in effect collecting a portion of its own allocated grant, which is inefficient. Secondly VAT as a percentage of income constitutes a larger relative percentage to poor households. The Katz commission found that zero-rating VAT in RSA reduces the low income households VAT burden by 18%, compared to only 6% in the highest income households. It is for this reason that VAT is often referred to as a regressive tax, but it is an invaluable source of income to the state which in turn provides social services with the majority of its revenue.

There are a number of challenges in measuring the overall tax rate per household. Measuring VAT per household is difficult because not all expenditure occurs through registered vendors and some products are exempt of VAT, and thus the household consumer basket and channel will affect total VAT paid. It is difficult measuring income tax per household as income tax is applicable to individual income and not household income, thus there could be more than one provider of household income and calculating average tax from average household income becomes difficult. But calculating the percentage of income tax as well as consumption tax paid per household is critical to the success of this study. Although Statistics South Africa have published information on per decile tax, these figures differ from those published by national treasury and SARS (Stats SA IES 2006), therefore a theoretical construction of tax based on raw data, the SARS tax tables and consumer basket constitution will be constructed.

2.3.1 *Summation of household tax:*

The study will use published tax and rebate models from SARS 2010 in conjunction with national income and expenditure data to calculate a mean per decile tax rate, separated into income and consumption tax.

2.4 Concluding hypotheses from the literature review

2.4.1 *Hypothesis 1*

The first hypothesis is that the utility of earned income, as defined by the mean household dependency index, will be statistically different per income decile.

Figure 2 measures the source of income to the household by income quintile. The trend of the percentage of total income from social grants (pension, diability and child grants) is a downward linear progression ranging from 67% in income quintile 1 to around 2% in quintile 5. The model under consideration assumes that a negligible amount of social wages are saved, and as a result can be deducted directly from VAT adjusted consumption. And as discussed in the literature review, cash social grants are considered as income to the household for general consumption, due to their magnitude relative to earned income in some cases. If this is the case then all cash social grants can be deducted from consumption to get earned income. Thus the percentage of social grants to total income should differ per income decile.

Income from wages is also taxed at various rates depending on the income category of the earner. Thus it is expected that the amount of income tax paid will increase exponentially in the last few income deciles, whilst being zero for the rest. This is the case as there are only 5.9 million registered income tax payers in a population of 50 million people. VAT is also charged at a differential rate depending on the income group of the household, as the constitution of the basket changes considerably and certain goods are VAT exempt. Finally households in lower income groups are provided with an amount of free water and electricity per month, which could be a significant component to utility in the lower income groups.

Thus as the proposed index will consider consumption, total social grants, income tax, consumption tax and free state subsidised services, all of which are different per income group. It can be assumed that the index itself will be different per income group.

This leads to the first hypothesis:

H1: The mean Dependency Index is statistically different per income decile.

2.4.2 *Hypothesis 2*

The second hypothesis is that the trend of the mean dependency index per income decile will fit an inverse curvilinear curve, increasing from decile 1 to 5/6 and decreasing to decile 10, with significance and accuracy.

Having hypothesised that the dependency index will be different per income decile, this hypothesis provokes the question of the trend of these differences. As the delivery of state services (social and infrastructural) are dependent on access and localised resources, they are not distributed equally amongst the poor. For example, a household living in rural Transkei (rural Eastern Cape) could be no different in personal wealth generation to a household in Soweto (Gauteng, Johannesburg). But it would be less likely that the rural household would have free basic services and easy access to cash social wages. Thus the household in Soweto could be in a higher income decile without a fundamental difference in earning capacity other than access. It is assumed that households in income decile 1, below the poverty line, would have very little access to such services (IES Analysis), and thus their index would remain relatively low compared to households in a more urban location. Similarly households in income decile 7-10 who have purchasing power eroded by income tax, whilst receiving little social assistance, are likely to have the lowest index scores. Resultantly income decile 2-6 should have a higher mean dependency index than those at the margins at income decile 1 and 10.

Also due to the progressive /regressive nature of the inputs, it is assumed the trend will be relatively smooth, as the state mechanisms target particular income deciles, which will be located next to one another. For example income tax targets the top three income deciles with increasing rates, thus there is an exponential burden of tax from decile 7-10. All the inputs seem to follow a trend clustered around the income categories, thus it is assumed that these trends will result in a smooth mean index trend with a particular inverse curvilinear shape. As the index will likely be a minimum at income decile 1 rising quite sharply through to income decile 6 (because of increasing access to social grants as well as social delivery), before beginning to taper off to an absolute minimum at income decile 10 (because of increasing tax incidence with little or no access to grants), an inverse quadratic or PolyRatio(2,2) curve is expected to fit the model with significance and accuracy. Suffice to say that if the index does follow a smooth inverse trend along these lines, the dependency index is measuring its intended construct as outlined above.

This leads to the second hypothesis:

H2: The mean dependency index per income decile will follow an inverse curvilinear trend increasing from income decile 1 to 5/6 and then decreasing to decile 10.

2.4.3 *Hypothesis 3*

The third hypothesis is there will be a strong downward trend in the mean dependency index within income decile 10, which will be measured as the top ten economic percentiles.

The challenge with measuring income decile 10 as one discreet decile, is there are likely large variations in per household index scores within this decile. The reason for this assumption is that any

household receiving more than R180, 000 per annum in income is considered to be in the top income decile. This total household income could come from more than 1 individual in the household. For example the poorest possible household in income decile 10 earns R180, 001 per annum, if there are two equal providers of income to that household the household would have the minimum possible income tax rate of 18% (Annexure A). But the richest household in income decile 10 with a single earner will be taxed at a rate of 40%. As discussed social assistance is not significant in this decile, and it is likely that the consumption tax rate would not be considerably different either. Thus as the entire spectrum of the SARS progressive tax rate could be observed in income decile 10, ranging from 18%-40% (122% increase), it is advisable to break this income decile up for greater granularity of analysis. From this information it is assumed that the top income decile, broken into ten percentiles equal in magnitude, will demonstrate a significant downward trend from the 90th to the 100th percentile.

This leads to the third hypothesis:

H3: It is likely that the trend of the mean dependency index within Decile 10 alone will be downward sloping ending at a minimum in the 100th percentile.

3 RESEARCH METHODOLOGY

This section outlines the methodology that will be used to conduct this research and answer the problem statements outlined in chapter 1 and the hypotheses constructed in chapter 2. The source of data for this quantitative analysis is the 2005/2006 IES data cleaned by DataFirst, a division of the University of Cape Town specialising in survey research in Africa (www.datafirst.uct.ac.za). From this raw data set a number of decile specific constructs were mathematically modelled. This section gives an overview of the research design and the model that will be created to answer the main problem statement. Issues surrounding the data capturing of the original samples will be discussed, and the internal and external validity of the study will be assessed.

3.1 Research methodology

The outcome of this study will be a multiplicative index score that can determine the utility of each household relative to each rand earned depending on the income decile in which a household is located. Thus for this index to have any value commercially or interpretatively, the research must be rooted in the quantitative paradigm.

The use of quantitative analysis is fit for purpose for a number of reasons. It allows us to draw direct cause and effect relationships without considering additional webs of influence (Johnson, and Onwuegbuzie 2004). It also allows us to generalize our findings as long as our sample is of sufficient

size, and is also therefore the best method for analysing large numbers of people (Johnson and Onwuegbuzie 2004). And importantly the results tend to have higher credibility with people in power, administrators, politicians etc. (Johnson and Onwuegbuzie 2004).

Despite the appropriate use of quantitative analysis, there are some caveats that need to be considered. In the construction of the categorical variables and underlying quantitative assumptions, there can be disparities between the researcher's theories and assumptions and reality (Johnson and Onwuegbuzie 2004). This can be avoided through careful consideration of published theory and measuring the outcomes without intent or bias. But the key challenge is ensuring that the results are applicable contextually and not just as an abstract model of very little real world consequence (Johnson, and Onwuegbuzie 2004).

3.2 Research Design

This study uses quantitative mathematical modelling to derive unique insights from already published data. A useful understanding of mathematical modelling was defined by Berry and Houston in 1995 as the "...translation of a problem into mathematical form to solve real-life problems" and by Keramati in 2011 as "...the art of translating problems from an application area into tractable mathematical formulations whose theoretical and numerical analysis provides insight, answers, and guidance". Similarly the hypotheses proposed can only be solved through the modelling of decile specific data of income, expenditure, tax and social grant averages in South Africa.

It is appropriate to use this methodology because the data is available at an acceptable level of integrity, and constructing a mathematical model of the relative buying power of earned income will add value to the current body of knowledge. In this case income inequality is a real life challenge in South Africa, and if this model can add insight into the perpetuation of this phenomenon then it will have fulfilled both definitions of effective mathematical modelling mentioned above.

There is no published literature on the construction of the proposed model, and thus cross literature referencing has proved difficult. There is supporting literature for the underlying constructs, however their combination is novel. This research report therefore seeks to design a simple construct that introduces the concept of the relative utility of income for further study and comment. It is by no means meant to be an exhaustive analysis of the forces and reasoning for income disparity, or mechanisms by which these challenges can be overcome.

3.3 Population and sample

3.3.1 *Population*

The population in this study is that of the Republic of South Africa. The population of RSA was 47 million in 2005 and 50.59 million in 2011 (Statistics SA). The country is separated into 9 provinces with the majority of people living in Kwazulu-Natal, Gauteng and the Western Cape. The demographics of the country show 79.3% of the population are African, 8.9% Coloured, 9.2% White and 2.6% Asian / Indian (Liebbrandt, Woolard and Finn 2009). Wide income disparity exists across these demographic groups, with 97.6% of income decile 1 being constituted by the African majority but only 27.5% of income decile 10 (Liebbrandt, Woolard and Finn 2009). Conversely the White minority constitute less than 1% of income decile 1 but over 56% of income decile 10 (Liebbrandt, Woolard and Finn 2009). These figures have been taken from the NiDS Wave dataset, and will be used as a benchmark to compare the internal validity of the Stats SA sample to ensure there is representivity across these groups in the sample.

Although the population being analysed is that resident within South Africa, it might be possible to generalise these results to other countries with similar socio economic landscapes. Care should be taken in analysing these results without considering fundamentals of these foreign populations however, as any bias in the core measurements in the dependency index could skew the data significantly. For example if aide is distributed as physical goods and services and not cash equivalents, a stark difference in the index could be observed. This could be adjusted for by considering the cash equivalents of those goods and services, but further assumptions and considerations would be required.

3.3.2 *Sample and sampling method*

The sample is a selection of 24,000 households taken from 3,000 distinct primary sampling units (PSU) to represent the spread of household income and expenditure (8 samples are taken from each of the 3,000 PSU's). These households are then divided into 10 distinct groupings each containing roughly 2,100 households from the poorest to the richest households (lowest income to the highest, or income deciles 1 -10 respectively). The reason that income deciles were constructed instead of evaluating the raw household data was that the implications of this study are most relevant for policy formation and evaluation. In this context income deciles are a good measure of spread of disparity within an economy, and a methodology adopted by Statistics South Africa the public statistics bureau (Statistics SA).

The sampling method for the primary data was a diary method where a fieldworker visited all sampled households on five separate occasions, filling in an aspect of the survey questionnaire each visit. Households were required to account for all goods and services procured over the 11 month duration (IES 2008). The households were required to keep a daily diary for the purpose of recording the data more effectively and accurately. These samples were then consolidated into the Statistics SA IES. The data used is a cleaned version of the Stats SA IES data, acquired through DataFirst research institute in Cape Town (UCT). The remaining data file consists of only 21,144 data records, as the balance were removed in the cleaning process.

It is important to compare the sample to other published datasets to ensure that the sample is representative of the South African population, and will retain internal validity in this regard. The percentages of population per demographic group, as well as per income decile of the sample were compared to those published in the NiDS 2009 studies (Liebbrandt, Woolard and Finn 2009). The proportions were very similar (Table 20, Annexure A) and indicated the sample was a good representation of the population. For example the percentage of African households in income decile 1 was 94% in the sample and 97% in the NiDS study, and 24% in the sample versus 27% in the NiDS study for income decile 10. Similarly White headed households constituted 0% of income decile 1 in the sample and 0.6% in the NiDS study, and 58% in the sample versus 57% in the NiDS study for income decile 10. The overall population demographics were also similar. This shows that the sample, for the metrics under consideration, is a suitable representation of the population under consideration.

3.4 Measures

The research instrument will require the construction of several metrics for each household per income decile, namely average income from wages, average expenditure, average tax paid (separating income and consumption tax) and average social grants received (separating service subsidization and cash contributions). Although some of these metrics are available directly through the data import, some of the metrics need to be transformed or created as new fields for statistical analysis. The purpose of these fields is to create the dependency index per household.

$$Dependency_Index = \frac{E*(1-Tc)+SSi}{(E-CSi)/(1-Ti)}$$

E = total household consumption (R/annum)

Tc = consumption tax as (%)

Ti = income tax (%)

CSi = cash social transfers by state (R/annum)

SSi = subsidised water and electricity (R equivalents / annum)

3.4.1 *Average total expenditure per decile (E)*

Consumption is the key measure of utility in this study, and thus annual per household consumption needed to be determined. There appear to be artefacts between income and expenditure due to generally underreported income, savings, credit and capital (Oosthuizen 2007). Thus in a number of instances expenditure exceeds income, which seems counterintuitive. There are acknowledged biases by Statistics South African in this data set (IES Analysis), including underreporting of income, and the separation of income into its correct sources. The method of measuring consumption was made more robust however, through the introduction of the diary method, and increasing the number of household consultations. Also non-income sources of consumption such as subsistence farming and bartering of services and goods, although outputs of work, are captured in the consumption variable and not the income variable. Thus in the index the expenditure variable **(E)** is used as the most significant measure of work.

3.4.2 *Income earned from wages per economically active individual in the household (Calculation of Ti):*

It was necessary to have a measurement for each individual within the sample of households that earned income through wages. This was necessary in order to create the field for tax rate per individual (Appendix A). These tax rates were then weighted and averaged in the consolidation of individual data records into household data records through the unique identifier Uqno. This then gave the average income tax rate **(Ti)** for that household, which is one of the key metrics for the dependency index.

It was also necessary to choose a starting point for income tax, as the first tax bracket is theoretically from an income of R0. There are however only 5.9 million income tax payers in the country, therefore decile 1-6 were exempt from all income tax (earning between R600 – R5,000 per household per month). It is likely that there are tax payers in income decile 6, and even potentially decile 5 which would reduce the value of each rand of income that they earn, resulting in a lower index score. Similarly there will be individuals within the top three income deciles that do not pay tax. Thus to match the number of tax payers to a representative sample, income tax was only calculated for income decile 7-10. Unfortunately the tax figures collected by Statistics SA in the 2005/2006 IES were inaccurate, and could not be used with any reliance (Stats SA, National Treasury). Thus income tax

was constructed for the top 3 income deciles based on wages earned per individual, and then consolidated into per household tax paid. The integrity of the tax results would be further increased if the rand values of tax paid, obtained from SARS, per household were used instead of a flat fee, as tax is notoriously complicated.

3.4.3 *Average consumption tax paid per decile (Tc)*

It was necessary to have both a measurement for income tax and consumption tax (VAT), as they affect consumers in different ways. Consumption tax is paid for with after income tax income, and despite its official value of 14%, many products and services are exempt from VAT in order to assist the poor. Thus the average consumption tax per household will be different per income decile depending on the specific spending within that consumers basket of goods.

Data from the IES analysis was used to break down the actual spending per income decile into categories of spend, such as food, fuel, clothing etc. Discounted consumption tax weightings were then allocated to the various categories and multiplied through by consumption to obtain the average consumption tax percentage per household (Annexure A). This forms the metric **Tc** in the dependency index.

3.4.4 *Average cash transfer social grants received per household per decile (CSi)*

As discussed in the literature review there is a large occurrence of benefit dilution, where cash transfer social wages **(CSi)** become part of the meagre household income for general household consumption. Thus for the calculation of this metric all cash grants were added into one non-discriminating field called **CSi**. This includes cash transfers from child grants, social pensions, disability grants and grant in aid. An insignificant value of CSi is assumed to be saved per annum, according to the nature of the equation.

3.4.5 *Average subsidised social services in cash equivalents (SSi)*

Utility is not just defined as consumption in this study, but rather consumption less consumption tax plus subsidised social services (water and electricity). These figures were not available as such, and needed to be calculated accordingly. The state aims to provide free 50Kwh of electricity and 6,000l of water to each poor household in the country. Much of the information on the provision of such services was in 2007, so some regressive assumptions have been made. The average unit price of these free services was multiplied by the free service provision for every household that had access to an

electrical terminal and was between income decile 1-6 (Annexure A). These figures were then summed as total subsidised services or **SSi**.

The equation can be simplified into two concepts, household utility divided by required earnings for that utility.

3.4.6 *Mathematical model*

The equation is defined by each of the grouped terms independently, and then explained in terms of the resultant index. There are 2 terms in the numerator of the equation which will referred to as term A and term B respectively. Term A is the total per household expenditure **(E)** less consumption tax **(Tc)** to show the true utility of consumption. Term B then adds on the social service income **(SSi)** as utility derived by the household similar to after consumption tax expenditure.

The denominator also consists of two terms, which are referred to as term 1 and 2 respectively. Term 1 is total per household expenditure **(E)** less the full value of household cash transfer social wages **(CSi).** This results in term 1 defining the equivalent amount of earned income to pay for the balance of expenditure **(E)**. Term 1 is then divided by term 2 (1 less the income tax rate in that decile) to give the amount of income required to earn the balance of expenditure less **CSi**. Thus term 1 transforms expenditure to the amount of income required before tax to afford such expenditure. Term 2 reduces the total amount of income required for expenditure by the Cash Social Income received by that household which is used directly for expenditure. Thus the denominator represents the total amount of income a household in that decile is required to earn through wages in order to afford their expenditure. When the numerator is divided by the denominator a factor representing the buying power of each household relative to income earned is created. This factor is defined as the dependency index when measured per income decile.

3.5 Mathematical and statistical procedures

In order to construct the model and then to test the three hypotheses several statistical procedures were used. This section will give an overview of the procedures used in this study, as will be presented in the results section. The following are broad statements of the necessary procedures:

1. The fundamental issue was to create the Dependency Index for further analysis. This step is a simple arithmetic implementation of the index equation (see page 14) for each household. Once

the dependency index is created, some data cleaning is required for odd outliers, as discussed below.

2. Hypothesis 1 requires a comparison of the dependency index between each income decile. This requires an ANOVA-type approach.
3. Hypothesis 2 requires an analysis of the shape of the dependency index curve across all deciles with the assumption of inverse curvilinearity. This requires curve fitting procedures.
4. Hypothesis 3 requires an analysis of the shape of the dependency index for the top ten percentile, with the assumption of a downward trend. This requires curve fitting procedures.

Each of the methodological issues in executing these steps is explained in the sections that follow.

3.5.1 *Step 1: Creating the Dependency Index*

As indicated above, the first step is to create the actual dependency index which is a non-trivial process of data acquisition, creation and manipulation using ultimately the equation for the dependency index on p 14. I do not go through the exact procedure here, instead it is explained in depth in the results chapter.

In order to interpret the dependency index it was necessary to adjust the data for outliers, as they were skewing the data considerably. Due to the nature of the dependency index equation any household that receives more than 83% of its household income from social grants is considered an outlier. In this study the statistical technique of winsorizing was used, with a pre-determined rule determining which observations were outliers, these outliers were then capped at a maximum of 5 (discussed in greater detail later). This type of winsorizing is subjective to a degree, as the parameters for the winsorising are pre-determined and not dependent of the sample fundamentals. It is appropriate in this context however as the index score represents a ratio which past a particular point loses its meaning, for example when a household receives more social grants than total consumption the index is negative, and when 98% of income for consumption comes from social grants the index value exceeds 50. Rule based winsorizing is not uncommon (Hawkins, 1980), and if checked for assumption bias can be a very effective means of data manipulation. I chose a winsorized index score of 5, which allows the capturing of over 80% of the data whilst minimising the radical effect of outliers.

3.5.2 *Step 2: Comparing income deciles using ANOVA procedures*

a. *GLM ANOVA*

The GLM ANOVA was a critical statistical step in the study, as it was required to determine whether the income deciles had index values that were statistically different from one another. The GLM ANOVA performs an analysis of variance on the dependency index of each household per income decile, and then compares the means per income decile to see whether they are statistically different from one another. The only relevant limitation is that the data is assumed to be a representative sample of the population, which has been demonstrated in the population and sample section. This procedure was essential to answer the first and last hypotheses. Two different sample iterations were considered using the GLM ANOVA, a winsorized iteration and an iteration with all the outliers removed. A threshold index value of 5 was chosen (will be discussed in detail) for both of the iterations, as this represented an appropriate cut off point after which 19% of the observations are present. The windsorized iteration capped all of these index values at 5, whilst the less outlier iteration removed all 19% of the observations entirely.

b. *One-way ANOVA Kruskal-Wallis*

In order to ensure that the groups were indeed different without any adjustment of the data (winsorized or without outliers), a nonparametric one-way ANOVA procedure was run. As this analysis was performed on the data including all the outliers, normality could not be assumed, but a random sample per decile could be assumed (discussed under sample and population). This including a large enough sample size and independent categorical groups allowed the use of the Kruskal-Wallis Z test with Bonferroni multiple comparisons (most stringent test). This test allocates ranks to the medians of the un-adjusted dependency index per household per income decile. This removes any large outliers, and gives a conservative trend of the index per decile. This procedure is primarily used to see whether the groups are different from one another and whether they follow a specific trend without any assumption bias.

3.5.3 ***Steps 3 and 4: Establishing the shape of the Index using curve-fitting***

This procedure seeks to fit a curve to the mean dependency index per income decile, so that the household dependency can be predicted simply by knowing what income decile a household is in. Thus in order to find a relationship between the index and the household income decile an appropriate mathematical equation will be mapped using the most appropriate curve. The result of this procedure will be a unique equation per index iteration, that calculates the dependency index score of a household depending which income decile that household is in. The NCSS general curve fitting

function will be used to determine which curvilinear iteration best fits the dependency index per income decile.

3.6 Interpretation

The index produced could constitute a useful metric for international macroeconomic comparisons, as well as microeconomic monitoring and recommendations. The model would be able to demonstrate a cross-section of the fiscal relationship between government and society per income decile. And in turn can be translated into the relative dependency of populations on the state, and indeed the relative dependency of the state on populations. In one graph the dependency index can show the impact of several metrics on one directly internationally comparable set of figures. The interpretation of the index will be done in the discussion section.

3.7 Limitations of the study

- The study is purely based on data that has been collected by Statistics SA. Although the cleaned data has greater integrity, it is still likely that inaccuracies will exist.
- Adjustments to outliers were required to build the dependency index, and although the non-parametric analysis was used to test the overall trend and differences, it is likely that the index will suffer from assumption bias around the thresholds selected for the index.
- The index does not take into account savings and wealth, only income for consumption purposes. Thus the index is not appropriate to demonstrate income disparity, rather the purchasing power of earned income.
- Close on 20% of the data samples were removed or adjusted in order to counteract the influence of outliers. It was not considered whether these adjusted or removed variables could have affected certain population demographics more than others, resulting in a non-representative sample group. It was shown that the outliers were spread across the deciles, but with concentrations near the bottom / middle of the grouping.
- Assumptions on tax compliance and income earned through wages had to be made in order to obtain the average tax payable per household. Income is known to be underrepresented, and not all eligible household pay tax, so there will be some discrepancy here. It is likely that integrating SARS information with the IES would yield better results, although the figures are not strictly comparable.

- More information is available on the state of the poor in RSA currently, including levels of service delivery and access. Thus this study would benefit from being redone using the 2010 IES when it is released in conjunction with more available information.
- The simplicity of the model in using income deciles as the only categorical separator. There is likely a large number of additional separators such as geographic, race and gender variables that could add granularity to the study and inform the context wherein each of these groups operates.

3.8 Validity and reliability

3.8.1 *External validity*

External validity looks to ascertain the degree of generalisability of results from one research sample to others, as well as how applicable it is to generalise the results of the sample to the population outlined in the research hypothesis (Kidder Judd and Smith, 1986). In this case the generalisability of the sample to the national population of South Africa is appropriate, as the data set was designed for this purpose and the population proportions are similar to other studies. Thus the model is appropriate, but individual households might find it difficult to characterise themselves in terms of income and expenditure deciles. Also households might find themselves at the margins of a decile, and the results could differ due to the deviation of the household from the average of the decile. The mathematical model will not provide insight into populations outside of South Africa however, as the metrics used are specific for the South African context. Having said this, the model itself is robust, but the data specific to that region needs to be integrated accordingly. Thus the study is externally valid in terms of the South African population in question in the study, but not externally valid for other populations without changing the underlying data.

3.8.2 *Internal validity*

Internal validity is defined as the integrity "with which statements can be made about whether there is a causal relationship from one variable to another in the form in which the variables were manipulated or measured" (Cook & Campbell 1979, p. 38). In this case the data is not being collected by us, so in terms of data collection we are assuming internally validity. The construction of the mathematical model will require simplifying assumptions however, and herein is the greatest potential for invalidity. The study attempts to construct the model without bias, and with all significant assumptions stated and explained. Statistical significance of a number of iterations of the construct will

also be attempted such that readers can make up their own minds as to the impact of the results. Internal validity is a central theme in the discussion of this study, and will be discussed at length in the remainder of the document.

3.9 Reliability

Reliability is a measure of repeatability and consistency of empirical measurements (Zeller and Carmines 1980). In this case the reliability of the outcome will be as reliable as the input data and metrics. The experiment in terms of methodology is highly repeatable, but a separate question could be asked of the data collection process. There has been large variation in results between the three published IES reports, which indicates that reliability of the source data could be called into question. But having accurate input data would result in reliable outcomes. Thus to increase the reliability the data was sourced from a company that cleans statistical information for research in Africa. DataFirst based at UCT.

4 CHAPTER 4: PRESENTATION OF RESULTS

4.1 Introduction

In this section the results of the study are presented and described in context of the main research problem and three hypotheses. The research problem was to construct a model that predicts an earned income utility index, the dependency index. This model is applied to every household in the 2006 IES and then statistically analysed. The process of data importation, manipulation, transformation and integration for the calculation of the above mentioned index will be presented below in figure 6. Some relevant GLM ANOVA tables are presented to give context to the analysis as well as an explanation of key variables. Finally, the three hypotheses will be tested and presented through GLM ANOVA, one-way Kruskal-Wallis non-parametric ANOVA and curve fitting statistical procedures. Three iterations of the index are presented and explained in line with the objectives of the study.

4.2 Creating the dependency index (Main problem 2.4.1)

To construct and analyse a model (the dependency index) that predicts average household utility relative to income earned per income decile in South Africa.

The data used in this study was cleaned data from DataFirst available in two relevant files. The first file represents every economically active individual (n=47,150) within the household sample group and their sources of income The other file contains consumption and income totals as well as many other service metrics per household (n=21,144). Figure 6 below shows the process and extent of the data consolidation and transformation undertaken by the current author. In order to save space, I do not repeat the contents of Figure 6 in text, the process is easily interpretable from the figure.

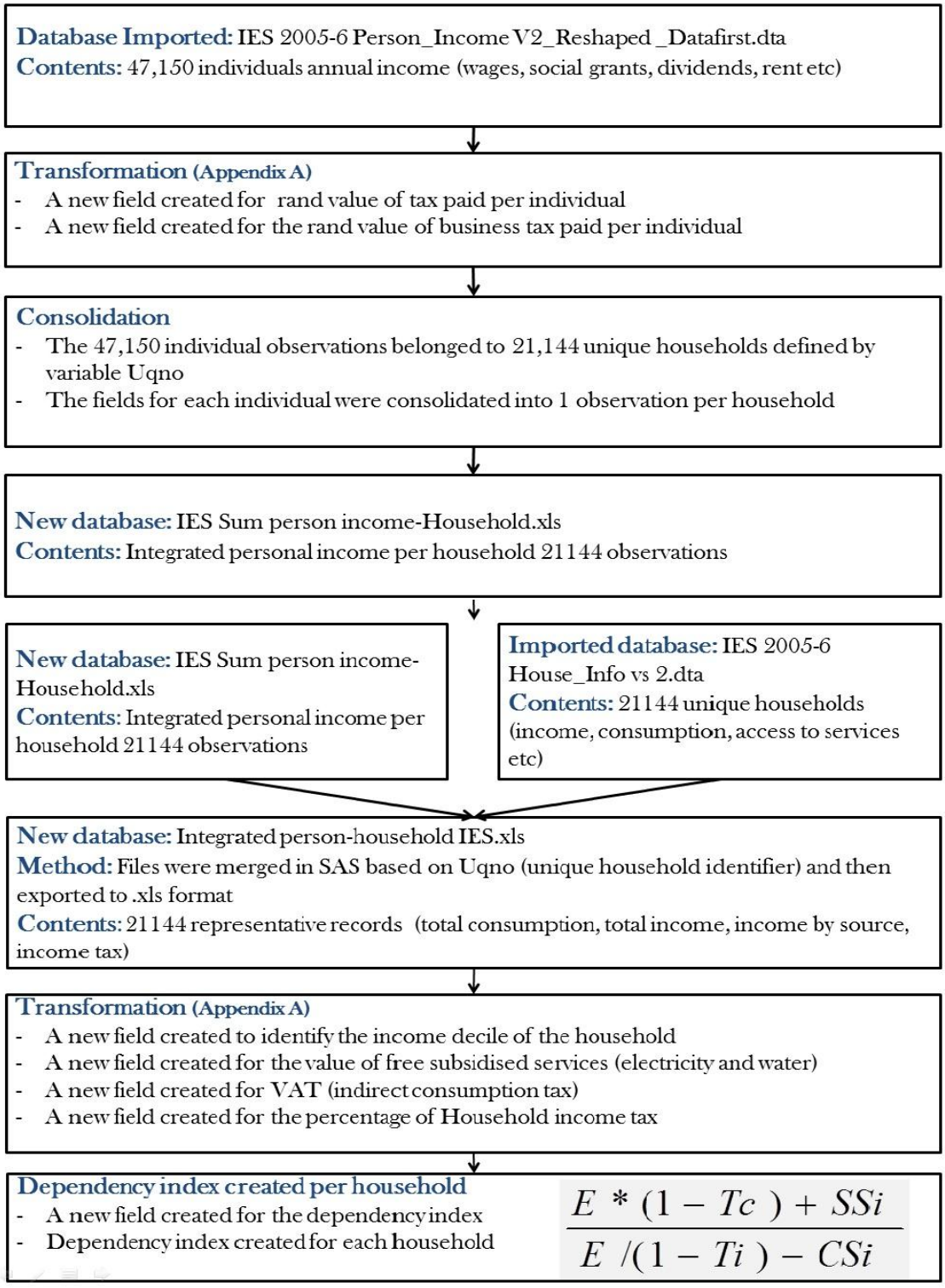


Figure 6: Process of data integration, manipulation and transformation

Critical to interpreting the statistics is understanding the income utility index that forms the dependent variable in all the hypotheses below.

$$Dependency_Index = \frac{E*(1-Tc)+SSi}{(E-CSi)/(1-Ti)}$$

E = total household consumption (R/annum)

Tc = consumption tax as (%)

Ti = income tax (%)

CSi = cash social transfers by state (R/annum)

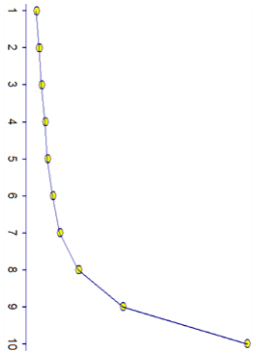
SSi = subsidised water and electricity (R equivalents / annum)

4.2.1 *Consumption (E)*

E represents total consumption, as it is a better proxy for utility than income for the purpose of this study, as shown in table 1 and discussed below.

Table 1: Consumption (E) / Decile:

Income Decile	Consumption as % of Income	Count	Mean	Standard Error
1	233%	2103	10431.88***	1269.98
2	140%	2461	13463.28***	1173.98
3	122%	2500	16146.96***	1164.79
4	112%	2309	19851.40***	1212/01
5	97%	2272	22216.47***	1221.84
6	91%	2151	27813.47***	1255.73
7	81%	1981	35414.92***	1308.50
8	79%	1952	55160.08***	1318.19
9	80%	1848	102343.60***	1354.77
10	62%	1566	233620.60***	1471.71



***P<0.01*

As can be seen from table 1 above, the mean consumption per income decile increases exponentially toward decile 10. But the *consumption as % of income* column shows that until income decile 5, consumption is greater than income. Conversely from decile 5 until decile 10, consumption is less than income, which indicates savings to a maximum of 38% in income decile 10.

There are acknowledged biases by Statistics South African in this data set (IES Analysis 2008), including underreporting of income, and the separation of income into its correct sources. The

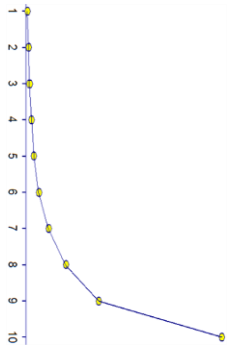
method of measuring consumption was made more robust however, through the introduction of the diary method, and increasing the number of household consultations (Stats SA). Also non-income sources of consumption such as subsistence farming and bartering of services and goods, although outputs of work, are captured in the consumption variable and not the income variable. Thus in the index the expenditure variable is used as the most significant measure of work (less any direct cash transfers) and utility (plus subsidized services).

4.2.2 *Income Tax (Ti)*

Interestingly when wages are measured as a function of total consumption, they never exceed 100% as can be seen in table 2 below.

Table 2: Wages / Decile

Income Decile	Wages as % of Consumption	Count	Mean	Standard Error
1	11%	2103	1095.39***	1421.08
2	20%	2461	2752.31***	1313.66
3	21%	2500	3449.24***	1303.38
4	29%	2309	5805.81***	1356.21
5	37%	2272	8160.45***	1367.21
6	49%	2151	13512.67***	1405.14
7	68%	1981	24162.89***	1464.19
8	78%	1952	42799.36***	1475.03
9	77%	1848	78733.60***	1515.96
10	91%	1566	211888.00***	1646.81



***P<0.01*

This indicates that all excess income as represented in table 2 between income decile 5 and decile 10 (3% - 38% savings respectively) cannot come from income from wages, as these never exceed consumption as shown in *percentage of total consumption* above. This is also potentially a sign of underreporting and misallocation of income, as well as individuals who own businesses having income from other locations other than wages.

Income tax (Ti) is another very important variable within the index. This variable was very poorly captured by Statistics South Africa in the 2006 IES, and it differed in orders of magnitude from those published by national treasury (National treasury 2008, Stats SA IES 2008). But in order to ensure the income tax was appropriate for the data set, it was constructed per individual on wages, and then summed and averaged across the contributing members of the household (Annexure A). Thus the

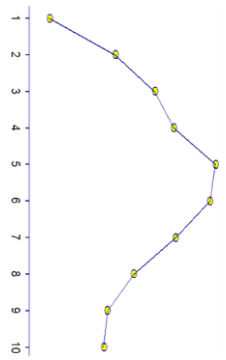
wages variable was transformed from the wages measurement in the data set, against SARS tax brackets (SARS 2010). It is likely that tax is thus underreported due to the underreporting of income and wages.

4.2.3 *Income from cash social transfers (CSi)*

South Africa has a progressive welfare policy that seeks to assist those who are in the greatest need of financial aid. The mean of all cash social transfers has been plotted below in table 3.

Table 3: Social Grants (CSi) / Decile

Income Decile	CSi as % of Consumption	Count	Mean	Standard Error
1	8%	2103	871.25***	194.15
2	28%	2461	3704.39***	179.47
3	34%	2500	5426.04***	178.06
4	31%	2309	6252.27***	185.28
5	36%	2272	8063.25***	186.78
6	28%	2151	7822.11***	191.97
7	18%	1981	6327.91***	200.03
8	8%	1952	4512.79***	201.51
9	3%	1848	3365.08***	207.11
10	1%	1566	3209.41***	224.98



***P<0.01*

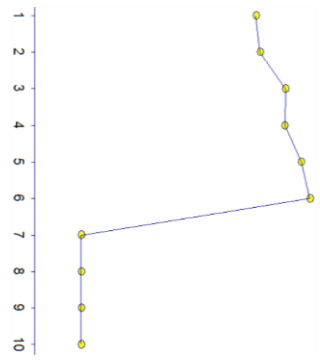
As discussed earlier and inferred by Hypothesis 2, the peak of social transfers only occurs in income decile 5, and constitutes the maximum percentage of total household consumption at 36%. Thus the state seems unable to deliver services and aid to the most needy population deciles, and the value of social wage provision seems to be distributed about the economic mean of national population. Interestingly income decile 9 and 10 receive similar social transfers to those in dire need at income decile 2 and more than those in income decile 1, which seems counter intuitive until viewed as a percentage of total consumption.

4.2.4 *Subsidized services (SSi)*

The cost of subsidized social services is primarily in the infrastructure and not the actual allocation of free water and electricity (McDonald, Pape and Council (2002). In this regard the free service constitutes a relatively small portion (maximum of 3.6 % in decile 1) of house hold consumption until decile 6, after which services are paid for as can be seen from table 4 below.

Table 4: Total Subsidized Services (SSi) / Decile

Income Decile	SSi as % of Consumption	Count	Mean	Standard Error
1	3.6%	2103	380.55***	4.95
2	2.9%	2461	390.03***	4.58
3	2.8%	2500	445.52***	4.54
4	2.2%	2309	444.58***	4.72
5	2.2%	2272	480.35***	4.76
6	1.8%	2151	498.97***	4.90
7	.0%	1981	.00***	5.10
8	.0%	1952	.00***	5.14
9	.0%	1848	.00***	5.28
10	.0%	1566	.00***	5.73



****P<0.01*

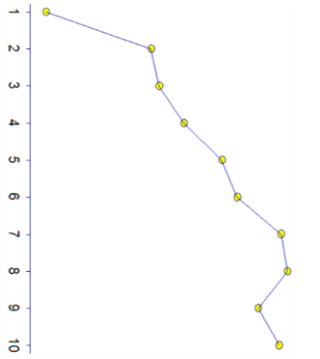
The relative contributions of free social services are meagre (*percentage of total consumption*, table 4), and will not make as significant an impact as initially anticipated in the design of the dependency index. With the increase in the price of utilities, and the increasing access provided by the state this could become an important measure in the long run however.

4.2.5 *Validity of earned income assumption*

The largest assumption in the creation of this model is that the balance of consumption **(E)** less all cash social grants **(CSi)** is equal to earned income. This assumption needs some statistical consideration if the internal validity of the model is to be scrutinized. Below in table 5 the GLM ANOVA of household wages plus **CSi** to consumption is presented.

Table 5: (Wages+ CSi) / Decile

Income Decile	(Wages+Csi) as % of Consumption	Count	Mean	Standard Error
1	27%	2103	0.27***	.02
2	72%	2461	0.72***	.20
3	75%	2500	0.75***	.20
4	86%	2309	0.86***	.20
5	102%	2272	1.02***	.21
6	109%	2151	1.09***	.21
7	128%	1981	1.28***	.22
8	130%	1952	1.30***	.22
9	118%	1848	1.18***	.23
10	127%	1566	1.27***	.25



****P<0.01*

As can be seen from table 5 above the mean percentage of Wages plus CSi (cash social grants) to consumption is .97. Further-more as can be seen from table 1 households in income deciles 1-4 consume significantly more than they earn through total income. Considering that income deciles 1 and 2 essentially represent the poverty line, it is understandable that significant "income" could be earned through informal work, begging or subsistence living that is not captured as income to the household. It is argued in this study that these forms of income do in fact constitute work on behalf of the household, and as there is no tax burden until income decile 7, considering earnings in this way does not render the dependency index redundant. From decile 5 social grants and wages constitute more than 100% of consumption, and as already assumed almost all social grants (CSi) are consumed and not saved, the balance is earned income. There are additional forms of income such as dividends and investments that are significant at upper income deciles, but these form largely the savings component whilst wages are used for consumption as will be elaborated on in the discussion.

4.2.6 *Summary problem statement*

This section has formed the underpinnings of the first problem statement, namely to create the dependency index itself. The means of the relevant components have been presented above to show their impact on the final index, and for context. These elements are then utilised as per the Dependency Index equation (pg. 14) to create the final index, which is then used in further hypotheses testing as discussed in Sections 6 to 7 and presented in the following ANOVA-based sections.

4.3 Results showing differences in the index per income decile (Hypothesis 1)

H1: The mean Dependency Index is statistically different per income decile.

As discussed in the methodology, since the dependency index is a continuous variable and this hypothesis seeks to ascertain whether it differs across deciles, ANOVA-type analyses are appropriate.

4.3.1 *Index manipulations*

Due to the nature of the equation, there are several observations that are significant outliers needing to be winsorized, statistically adjusted for or removed. Three methods of index adjustment will be presented here for contrast.

i. **Index_raw**

The first method used was to take the index in its raw form with all the outliers, and run them though a non-parametric Kruskal-Wallis one-way ANOVA. This controls for outliers by allocating the

values median ranks and not means, thus removing all significant outliers. These medians are then treated as means when comparing the deciles.

ii. **Index_winsorized (if 0 >index>5 then index = 5)**

The top half of the equation (pg.14) represents household utility, and the bottom half represents the money required through earnings to afford such utility. On a number of occasions households received several thousand rand more in social cash transfers than the total consumption of that household. This results in index outputs that are negative and some time in excess of 100 units, compared to an index average of 1.56 these outliers skew the data greatly. Similarly if households receive more than 83% of their income from cash social transfers the index rises above 5 and starts to distort the results. Thus a windsorised version of the index takes any negative index value and gives it a value of 5 (assumes maximum CSi to E ratio of 5:1), and similarly any index value greater than 5 is winsorized to 5. There were 1,681 index values below zero (more consumption that social grants received), and 2,547 index values above 5 (over 83% of consumption can be from social grants), thus 19% of the households were affected by adjustments to outliers. This approach allows for use of a parametric GLM ANOVA procedure.

iii. **Index_Less_outliers (if 0 >Index>5 then observation removed)**

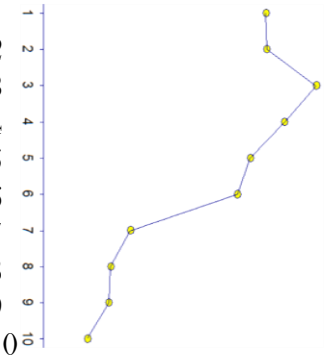
The final approach toward the outliers was to remove all 4,228 of the outliers discussed from the sample. As the observations with the greatest values are removed, it is expected that this index iteration will represent the lower limit of the index. This was possible as the sample size is large (n=21,144) enough to retain statistically significant results even after the removal of the outliers. However the outliers were spread mainly between decile 2-6 which could cause some population specific error. This approach also allows for use of a parametric GLM ANOVA approach.

The results for the three index iterations discussed above will now be presented.

4.3.1 *Non-parametric Kruskal-Wallis analysis of the raw index*

A one way Kruskal Wallis ANOVA of the raw index values was run in NCSS for the 21,144 observations, the results of which are shown in table 6, 7 and 8 below. Table 9 shows the fit of the ANOVA at a 99% level of confidence. Table 6 shows the basic differences between the deciles on the median reversed rank of the dependency index. As can immediately be seen, the index peaks at income decile 3 and falls steadily to income decile 6 after which the index drops sharply towards income decile 10.

Table 6: Raw Index / Decile : Kruskal-Wallis non-parametric ranked medians

Income Decile		Count	Median
1		2103	1.01***
2		2461	1.01***
3		2500	1.10***
4		2309	1.04***
5		2272	0.98***
6		2151	0.95***
7		1981	0.76***
8		1952	0.72***
9		1848	0.72***
10		1566	0.68***

***P<0.01*

Table 7 below shows the overall Kruskal Wallis ANOVA fit: as seen there, the differences between deciles are significant at the 99% level of confidence.

Table 7:The Kruskal Wallis one-way ANOVA statistical fit

Method	DF	Chi-Square (H)
Not Corrected for Ties	9	6250.16***
Corrected for Ties	9	6250.85***

****p < .01. Number sets of ties = 140*

Table 8 then shows which groups are statistically similar to one another through the use of grouping brackets. To identify differences between deciles I performed a z-value Bonferroni multiple comparison test (>3.26). The results are depicted graphically in table 8 below.

Table 8: Similar groups identified by the Kruskal Wallis ANOVA (Dunn's test)

Decile	Median rank	Z-score
Decile 3	1.10	25.69
Decile 4	1.04	21.95
Decile 1	1.01	21.62
Decile 2	1.01	19.26
Decile 5	.98	11.69
Decile 6	.95	11.25
Decile 7	.76	-13.25
Decile 8	.72	-25.59
Decile 9	.72	-39.27
Decile 10	.68	-47.55

(P<0.01, Bonferroni test: Medians significantly different if z-value > 3.26)

As can be seen from table 8 above, income deciles 7 through 10 are significantly different from all other groups, whilst deciles 5 and 6 are similar to one another, and deciles 4 and 1 are similar to one another flanked by similarities to deciles 3 and 2. The Bonferroni test was used as it was the most stringent of the multiple comparison z-value tests available on NCSS.

These results provide support for hypothesis 1, since it shows differences between deciles on the dependency index.

4.3.2 *GLM ANOVA of the winsorised index*

A GLM ANOVA of the winsorized index values was run in NCSS for the 21,144 observations, the results of which are shown in table 9, 10 and 11 below. As can be seen from table 9 below, the means of the index per income decile first increase to a maximum at decile 3, dipping and rising to decile 5, then dropping steeply toward decile 10.

Table 9: Winsorized Index/ Decile : Summary of Means

Income Decile	Count	Mean
1	2103	1.22***
2	2461	2.09***
3	2500	2.25***
4	2309	2.11***
5	2272	2.25***
6	2151	1.92***
7	1981	1.27***
8	1952	.95***
9	1848	.82***
10	1566	.73***

***P<0.01*

Table 10 below shows the fit of the ANOVA at a 99% level of confidence, including the power score and other relevant metrics

Table 10: Winsorized Index/ Decile: GLM ANOVA regression

	Sum of Squares	*df*	Mean Square	F	Power
Regression	7069.33	9	785.48	483.86***	1.00
Residual	34306.6	21133	1.62		
Total	41376.1	21142			

¨P<0.01

The TukeyKramer multiple comparison test in table 11 below has shown some groups to have statistically similar index values, but largely the deciles are statistically different from one another.

Table 11: Similar groups identified in the winsorized index (Tukey Kramer pairwise comparison)

Decile	Mean	Standard error	
Decile 3	2.25	.026	]
Decile 5	2.25	.027	]
Decile 4	2.11	.027	]
Decile 2	2.09	.026	]
Decile 6	1.92	.028	]
Decile 7	1.27	.029	]
Decile 1	1.22	.028	]
Decile 8	.95	.029	]]
Decile 9	.82	.030	]
Decile 10	.73	.032	]

Alpha 0.05, Critical Value = 4.47, DF = 21142, MSE = 1.62

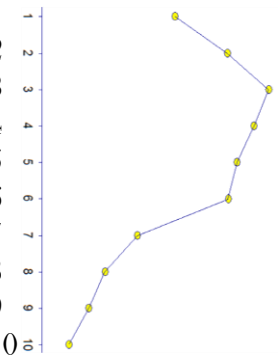
As can be seen from table 11 above decile 1 and 7 have a similar index score, as do decile 3 and 5, 4 and 2 and 9 and 10. Decile 6 is statistically different from all other groups.

These results provide support for hypothesis 1, since it shows differences between income deciles on the dependency index.

4.3.3 *GLM ANOVA of the Index less outliers:*

A GLM ANOVA was run in NCSS for the 18,595 remaining observations, the results of which are shown in table 12, 13 and 14 below. As can be seen from table 12 below, the means of the index per income decile first increase to decile 3 and then tapers off to decile 6, before dropping quickly to a minimum at decile 10.

Table 12: Index Less Outliers / Decile : Summary of means

Income Decile		Count	Mean
1		2065	1.14***
2		1967	1.36***
3		1979	1.53***
4		1887	1.47***
5		1735	1.40***
6		1822	1.36***
7		1840	.99***
8		1910	.86***
9		1833	.79***
10		1558	.71***

***P<0.01*

Table 13 shows the fit of the ANOVA at a 99% confidence level, as well as the power and other relevant metrics.

***Table 13:** Index less outliers/ Decile: GLM ANOVA regression*

	Sum of Squares	df	Mean Square	F	Power
Regression	1492.27	9	165.80	440.01***	1.00
Residual	7003.77	18586	.3768		
Total	8496.03	18595			

***P<0.01*

The TukeyKramer multiple comparison test shown in table 14 below has shown some groups to have statistically similar index values, but by in large the deciles are statistically different from one another.

Table 14: Similar groups identified in the index less outliers (Tukey Kramer multiple comparison)

Decile	Mean	Standard error	
Decile 3	1.53	.0138	A
Decile 4	1.47	.0143	A
Decile 5	1.40	.0147	A B
Decile 2	1.36	.0138	B
Decile 6	1.36	.0144	B
Decile 1	1.14	.0135	C
Decile 7	.99	.0141	D
Decile 8	.86	.0143	E
Decile 9	.79	.0143	F
Decile 10	.71	.0156	G

Alpha 0.05, Critical Value = 4.47, DF = 18,533, MSE = .38

As can be seen from table 14 above decile 2, 5 and 6 have a similar index score, as do decile 3 and 4. Decile 1 and 7 through 10 are significantly different from all other groups.

These results provide support for hypothesis 1, since it shows differences between deciles on the dependency index.

4.3.4 *Visual comparison of the factors*

Although the trends of the three index iterations appear to be similar, they differ by a large degree in terms of magnitude, as can be seen in figure 7 below.

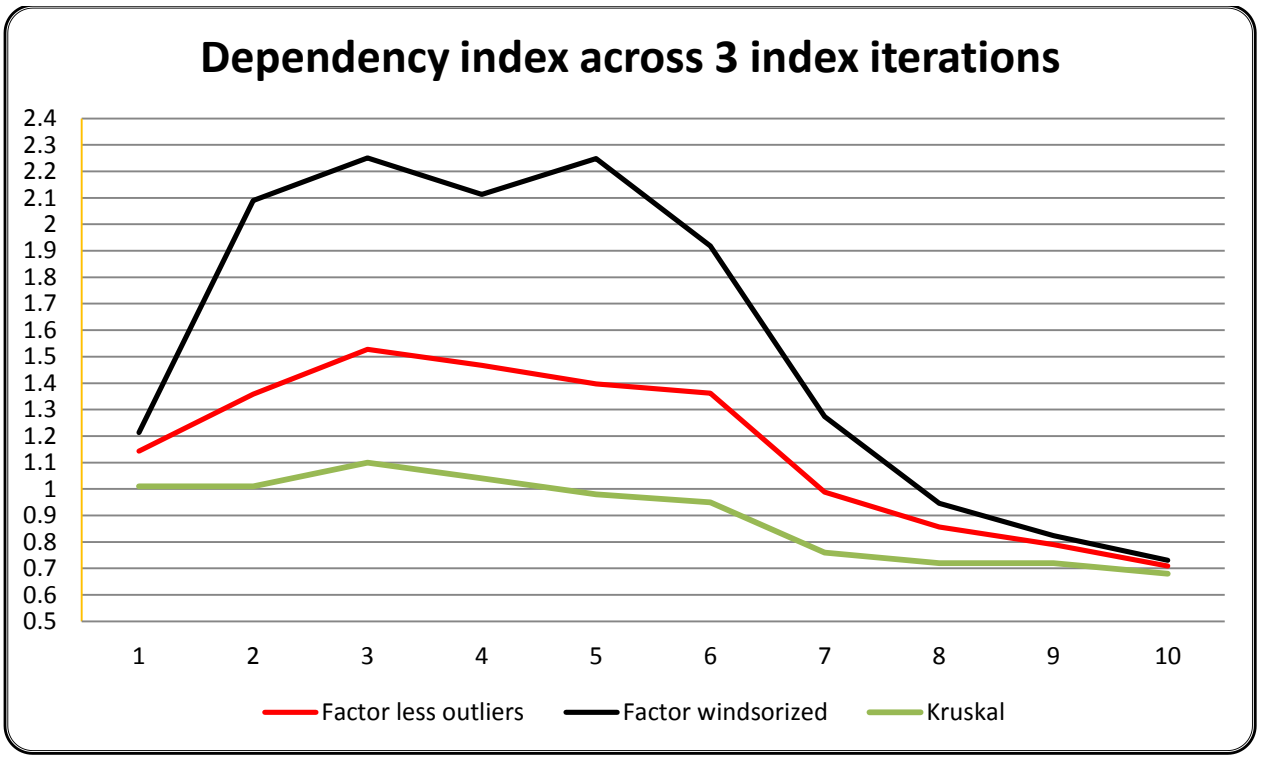


Figure 7: Means of the three index iterations per income decile

The index shown by the Non-parametric analysis is in medians, and not directly comparable, thus it is shown for trend and not magnitude. The index less outliers has a maximum mean of 1.53 in decile 3, whilst the winsorized index has a maximum mean of 2.25 in decile 3 (32% increase). This is not unexpected as the large outliers have been removed completely, and therefore their influence in non-existent. Whilst the winsorized index included all the outliers, of which there are almost 20%, but limited them to a maximum utility to earned income ratio of 5:1. In both cases decile 10 has the minimum index value at .71 and .73 respectively. Figure 7 shows the magnitude of the iterations, but it is also of interest to consider the absolute trend of the three iterations through a standardised scale as seen in figure 8 below.

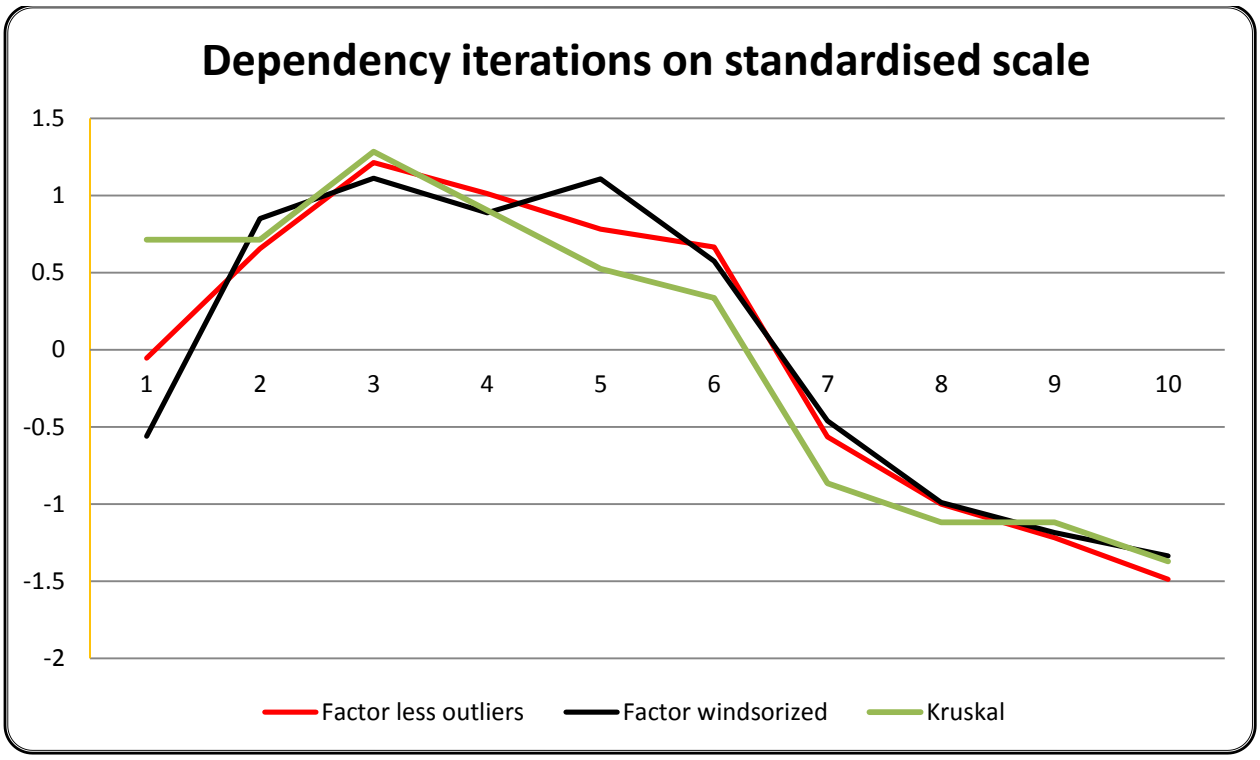


Figure 8: Standardised dependency index iterations for trend analysis

As can be seen the factor less outlier iteration and the windsorized iteration have a remarkably similar trend. Both are represented by a minimum at income decile 1 and 10 as expected, whereas income decile 1 in the one-way ANOVA is far greater than zero in decile 1 and is slightly off the observed trend. However the index iterations follow the same basic trend as expected, which adds support to hypothesis 1 and 2.

4.3.5 *Summary of hypothesis 1*

Hypothesis 1 was that income deciles would have a significantly different dependency index value. This is not strictly the case, as some deciles have similar index values, but by in large the values are different per decile and the hypothesis has support, or at least the null hypothesis rejected. Similar deciles appear to be on either side of the decile range 3-5, when the index drops toward their respective minima at decile 1 and 10. Thus the constitution of the index, as expected, results in significantly different index scores across income deciles in South Africa, with similar index values providing points of discussion.

4.4 Results pertaining to Hypothesis 2

H2: The mean dependency index per income decile follows an inverse curvilinear trend increasing from income decile 1 to 5/6 and then decreasing to decile 10.

To test this hypothesis, a curve fitting programme was applied via NCSS, which tests the relative fit of multiple different curves.

4.4.1 *Curve fitting of the winsorized Index*

When fitting curves to the winsorised index, several non-linear iterations were attempted. In fact, the PolyRatio(3,3) had the best fit and is thus the only results presented.

- Best fit curve: Y = PolyRatio(3,3)
- Index=(A+B*(Decile)+C*(Decile)^2+D*(Decile)^3)/(1+E*(Decile)+F*(Decile)^2+G*(Decile)^3)
- Bootstrapped R^2= .98

The equation parameters have been defined below in table 15.

Table 15: Equation parameters for PolyRatio(3,3) winsorized index/ income decile

Parameter Name	Bootstrapped Mean	Standard Error	Lower 95% C.I	Upper 95% C.L
A	-102.88**	1.32	-102.04	-98.17
B	210.57**	2.74	200.83	208.62
C	-42.23**	.57	-41.83	-40.25
D	2.37**	.03	2.26	2.35
E	67.95**	.94	64.68	67.34
F	-15.35**	.23	-15.23	-14.57
G	1.05**	.02	.99	1.04

**P<0.05

All the factors A through G are significant at 95% confidence, and the fit of the curve to the mean index per income decile is very good. Thus the equation can be used to define the index as long as the income decile is known (Appendix A). The graph of the curve from NCSS is shown below in figure 9.

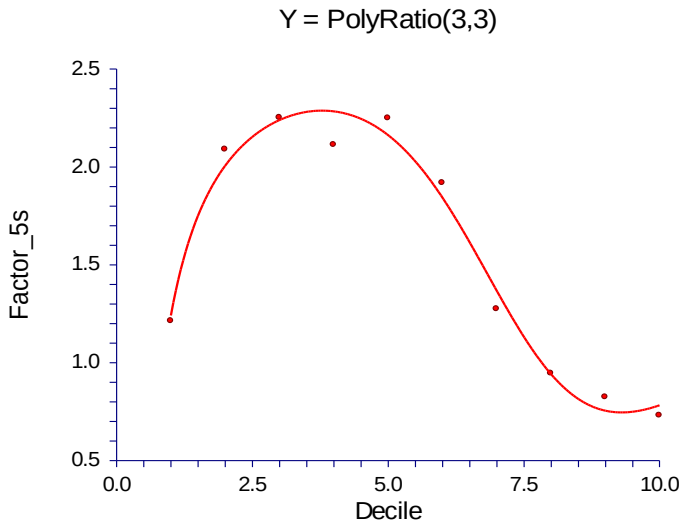


Figure 9: PolyRatio(3,3) graph fitted to winsorized index per decile

The equation for the fitted line is given in equation 1 below:

Winsorized Index equation=(-102.88+210.57*(Decile)-42.23*(Decile)^2
+2.37*(Decile)^3)/ (1+67.95*(Decile)+ 15.35*(Decile)^2 +1.05*(Decile)^3) Equation 1.

4.4.2 *Curve fitting of Index less outliers*

Several non-linear iterations were attempted, but again the PolyRatio(3,3) had the best fit and thus are the only details demonstrated.

- Best fit curve: Y = PolyRatio(3,3)
- Factor=(A+B*(Decile)+C*(Decile)^2+D*(Decile)^3)/(1+E*(Decile)+F*(Decile)^2+G*(Decile)^3)
- Bootstrapped R^2= .98

The equation parameters have been defined below in table 12.

Table 16: Equation parameters for PolyRatio(3,3) Index less outliers / income decile

Parameter Name	Bootstrapped Mean	Standard Error	Lower 95% C.L.	Upper 95% C.L
A	.29**	.001	.29	.29
B	2.34**	.004	2.34	2.34
C	-.51**	.001	-.51	-.51
D	.031**	.00006	.032	.031
E	1.16**	.171	.72	1.60
F	-.30**	.001	-.30	-.30
G	.023**	.00005	.023	.023

**$P<0.05$

All the factors A through G are significant at 95% confidence, and the fit of the curve to the mean factors per income decile is very good. Thus the equation can be used to define the utility factor of income as long as the income decile is known (Appendix A). The graph of the curve from NCSS is shown below in figure 10.

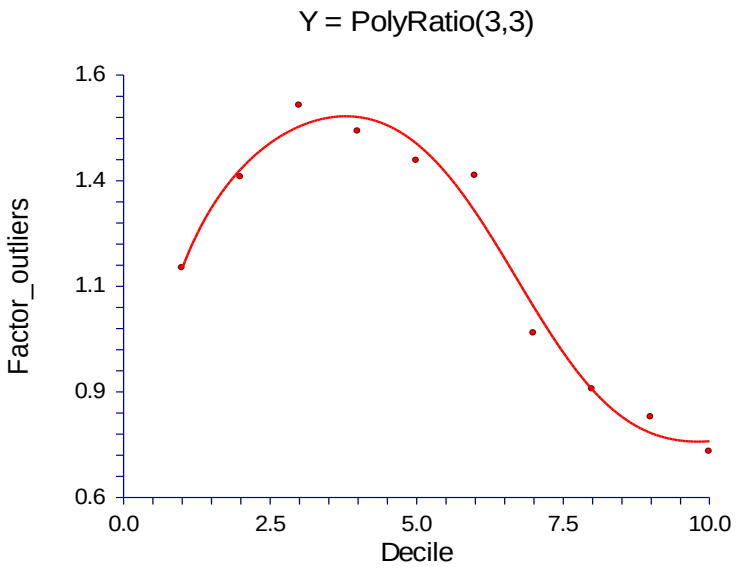


Figure 10: PolyRatio(3,3) graph fitted to per decile mean index less outliers

The equation for the fitted line is given in equation 2 below:

Index equation less outliers = (.29+2.34*(Decile)-.5*(Decile)^2
+.03*(Decile)^3**)/(**1+1.16*(Decile)+-.3*(Decile)^2+.023*(Decile)^3**)** Equation 2

4.4.3 *Summary of per-decile curve fitting*

The PolyRatio(3,3) curve fits both data iterations remarkably well with R^2 of .98 for both, and all bootstrapped factors being significant at the 95% confidence level. A quadratic or PolyRatio (2,2) curve was expected to fit the data better, as there is a minimum at decile 10 in all cases. However, the data fits the PolyRatio (3,3) by an R^2 differential of at least .05 in both cases. Therefore the PolyRatio(3,3) curve was accepted for as accurate results as possible, but the outcomes from hypothesis 3 must be considered when analysing decile 10 to avoid the overestimation of the factor in this decile. Thus the question is more about which factor iteration is more appropriate, as both of them are well represented by the equations calculated above.

Having said this, the cubic equation implied by the PolyRatio (3,3) basically fulfils the requirements for support of Hypothesis 2, since the bulk of the equation forms an inverse quadratic curve with only a small flattening out towards Decile 10 suggesting the elaboration. Hypothesis 1 had expected a peak at Deciles 5/6, in this case the literal data peak is at Decile 3 and the fitted curve peak is at Decile 4, slightly earlier than expected. I discuss these results in the next chapter. It is noteworthy that, as expected, the dependency index starts falling drastically after Decile 6.

4.5 Results pertaining to Hypothesis 3

H3: It is likely that the trend of the mean dependency index within Decile 10 alone will be downward sloping ending at a minimum in the 100th percentile.

Any household earning more than R180,212 is classified as belonging to income decile 10. Therefore assuming there is only one contributor to household income the top three tax brackets (Annexure A) are represented by the top decile. It is likely that there are other contributors to household income however, and thus income decile 10 should have quite a large internal trend owing to the differential tax rate.

Households in income decile ten were divided into ten equal groups based on total income earned. These ten groups then represent the top 10 percentile of the national sample, the degree of significance of this GLM ANOVA is presented below in table 17.

Table 17: Top 10 Percentile of index less outliers: GLM ANOVA regression

	Sum of Squares	df	Mean Square	F	Power
Regression	1.90	9	.21	8.50¨	1.00
Residual	38.44	1548	.025		
Total	40.34	1557			

¨p < 0.05

There are still 1,557 observations in the top ten percentile, with 156 observations per percentile, so the fit remains good for all but one of the factors. The shape and means of the index are shown in table 18 below.

Table 18: Index Less Outliers / Top 10 Percentile: Summary of means

Top 10 percentiles	Count	Mean
1	156	.74**
2	156	.74**
3	156	.73**
4	156	.73**
5	156	.75**
6	156	.71**
7	156	.70**
8	156	.68**
9	156	.67**
10	154	.64**

¨*P<0.05*

For the purpose of the top ten percentile analysis the 'iteration less outliers' was chosen, although in this decile there were very few outliers. Interestingly there is a spike at the 95th percentile, where some large social and disability grants were noticed, and more duel income earning households were located (lower effective tax rate). But other than the 95th percentile, the means trend downwards towards the 100th percentile. The trend is as expected and is driven by the progressive income tax system. Thus for the index to have meaning for more affluent individuals, a curve needed to be fitted to the index means of the top ten percentile. This is also very relevant because the cubic graphs fitted for the full decile range (PolyRatio (3,3)) trends slightly upward in decile 10, whereas the true trend as can be seen in figure 11 is downwards to a minimum at decile 10.

- Best fit curve: Y = Polyratio(3,3)
- Factor=(A+B*(Decile)+C*(Decile)^2+D*(Decile)^3)/(1+E*(Decile)+F*(Decile)^2+G*(Decile)^3)
- Bootstrapped R^2= .94

The equation parameters have been defined below in table 19.

Table 19: Equation parameters for PolyRatio(3,3) Index less_outliers / top 10 percentile

Parameter Name	Bootstrapped Mean	Standard Error	Lower 95% C.L.	Upper 95% C.L.
A	.76**	.0005	.76	.76
B	-.29**	.0007	-.29	-.28
C	.03**	.0007	.03	.033
D	-.0003*	.00007	-.0003	.00004
E	-.36**	.001	-.36	-.35
F	.036**	.001	.03	.04
G	.0004**	.0001	.0003	.0008

**P<0.05, *P<0.1

All but parameter D are significant at the 95% confidence level (D significant at the 90% level), and the R^2 is high at .94. Therefore the curve can be used as an accurate measure of the utility factor per percentile, despite D not fitting the curve with 95% significance. The curve from NCSS has been added below as figure 11.

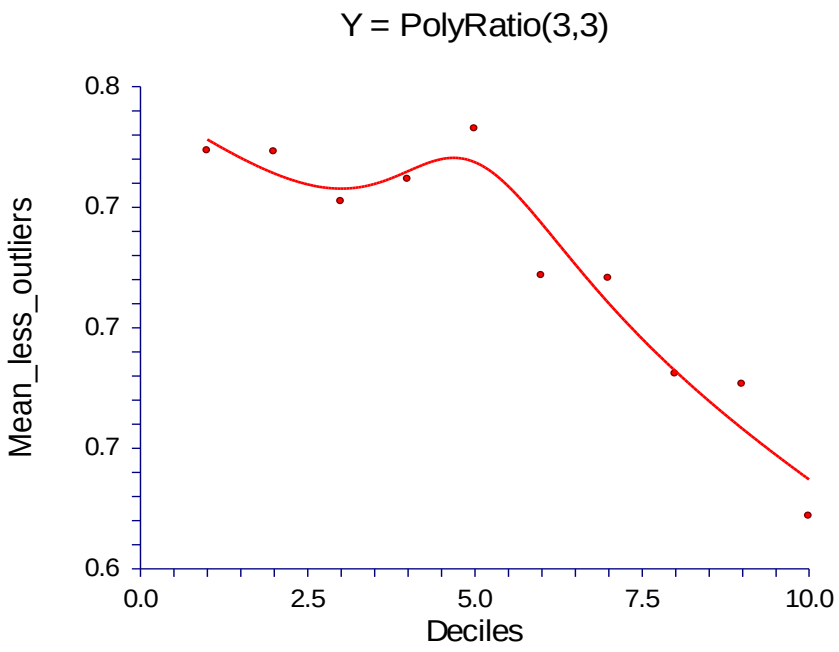


Figure 11: PolyRatio(3,3) graph fitted to the mean index of the top ten percentile less outliers

The equation for the fitted line is given in equation 3 below:

$$\text{Index of top ten percentile} = \frac{(.76 + .29*(\text{Percentile}) + .03*(\text{Percentile})^2 + .00027*(\text{Percentile})^3)}{(1 + -.36*(\text{Percentile}) + .04*(\text{Percentile})^2 + .00035*(\text{Percentile})^3)} \qquad \text{Equation 3}$$

4.5.1 *Summary of hypothesis 3*

It was hypothesized that there would be a strong downward slope in the top ten percentile within RSA. Interestingly decile 4 and 5 seem to break this trend and create the peak for the PolyRatio (3,3) graph to fit the data. Thus there is a strong downward trend as expected, but there is a bulge in earned income utility in decile 4 and 5 where some high disability and grant in aid figures were noted as well as on average lower income tax. This results in the overestimation of the factor for the 100th percentile. For the analysis of the top ten percentile outliers were completely removed, although there were not many outliers in this decile. The curve fits the graph well and is significant at the 95% level in all cases but for parameter D which fits at the 90% confidence level.

I provide discussion on and interpretation of these results in the next chapter, suffice to say that these results serve to support hypothesis 3 as the trend observed is largely downward to a minimum at the 100th percentile.

4.6 Summary of the results

The results have shown the index is able to take into account various influences of household utility versus earned income and translate them into a single measure per income decile. This allowed us to compare the index across income deciles for statistical analysis. Despite the fact that two iterations of the index differing in magnitude have been used, their trend is consistent. The trend is also consistent with the non-parametric analysis which reinforces the observed trend.

The index shows how households gain income utility as they gain access to social services peaking in decile 3-5, and how income and consumption tax then begin to erode utility toward decile 10. The effect of subsidised social services is surprisingly low in its cash equivalent units and constituted a maximum of 3.6% of consumption in decile 1, resulting in this metric not affecting the index trend significantly.

The results are highly significant and powerful due to the size of the dataset and appropriate iterations, and therefore can be used to make useful real world comparisons for discussion purposes. The combination of the three equations for the index and the top ten percentile breakdown allow a full comparison of each decile's interaction with the economy and the state.

Moreover the index has come to represent the dependency of households on the state when above 1, and the dependency of the state on households when below 1.

5 DISCUSIION OF THE RESULTS

5.1 Introduction

This chapter mirrors the structure of the results section. Implications of the results are discussed in context of the body of this report. Conclusions of the discussion will then be consolidated and presented in summary of the research.

5.2 The dependency index

As discussed in the beginning of the research report, the main problem of the research is *to construct and analyse a model (the dependency index) that predicts average household utility relative to earned income by income decile in South Africa.*

The main research problem was to define and create the model for the dependency index. This has been achieved in this study, albeit through different factor iterations. The largest element for discussion around the model, before the discussion of individual hypotheses, is internal validity.

There are several critical assumptions that were made in order to create this model. The largest assumption is that the balance of consumption (E) less all cash social grants (CSi) is equal to earned income. Table 2 and table 6 in the results section suggest that this assumption is approximately fair, as wages plus social grants (CSi) are for the most part equal to or greater than consumption (mean = .97, table 6). When this is not the case in deciles 1-4 there is a discrepancy where total consumption can be 230% more than total income (table 2), which accounts for much of the gap of earned income, and the balance can possibly be explained by informal work for consumption in the household. Thus because wages plus social grants make up essentially 100% of consumption, it is fair to assume that these two elements can be defined as the only inputs for this purpose.

As discussed there are other forms of income (investments, dividends, gifts, alimony etc.) to households, but these have not been considered, the reason for this is multi-faceted. For example, in income decile 10, wages plus social grants is 27% larger than total household consumption (table 6), whilst total income is 42% greater than consumption (table 2). Thus 15% of income is earned through other sources in this case. In income decile 6 however, wages plus social grants is 9% greater than consumption, and total income is 9% greater than consumption. Thus in this case 100% of income is earned through wages and social grants. But the factor does not take into account savings, and the only reason to delimit the balance of consumption less social grants to determine earned income, is for the calculation of the magnitude of that earned income pre-income tax. It would be difficult to explain the source of income for consumption for households in income decile 10 for example, if 15% of consumption could be paid for by dividends taxed at 12%, social grants could be saved, and the balance of consumption could be paid for in wages taxed at 40%. As can be seen, without complicating the dependency index immensely it is difficult to factor in all possible mechanisms of consumption and their relative sources of income, suffice to say that wages plus CSi is a good proxy for income for consumption purposes. This, in conjunction with the assumption that savings from social grants are negligible, leads to the inevitable conclusion that the balance of consumption less CSi is post income tax earned income for consumption purposes. Once these assumptions are accepted the model holds significant internal validity and the discussion of the hypotheses below is of great interest.

The creation of the final data set for statistical analysis involved numerous steps (figure 6), that involved the integration of data into discrete household values from 2 different databases. These datasets were a cleaned version of the Statistics SA IES 2005/2006 by DataFirst. Despite the cleaning of the data there are inherent inaccuracies in the representativeness of the dependency factor, as there

are acknowledged errors and biases in the dataset from Statistics SA. Therefore the integrity of the data in many ways will determine the internal validity of this study, and an opportunity exists to repeat this study with the new IES information to be published in 2012 for comparison.

5.3 Discussion pertaining to Hypothesis 1

H1: The mean Dependency Index is statistically different per income decile.

Hypothesis 1 postulated the existence of three possible factor iterations in order to deal with the problem of outliers, where more than 83% of household consumption was derived from social grants. These outliers represent nearly 20% of the total sample, and thus have a large effect on the trend of the dependency index across income deciles. It is possible that there are households that receive more social assistance than they consume (1,681 according to the data), which in turn results in this form of social dependency becoming a mechanism of wealth generation (households saving social grants). The dataset before the outliers have been removed or adjusted could be of interest to policy makers for a view of where social grants are becoming a mechanism of wealth generation. It is likely that such households will not see the need to escape dependency on the state as structural poverty traps can skew incentives of the incumbents (Bourguignon 2004). Although outliers are largely ameliorated for in this study their examination could assist in more effective allocation of scarce state resources to needy groups, as opposed to concentration on a small number of income deciles to the exclusion of others.

As presented in the results there were three dependency factor iterations explored, which will be discussed briefly below and then contrasted for discussion purposes.

5.3.1 *Non-parametric Kruskal-Wallis analysis of raw index*

The reason for doing the non-parametric analysis was to create an index iteration without assumption bias. This is useful in the discussion and interpretation of the other two iterations, but it cannot form the dependency index itself, as it does not represent the index means with magnitudes, rather the median of the ranks of the index.

5.3.2 *GLM ANOVA of the winsorized index*

A winsorized ratio of 5:1 was chosen so that the outliers would have some influence on the index, but the trend of the majority could still maintain its integrity. This iteration is considered to represent the upper limit of the dependency index in this study.

5.3.3 *GLM ANOVA of index with removed outliers*

As the upper limit of the index has been defined in the winsorized iteration, this iteration will represent the lower limit of the index. This is the case because any observation that has in excess of 83% of is consumption coming from social grants (index score exceeds 5) is removed from the dataset altogether. Thus this iteration of the index is most likely the closest representation of the average household in South Africa, as it contains 81% of the data that is the most closely spread.

5.3.4 *Discussion on hypothesis 1*

The Tukey-Kramer multiple comparisons presented in the results demonstrate that although the household index scores are distinct in their constitution, they are often statistically the same. In order to illustrate this more effectively figure 7 presented in the results section has been adapted to figure 12 below, showing statistically ($p<0.1$) similar groups with similar shape and colours icons.

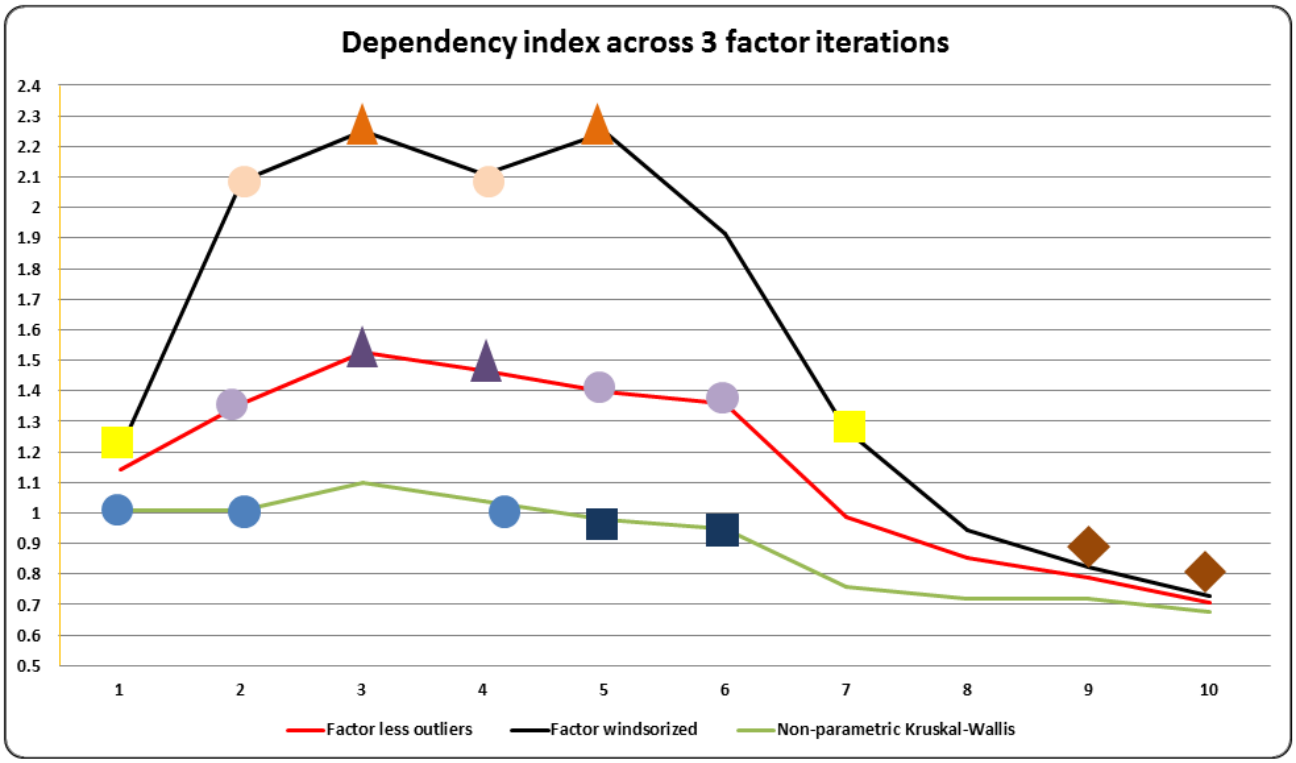


Figure 12: Dependency index across three different index iterations, showing similar groups

This is an effective visual illustration of similar groups. Each index iteration has been plotted against income deciles, with similar groups in each line graph being represented with the same shape and colour icon. If the mean is different from all groups (or inappropriate for grouping) there is no coloured icon present. As can be seen in figure 12, each index iteration shows statistical differences between various income deciles, which disproves the null hypothesis that no differences exist.

Therefore, hypothesis 1 (H1) that statistical differences exist between income deciles can be confirmed in all three index iterations.

There are of course several deciles which have statistically similar index values, despite the different constitution of the elements in their index calculations, their net purchasing power of earned income is the same. For example, income deciles 1 and 7 are statistically similar to one another ($p<.05$) although they fall on either side of the 'social grant bubble', and have fundamental differences in the constitution of their income. Income decile 1 is so impoverished and isolated that most households cannot even access social assistance. These households earn very little in the way of formal wage, and their consumption is mainly through subsistence agriculture, bartering, begging and odd jobs (Muller 2003, Thomas 1994). Thus as these households receive very little in the way of social benefit, virtually all of their consumption is driven through non-taxable earned 'income' (in this case subsistence agriculture is considered as income in its rand value of consumption, as there is no tax paid at this level).

The mean index value in decile 1(winsorized) is 1.22, meaning that for every R1 of income earned this household can spend R1.22. Households in this decile are thus a net burden to the state to the extent of 22c for every rand that these households spend of earned income. Similarly income decile 7 has a mean index value (winsorized) of 1.27, meaning that for every R1 that a household in income decile 7 spends of earned income the state sponsors that household and extra 27c. But income decile 7 pays income tax and more consumption tax. How can they be the same burden to the state as income decile 1? The answer lies in the fact that each household in income decile 1 receives on average 13.8% of the social grants (table 4) that the average household in income decile 7 receives, despite households in income decile 1 residing in dire poverty. So, despite the fundamental difference between households in income deciles 1 and 7, they have the same purchasing power of each rand of earned income spent.

There are other examples of similar dependency index values between income deciles such as income deciles 2, 5 and 6 in the iteration less outliers, and factors 1, 2 and 4 in the one-way ANOVA Kruskal-Wallis analysis. There are several other similar groups as well, but on the whole the mean dependency index iterations all show significant differences between one another, confirming hypothesis 1.

5.3.5 Why the indices differ in magnitude

The critical comparison in magnitude is between the index with removed outliers and the winsorized index, as these would be used directly as the dependency index. It is important to note that there is a large difference in the magnitudes between decile 2-6 between the two models. This aligns with the increased number of outliers observed in these deciles. These deciles represent the bulk of

social grant assistance as a percentage of consumption (table 4), and point to a very important consideration. The index is sensitive to social grants as a percentage of consumption, and of the 1,681 observations in the sample who earned more income in social grants than they consumed, the majority of these existed in income deciles 2 and 5 and to a lesser degree 3. From figure 12 it can be seen that the greatest range between the indices exists between income decile 2 and 5 and to a lesser extent 3 and 4. Therefore the areas in figure 11 with the greatest difference are likely to have the greatest number of households who consume less that the value of their social grants. This is of concern when groups in income decile 1 are firmly below the poverty line (Oosthuizen 2007), and as shown by the study receive very little social assistance even as a percentage of earned income. The graphs converge from income decile 6 in all cases, as social assistance reduces as a percentage of consumption and hence outliers are minimised.

5.3.6 *Conclusion of Hypothesis 1*

Hypothesis 1 (H1) is largely supported as the varying inputs to the dependency index per income decile have resulted in statistically different index values (in most cases) for all the index iterations. When the index values are similar, interesting comparisons can be drawn which strengthen the understanding of the relative socio economic position of population groups.

The relevance of the three different factor iterations has been elucidated through the analysis of hypothesis 1 which assists in defining real world applicability. One index does not have to be chosen over the other however, in fact two (potentially more) separate snap-shots of varying degrees of outlier influence can add granularity to the analysis. The index less-outliers can be used as a conservative estimate of the factor, and also the majority trend of 81% of the data-set. The winsorized index shows the effect of these outliers on the average index, and indicates groups who have more outliers than others. There could be a lot of value in running several winsorized iterations from a ratio of 2:1 to a ratio of 15:1. This will show the true magnitude of each of the outliers through the total increase in amplitude, and give a fiscal 'topography map' of the relationship between the state and its citizenry. The differing inputs to the dependency index equation at each level fundamentally change the mean index per income decile, but two different deciles can have statistically similar index scores with fundamentally different economic situations and inputs.

5.4 Discussion pertaining to Hypothesis 2

H2: The mean dependency index per income decile should follow an inverse curvilinear trend increasing from income decile 1 to 5/6 and then decreasing to decile 10.

It was expected that the shape of the curve would rise to around decile 4 as access to social grants increases, and then stabilize toward decile 6 as consumption tax and social grants begin to erode buying power, then the curve would drop steeply owing to income tax and decreasing social income to a minimum in decile 10. It was therefore assumed that one of the inverse curvilinear relationships in figure 12 below would have shown the best fit.

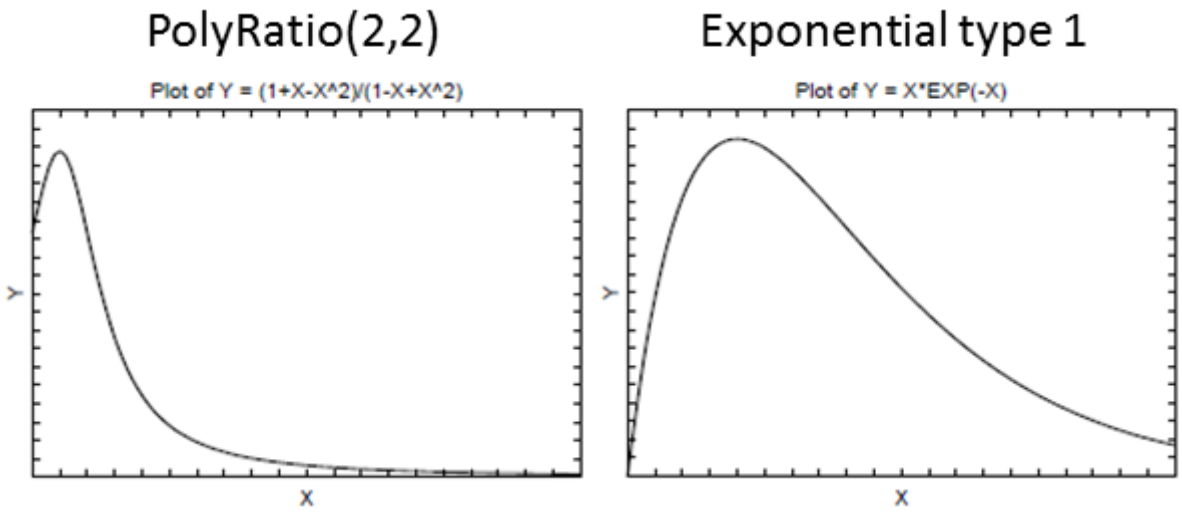


Figure 13: expected curvilinear relationships of the mean factor per decile

In reality the PolyRatio(3,3) curve fit the data set the best in all cases, as can be seen from the analysis of the results. This implies that the curve is beginning to rise past income decile 10. Of course there is nothing after income decile 10, as the categories are discrete? The likely reason for this fit is that there is a steep decrease in the index between income decile 6 and 8 after which the index levels out (figure 8 and 9). This causes the last increase predicted by the PolyRatio(3,3) curve to bisect income decile 9 and 10 in both index iterations, which gives the overall model good fit. The problem is that this decreases the resolution of the indices at income decile 9 and 10 (as can be seen by deviations in figure 8 and 9). Despite this both index iterations show incredibly good fit with the PolyRatio(3,3) curve, and as a result these models have been fitted for the estimation of a households buying power relative to earned income. To further resolve the issue of a potential over estimation of the index in income decile 10, this decile has been broken down into percentiles in the results of hypothesis 3. Thus the model with the PolyRatio(3,3) equation should be used to calculate the household index score, but if the household is in income decile 10, the percentile breakdown should be used for greater resolution and to offset the potential error introduced by the PolyRatio(3,3) curve fitting.

As outlined in the discussion on hypothesis 1, both index iterations are being considered in this study as they show different perspectives relative to outliers. Figure 13 below shows a decision tree diagram that can be used in order for a household to calculate its dependency index score.

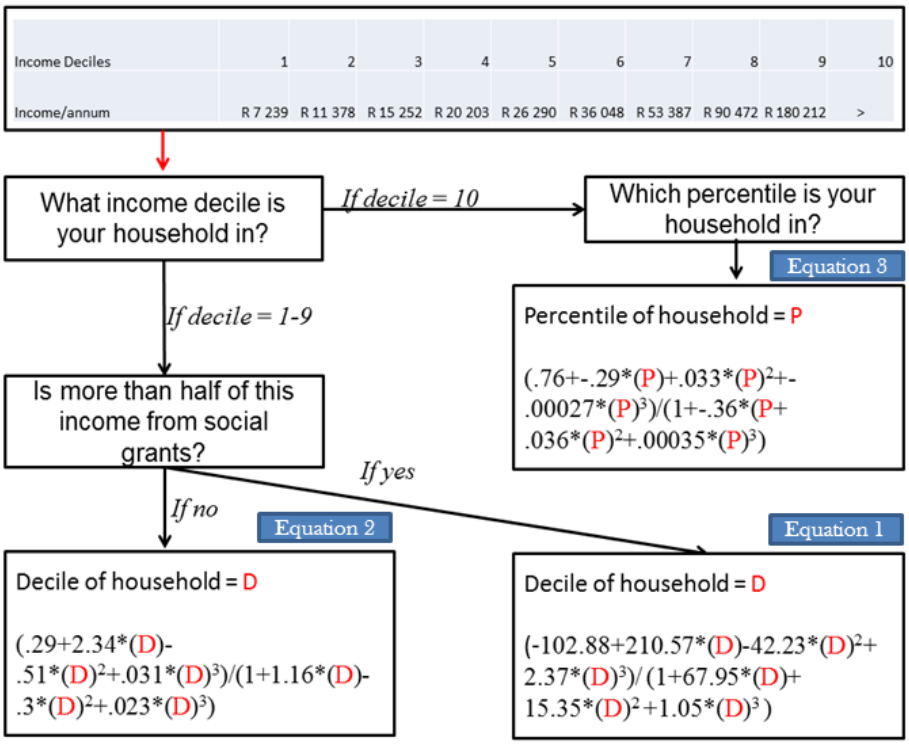


Figure 14: Decision tree for obtaining household dependency index

This decision tree can be illustrated through an example of a household that earns R23,000 per year before tax. It can be seen from figure 13 that this household is in income decile 4. Thus the first decision is directly downwards, which leads to the question about the magnitude of social grants. This household receives more than half of its income from disability and child care grants (which is above the average in all deciles and suggests an outlier). Thus the second decision is yes, which results in the use of ***Equation 1*** = $-102.88+210.57*(4)+42.23*(4)^2+ 2.37*(4)^3)/ (1+67.95*(4)+ 15.35*(4)^2 +1.05*(4)^3$

Therefore the dependency index according to the model = **2.28**

This indicates that for every rand of income earned and consumed by the household an additional R1.28 (128%) is provided by the state to that household for consumption. Thus through the equations and decision tree shown in figure 13, it is possible for households to calculate their spending power relative to income earned through wages.

5.4.1 *Conclusion of hypothesis 2*

It can be accepted that an inverse curvilinear line fits the index iterations with significance and accuracy. It was surprising that the best fit curve was from a PolyRatio(3,3) however, as in all cases income decile 10 represents a significant minimum. But in order to ameliorate any discrepancy in income decile 10 it was further separated into percentiles. Some of the results from hypothesis 3 have been considered here in order to create figure 13, which shows a decision tree on how a household can obtain its most accurate dependency factor.

Hypothesis 2 has resulted in the creation of a model that allows individual households to obtain and compare their dependency index, which can be useful in a couple of ways.

- A household could calculate their exact dependency index score, and compare that to the average index for their income decile. If their index calculation is lower than the average, it is likely that they could obtain extra relief from the state to subsidise consumption. However if their index calculation is higher, it is likely that they are receiving more from the state than the average household in that income decile.
- A household could do a cost benefit analysis of increasing their overall household income, and hence rising in income deciles. This could outline the real benefit derived in spending power versus the cost to obtain such additional pre-tax income. For example a new job could double wages resulting in a higher income tax burden, as well as making that household no longer eligible for certain social grants, whilst requiring more time dedicated to work, which could reduce the incentive to become 'wealthier'.

The example given of the household in income decile 4 above working out their dependency index and comparing it to the average is rather absurd however, as literacy and education in income decile 4 is very low (Van der Berg, Wood and Le Roux 2002, Statistics SA IES 2008). And in this context it is likely that the dependency index has greater use for policy makers and economists than the average household.

5.5 Discussion pertaining to Hypothesis 3

H3: It is likely that the trend of the mean dependency index in income decile 10 will be downward sloping ending at a minimum in the 100th percentile.

The top ten economic percentiles were created by dividing the 1,560 households in the top income decile by 10, resulting in 10 sorted groups of 156 households from the least to the most annual received income. This was necessary for a number of reasons. Firstly any household earning over R180,000 is in income decile 10, so much of the progressive tax burden is only experienced in this decile. Secondly as discussed in hypothesis 2, the PolyRatio(3,3) curve trends upwards in income decile 10, and could distort the results in this context. Also as CPI in South Africa is representative of the 95th economic percentile (Oosthuizen 2007), it is of interest to deconstruct this most economically productive group.

A strong downward trend in the mean index per percentile was expected, due mainly to the progressive tax system in RSA. This for the most part is true, aside from income deciles 4 and 5 which trend upwards. It was found that these income deciles had several households who received relatively large amounts of income through social grants, particularly for disability assistance. Also these households had lower than average income tax payments, as there were generally more contributors to household income, each with a lower effective tax rate. This pushed the mean index up in these percentiles and resulted in the best fit being the PolyRatio(3,3), and not the direct downward trend expected.

5.5.1 *Conclusion for hypothesis 3*

The best use of these results for households is outlined in figure 13 as a decision tree for the calculation of the dependency index in income decile 10. There is a strong downward trend as expected and hypothesised, but some households near the middle of the decile had higher social support and lower tax rates than expected. What is interesting to note is that the 100th percentile has the minimum factor of .63. This indicates that the wealthiest percentile have their spending power eroded by 37% which is collected through the state as income and consumption tax, whilst the 95th percentile with a factor of .75 only has their spending power eroded by 25%. It is likely that a more sophisticated measure of calculating real tax paid than the one employed in this study could affect these results, as this is the major contributing element in this decile.

5.6 Conclusion of discussion

The model for the dependency index has been defined and created in this research report. It has held up against internal scrutiny, and with the adjustments of the indices through the index iterations, it holds up to external scrutiny as well. The nature and purpose of each index iteration was discussed in detail, and it was concluded that both iterations are necessary in conjunction to draw a meaningful fiscal cross-section.

Through the GLM ANOVA procedures it was established that income deciles are largely statistically ($p<0.01$) different from one another, proving hypothesis 1. Income deciles with similar index values offer interesting comparisons to show how in some cases remarkably different economic situations can result in households having the same purchasing power of earned income.

Through curve fitting, the equations for the two index iterations were defined. A curve was also fitted to the top ten percentile, so as to evaluate this decile in greater granularity. Using these three equations a decision tree was created demonstrating how a household can calculate the dependency

index of their peer group (figure 13). This is possible as all three iterations were fitted with a PolyRatio(3,3) curve with incredible accuracy and fit, so an equation for each possibility was defined and added to the decision tree in figure 13. Using this decision tree it is possible for a household to get an accurate representation of the dependency index of their peers for comparison or planning purposes. The second hypothesis was thus accepted, that an inverse curvilinear curve would fit the data with accuracy and significance.

The third hypothesis of a downward trend in the top ten percentile can be accepted, as the data does show a strong negative trend toward the 100th percentile. But there is an increase in the index in the middle of the income decile which was not expected, as it was not considered likely that social grants would be a significant portion of income in this decile and taxes would be generally lower. There are several households that do receive large social grants in this spike, primarily though disability grants. Thus the trend is downwards, but due to the households near the middle of the income decile who received more social grants and paid lower tax the trend was a PolyRatio(3,3) and not the straight downward trend initially anticipated.

6 CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

The prior sections presented and thoroughly discussed the constitution and validation of the dependency index. The index is in itself a novel construction, and hence the majority of this paper has been to define its structure and why it can be used as a representative function for the dependency of the citizenry on the state. The significance of this study, as outlined in section 2.4, will now be discussed with potential areas for future research and utilisation of the dependency index suggested. In this concluding chapter an understanding of the various index iterations and how they were constructed and interpreted is assumed.

6.2 International comparison and benchmarking

Comparison of the dependency index across countries will likely be the most useful application of the model. The Gini coefficient, which measures income inequality through the Lorenz curve (Zagier 1983), is an example of a popular international comparator that is highly eloquent in its representation as a single coefficient for an entire population (Dalton 1920). This is possible because the Gini coefficient is either calculated using income or consumption, varying in some cases between total income and income from wages, on the shaped Lorenz curve which can be presented as a single figure

determining the shape of that curve (Zagier 1983). On the other hand, the Dependency Index needs to be viewed as a discrete value per income decile on a Cartesian scale, and is thus not as elegant for popular consumption. Indeed the Dependency Index is a representation of the fiscal relationship between the state and its citizenry, and will need to be contextualised for a particular country's growth path. Having said this, the Dependency Index considers an array of interesting metrics that can be very informative when analysing the fundamentals of an economy, and it is argued here that the index is a robust, novel and elucidating tool for the analysis of populations. The Dependency Index, similarly to the Lorenz curve, measures population in households from lowest to highest 'income' on the x-axis, and the dependent variable representing the index on the y-axis.

The reason that international comparison and benchmarking is likely the best use for the model is that the dependency index consolidates the interaction between the state and its citizenry into a single score per income decile. It would be extremely interesting to plot the dependency index for countries such as Brazil, the United States of America, India and Britain on the same axes as the dependency index for South Africa. In one snap shot a comparison of the fundamental state relationship with its citizenry across nations is possible, without purchasing power parity considerations. For example in Brazil personal income tax can reach 60%, as the state has well entrenched social policies, grants and assistance (Lierberman, 2001). Thus it is likely that the dependency index in Brazil could be similar to that in RSA, but potentially with even greater range. It would be interesting to see when access to social assistance peaks and how effective the state has been in reaching those in dire poverty. Similarly Britain has a very well entrenched social policy (Esping & Andersen, 1995), and a progressive tax rate, so the magnitude and shape of its dependency index would be of great interest for comparison. In developed economies such as Britain and the United States of America however, it is likely that social services are provided more effectively in the lowest incomes, as these countries have fewer rural populations without access to services (World Bank Data). Thus it is likely that access to social services would occur earlier than in RSA. Also as household employment is more likely in developed economies (World Bank Data), it is likely that the magnitude of the index would be less than in RSA. It is also likely that households in developed economies would show a sub 1 dependency factor (net money flow from household to state) in an earlier decile than income decile 8 (Figure 7) as employment rates are higher (World Bank Data). Thus the best area for additional research would be to compare and contrast several different economies on one Cartesian scale, having constructed their dependency indices from raw data in a similar census / national survey year.

The comparison of the index across developing and developed nations would likely show the greatest difference in magnitudes and shape, which is interesting in itself as a single graph can represent much about an economy. But most useful would be a comparison of the index across similar countries

whose policies and state implementation could be benchmarked. These benchmarks could form part of economic policy roadmaps in the transformation of the economy on a macro scale with micro inputs.

6.3 Provision of state services

Although the state attempts to provide services and assistance to those most in need the reality, as can be seen from the dependency index, is that the neediest groups in income decile 1 and 2 receive far less social assistance than those near the middle of the income divide (Table 3). The use of the winsorized index iteration with a number of different outlier ratios (2:1 – 15:1) can be plotted on the same Cartesian plane, so the income deciles with the greatest number of outliers can be visualised. This could assist policy makers in adjusting the implementation and eligibility requirements of social grants to smooth out this curve to maximise assistance to needy households and reduce dependency on state resources. For example a road map figure could be to increase the index score of income decile 1 to 1.5 by 2015, whilst reducing those outliers who receive more social assistance than they consume. Thus the Dependency Index shows how effectively the state is providing alleviation to poverty across the full income range, and can be used to benchmark progress of various initiatives.

6.4 The tax burden

Similarly to the above, the extent, depth and location of the tax burden can be visualised through the dependency index. Thus the index can indicate the extent to which income tax is eroding buying power per income decile. Threshold values could be identified in these deciles indicating when the burden is likely too large to encourage growth and investment. This could give an indication on when the pressure on a particular income category is starting to discourage legitimate tax returns and fiscal obedience (Kirchler, Hoelzl and Wahl 2008). The revenue stream from income tax is the largest contributor to the state fiscus (National Treasury and SARS 2011), and through the index various scenarios can be considered as to how this revenue can be maximised without affecting the purchasing power of households significantly.

6.5 Contributors or consumers of state capital

When an income decile has an index score below 1, it has a net positive effect on the state's fiscus. Conversely when the income decile has an index score above 1 it has a net negative effect on the states fiscus. The degree of divergence from 1 indicates the magnitude of that dependency. Thus if a household has an index score of 0.5, that household can only consume 50c for each rand that it earns,

so that household has a 50% utility of earned income. But if a household has an index score of 2.0, that household can consume R2 for every rand that it earns in income, so that household has a 200% utility of earned income. The balance is directly from the state as cash equivalents, so it can be determined which households, groups, or deciles are net consumers or contributors to the state fiscus. Similarly the state can scenario plan various initiatives to ensure it maximises revenue, whilst providing the largest amount of support to needy groups for human capital development. For example the index could inform the state that income decile 5 is receiving the greatest benefit from state, and in many cases are becoming wealthier from the state provision (consumption less than social grants). Thus the state could consider reducing these grants to be redistributed to households in income decile 1 and 2 which benefit the least from such state provisions.

6.6 Bottle necks in the economy

Economic bottlenecks are most often considered in terms of supply and demand relationships that limit the progression of nations (Mincer 1996, Valdes 2006). It is also true that bottlenecks can be experienced within the progression of citizenry from poor to rich, and progressive policies can have the unintended consequences of retarding growth and preventing economic inclusion (Bialer 1986, Hayek 1960). We show here through the dependency index that the progressivity of the RSA tax and social welfare system creates potential bottlenecks in the progression of citizens from poor to rich.

To illustrate this point effectively, the winsorised index to a ratio of 5:1 shown in table 10 can be used (the upper limit factor in this study). The index rises from 1.22 in income decile 1 to 2.09 in income decile 2, this illustrates that there is strong incentive for households to rise out of income decile 1. This is because each rand that they earn in income decile 2 will essentially have a greater purchasing power of 72%, as they will gain access to relatively large amounts of social assistance. Between income decile 2 and income decile 5 there are only small changes in the index of a maximum of 6%. Thus there is an incentive for households to keep rising in income as their purchasing power remains the same but their income is increasing per income decile.

Income decile 5-7 represent a strong disincentive to become wealthier however, as the index changes from 2.25 in income decile 5 to 1.27 in income decile 7. This bottle-neck in the progression of households can best be illustrated by comparing the required earned income to total income percentage across the two deciles. Income decile 5 has a maximum per household income of R26,290 and income decile 7 has a minimum per household income of R36,048, therefore R9,758 is the minimum difference in income between the two income deciles. This means that the household at the top of income decile 5 needs to increase its income by 27% in order to be in the bottom end of

income decile 7. But the required amount of earned income increases by 59% (adjusted for index) as social grants begin to drop off and income tax starts to be levied by the state. Thus in order for that household to move from income decile 5 to income decile 7, it needs to earn 59% more income not 27% more income. This could represent a significant disincentive on a return on effort basis.

Thus it is in the interest of the state to ensure that the slope of the index rises quickly from income decile 1, and then gradually decreases to income decile 10 so as not to create bottle necks in the ability of households to create wealth. If such a bottle neck exists, it is likely that transformative processes to increase the lot of black households could be limited to such bottle necks, as the vast majority of households migrating from income decile 5 to 7 will be black households (Stats SA). This may create further entrenchment of the status quo, where the richest households remain the richest households and rising largely black "potentially affluent" households burgeon around decile 5. This will likely further entrench socioeconomic disparity and racial tension, as the establishment of the rising middle class could be hampered by such bottlenecks.

6.7 Sensitivity testing of proposed tax and social policies

Proposed changes can be directly inputted into the data so that the net result on the index per decile can be seen. Changes in the cost of services, the value of child care grants, the exemptions from consumption tax, income tax rates and brackets etc. can all be considered for their effect on the purchasing power of households. This will allow sensitivity analysis on the various policy considerations to see the net effect on the economic wellbeing of the population. It would likely be possible for computer simulations to be run to allocate resources for the effective achievement of a policy road map derived through selecting index thresholds at different income deciles. For the consideration of comparing income deciles in terms of economic wellbeing it would be necessary to incorporate savings into the higher income deciles, as the index does not consider this.

6.8 Recommendations

The results from this study are not definitive in terms of recommendations derived from the dependency index, rather this study opens up a deep vein for further research. The purpose of this study was to propose this model as a novel mechanism of economic analysis of populations separated into income deciles. Much insight can be derived from this model, and by further increasing its integrity it could become a primary source of economic policy decision making. This study should be scrutinized to ensure representivity and then be considered for the creation of a high level macro-economic comparator index driven through microeconomic data.

6.9 Suggestions for further research

There are numerous avenues for further research, but a few particular studies could work well to increase the validity of the study, as well as show its usefulness for international and micro economic analysis.

- Perform the same study on the 2010/2011IES which will be released in 2012, preferably after data cleaning through DataFirst. This IES should be a better representative sample, and indicate much of the recent increases in the cost of services and social assistance as outlined in the literature review.

- Perform a study with the same methodology on other international markets (Brazil, USA, China, India etc.) for comparative purposes. These should be very interesting when plotted on the same Cartesian scale, where benchmarking could be possible.

- Perform qualitative studies as to the perceived value of each rand earned through work by households in different deciles. It would be interesting to note whether there is a perceived difference between the utility of income between deciles, and which deciles are more aware of this difference.

- It would be interesting to link the occurrence of social unrest, strike action, service delivery protests etc. to the income decile of households involved in such action and their respective Dependency Index values. This would indicate whether the dependency on the state of specific households leads those households to be more or less likely to increase their demands from the state and employers. Several covariates would likely need to be used in order to account for any causality in this matter (White and Chalak 2006).

- A causal study could be performed to see whether there is any link between households who evade tax and their relative dependency index score (White and Chalak 2006). It could be that as certain thresholds of tax are reached and dependency index scores begin to drop an increased propensity to evade tax is noticed (Poterba 1987). Similarly there could be a high level of tax evasion for households who are just becoming wealthy enough to pay tax, but are not accustomed to doing so.

- A study indicating the relative value obtained through state provided services relative to tax paid can be used as a weighting on the dependency index. For example, wealthy households often make use of private services (school, security, medical aid etc.) even when a portion of their tax is used to subsidize public services for the same function as they are paying private service providers (Landman 2002). Increased tax allocation to public services unused by the population paying for such revenue, in the form of income tax, is likely to persist with the implementation of initiatives such as the National Health Insurance (2012 National budget speech). On the other hand, lower income households do not pay income tax, but make use of the subsidised public services paid for by income tax revenue. Thus there could be a differential in the utility value of public services to different households running inverse to the amount of income paid to the state for the provision of such services. It is important to note that the private services paid for by income tax contributors are in most cases significantly superior to those provisioned by the state (Fiske and Ladd 2004). Thus the argument could be made that the further investment by these households in such private services adds to the development of human capital significantly, which will likely increases disparity between groups.

- The impact of the dependency index on inflation could be elucidated as a driver of inequality and wage distribution. For example a wealthy household might earn R500,000 per annum in income, but pay R200,000 of that income to the state for taxes. Thus the effective income for consumption to that household is only R300,000. If inflation is 6% on consumption, it is likely that the real inflation on income is higher than 6% for this household, as it would require more than 6% earned income to offset the inflation on consumption due to income tax. This could perpetuate higher wage increases at higher income deciles, which could be perceived as increasing income disparity whilst the real purchasing power of households remains the same.

- Determine the impact of indirect taxes on the utility of earned income, as this study focuses solely on income tax (direct tax) and VAT (indirect tax). In South Africa there are several indirect taxes that erode the purchasing power of post income tax income, which could have significant impacts on the utility of such income (Sørensen 1994). These exist in all countries in different forms from import and export taxes to capital gains tax etc. In the South African context there are several additional indirect taxes. An example of this is the newly introduced (2012 National budget speech) e-tolling tax, which taxes vehicles for travelling on certain public roads, whose provision is already allocated through income tax and a fuel levy

(www.sars.gov.za). Public transport and other special vehicles are exempt from paying these road taxes, and hence different population groups are affected in different ways (2012 National budget speech). This will likely result in material impacts on the dependency index, as the skewed influence would be experienced across income deciles. As the economic burden of increasing social welfare and above inflation public sector wage increases persists (Ndungu 2008), the state will need to find additional ways to increase its revenue, largely from the same economic minority group. This is likely to play out in the form of numerous indirect taxes targeted at those in higher income deciles (www.sars.gov.za) to increase state revenue streams. These indirect taxes could have significant effects on the dependency index, and it is recommended that these factors be integrated in further research.

7 REFERENCES

Aaron, H. and McGuire, M. (1970) Public goods and income distribution, *Econometrica: Journal of the Econometric Society*, pp. 907-920.

Artsev, Y., Roshal, V. and Finkel, Y. (2006) *Consumer Price Indices - Measuring Across Households*. Paper presented at Nineth Ottawa Group Meeting on Prices, 14-16.

Berry, J. S. and Houston, K. (1995) *Mathematical modelling*, Edward Arnold.

Bialer, S. (1986) *The Soviet paradox: external expansion, internal decline*, IB Tauris.

Bourguignon, F. (2004) The poverty-growth-inequality triangle, *Poverty, Inequality and Growth*, p. 69.

Case, A. and Deaton, A. (1998) Large cash transfers to the elderly in South Africa, *The Economic Journal,* 108(450), pp. 1330-1361.

Cook, T. D., Campbell, D. T. and Day, A. (1979) *Quasi-experimentation: Design & analysis issues for field settings*, Houghton Mifflin Boston.

Cornes, R. (1995) Measuring the Distributional Impact of Public Goods, van de Walle and Kimberly Nead edited, Public Spending and the Poor: Theory and Evidence, Johns Hopkins Press

Crawford, I. and Smith, Z. (2002) Distributional Aspects of Inflation. *The Institute for Fiscal Studies Commentary*, No. 90.

Dalton, H. (1920) The measurement of the inequality of incomes, *The Economic Journal,* 30(119), pp. 348-361.

De Wulf, L. (1975) Fiscal Incidence Studies in Developing Countries: Survey and Critique *Staff Papers-International Monetary Fund*, pp. 61-131.

Devereux, S. (2005) Can minimum wages contribute to poverty reduction in poor countries?, *Journal of International Development,* 17(7), pp. 899-912.

Devereux, S. (2007) Social pensions in Southern Africa in the twentieth century, *Journal of Southern African Studies*, pp. 539-560.

Devereux, S. (2010) Building Social Protection Systems in Southern Africa, *Background paper for the European Report on Development*.

Edwards, L. (2001) Globalisation and the skills bias of occupational employment in South Africa, *South African Journal of Economics,* 69(1), pp. 40-71.

Esping Andersen, G. K. J. (1995) Welfare states without work: The impasse of labor shedding and familialism in continental European social policy, *Estudios/Working Papers* (71)

Fedderke, J. W. and Luiz, J. M. (2008) Does human capital generate social and institutional capital? Exploring evidence from South African time series data, *Oxford Economic Papers,* 60(4), p. 649-682.

Fedderke, J. W., de Kadt, R. and Luiz, J. (2001) Indicators of political liberty, property rights and political instability in South Africa: 1935-97, *International Review of Law and Economics,* 21, pp. 103-134.

Fiske, E. B. and Ladd, H. F. (2004) Balancing public and private resources for basic education: school fees in post-apartheid South Africa, *Changing class: Education and social change in post-apartheid South Africa. Pretoria: HSRC*.

Finn, A., Leibbrant, M. and Woolard, I. (2009) Income and expenditure: Analysis of the NIDS Wave 1 Dataset Discussion Paper no. 5.

Garner, T.I., Johnson, D.S. and Kokoski, M.F. (1996) Experimental Consumer Price Index for the Poor, An. Monthly Lab. Rev. 119:32.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. (2004) Do institutions cause growth?, *Journal of Economic Growth,* 9(3), pp. 271-303.

Haggard, S. and Kaufman, R. R. (1997) The political economy of democratic transitions, *Comparative Politics*, pp. 263-283.

Hall, R. E. and Jones, C. I. (1999) Why Do Some Countries Produce So Much More Output Per Worker Than Others?*, *Quarterly journal of Economics,* 114(1), pp. 83-116.

Hawkins, R. H. (1980) 'Infiltration and curve numbers: some pragmatic and theoretic relationships', *A paper presented at*, held in, on the, organised by ASCE.

Heideman, L. (2011) Municipal Local Economic Development and the power of the local multiplier effect: piloting a community enterprise identification method in South Africa and Namibia.

Hobijn, B. and Lagakos, D. (2005) Inflation inequality in the United States, *Review of Income and Wealth,* 51(4), pp. 581-606.

Hoogeveen, J. G. and Özler, B. (2004) Not separate, not equal: Poverty and inequality in post-apartheid South Africa, *World Bank*.

Hayek, F. A. (1960) *The constitution of liberty,* Vol. 231 pp 306-332, University of Chicago Press Chicago.

Idson, T. and Miller, C. (1999) Calculating a price index for families with children: implications for measuring trends in child poverty rates, *Review of Income and Wealth,* 45(2), pp. 217-233.

Johnson, R. B. and Onwuegbuzie, A. J. (2004) Mixed methods research: A research paradigm whose time has come, *Educational researcher,* 33(7), p. 14.

Kidder, L. H., Judd, C. M., Smith, E. R. and Issues, S. f. t. P. S. o. S. (1986) *Research methods in social relations*, Holt, Rinehart and Winston Fort Worth, TX.

Kirchler, E., Hoelzl, E. and Wahl, I. (2008) Enforced versus voluntary tax compliance: The "slippery slope" framework, *Journal of Economic Psychology,* 29(2), pp. 210-225.

Klasen, S. and Woolard, I. (2009) Surviving unemployment without state support: unemployment and household formation in South Africa, *Journal of African Economies,* 18(1), pp. 1-51

Landman, K. (2002) 'Gated communities in South Africa: Building bridges or barriers', *A paper presented in Pretoria.*

Lichbach, M. I. (1989) Does economic inequality breed political conflict, *World Politics: A Quarterly Journal of International Relations*, pp. 431-470.

Liebbrandt, M. Woolard, I. and Finn, A. (2009) Income & expenditure inequality: analysis of the NIDS Wave 1 dataset, University of Cape Town: National Income Dynamics Study (NiDS).

Lieu, P. T., Chang, C. and Chang, J. (2004) inflation rate Variations across household: empirical evidence from taiwan, *international Journal of Business,* 9(1), pp. 103-124.

Lindholm, L. and Rosen, M. (2000) What is the "golden standard" for assessing population-based interventions?—problems of dilution bias, *Journal of epidemiology and community health,* 54(8), p. 617.

Lund, F. (1993) State social benefits in South Africa, *International Social Security Review,* 46(1), pp. 5-25.

McEwen, H., Kannemeyer, C. and Woolard, I. (2009) Social Assistance Grants: Analysis of the NIDS Wave 1 Dataset Discussion Paper no. 10.

McDonald, D. A., Pape, J. and Council, H. S. R. (2002) *Cost recovery and the crisis of service delivery in South Africa*, Zed Books London.

McGranahan, L. and Paulson, A. (2006) Constructing the Chicago Fed Income Based Economic Index–Consumer Price Index: Inflation Experiences by Demographic Group: 1983-2005, *Federal Reserve Bank of Chicago WP 2005,* 20.

Meth, C. and Dias, R. (2004) Increases in poverty in South Africa, 1999-2002, *Development Southern Africa,* 21(1), pp. 59-85.

Moaddel, M. (1994) Political conflict in the world economy: A cross-national analysis of modernization and world-system theories, *American Sociological Review*, pp. 276-303.

Morgan, J. (1962) The anatomy of income distribution, *The Review of Economics and Statistics,* 44(3), pp. 270-283.

Moss, T., Pettersson, G. and Van de Walle, N. (2006) An aid-institutions paradox? A review essay on aid dependency and state building in sub-Saharan Africa, *Center for Global Development Working Paper No. 74, Mario Einaudi Center for International Studies Working Paper No. 11-05.*

Muller, C. (2003) Measuring South Africa's informal sector: An analysis of national household surveys, *Development Policy Research Unit Working Paper No. 03,* 71.

National Treasury and SARS. 2011 Tax Statistics. Available (online): http://www.sars.gov.za

National Treasury. (2010) Pocket guide to South Africa: Energy and water.

Ngandu, S., Altman, M., Cross, C., Jacobs, P., Hart, T. and Matshe, I. (2010) THE SOCIO-ECONOMIC IM DOWNTURN ON RSA RESPONSES AND POL. Pretoria.

Ndungu, S. K. (2008) Collective Bargaining. Wage and non-wage settlement trends in the South African labour market, Johannesburg: NALEDI. Pretoria.

Oosthuizen, M. (2007) Consumer Price inflation across the income distribution in South Africa, *Development Policy Research Unit Working Paper 07,* 129.

Poterba, J. M. (1987) Tax evasion and capital gains taxation, National Bureau of Economic Research Cambridge, Mass., USA

Ravallion, M., Datt, G. and Chaudhuri, S. (1993) Does Maharashtra's Employment Guarantee Scheme guarantee employment? Effects of the 1988 wage increase, *Economic Development and Cultural Change,* 41(2), pp. 251-275.

Rogerson, C. (2009) Strategic review of local economic development in South Africa, *Pretoria: DPLG and AHI.*

Sabates-Wheeler, R. and Devereux, S. (2010) Cash transfers and high food prices: Explaining outcomes on Ethiopia's Productive Safety Net Programme, *Food Policy,* 35(4), pp. 274-285.

Samson, M. (2009) The impact of social transfers on growth, development, poverty and inequality in developing countries, *Building Decent Societies. Rethinking the Role of Social Security in Development*, pp. 122–150.

Samson, M., Babson, O., Haarmann, C., Haarmann, D., Khathi, M. G., Mac Quene, K. and van Niekerk, I. (2002) Social security reform and the basic income grant for South Africa, *Economic Policy Research Institute. Report commissioned by the International Labour Organization (ILO).*

Sen, A. (1995) The political economy of targeting, *Public spending and the poor: Theory and evidence*, pp. 11-24.

Sørensen, P. B. (1994) From the global income tax to the dual income tax: Recent tax reforms in the Nordic countries, *International Tax and Public Finance,* 1(1), pp. 57-79.

Statistics South Africa. (2008) Income and expenditure of households 2005/2006: Analysis of results. Report 01-00-01. Available (online): http://www.statssa

Statistics South Africa. (2008) Income and expenditure of households 2005/2006: Statistical release. P0100. Available (online): http://www.statssa

Taktek, N. (1998) Comparative Study of Analytical Consumer Price Indexes (CPI) for Different Subgroups of the Reference Population. *Analytical Series*, No. 13. Statistics Canada.

Thomas, R. H. (1994) Human rights and street children in South Africa 1, *Development Southern Africa,* 11(2), pp. 205-215.

Thorbecke, E. and Charumilind, C. (2002) Economic inequality and its socioeconomic impact, *World Development,* 30(9), pp. 1477-1495.

Van der Berg, S., Wood, L. and Le Roux, N. (2002) Differentiation in black education, *Development Southern Africa,* 19(2), pp. 289-306.

Whiteford, A. and Van Seventer, D. E. (1999) *Winners and losers: South Africa's changing income distribution in the 1990s*, WEFA Southern Africa.

White, H. and Chalak, K. (2006) Parametric and Nonparametric Estimation of Covariate-Conditioned Average Causal Effects, *UCSD Department of Economics Discussion Paper*.

Woolard, I., Harttgen, K. and Klasen, S. (2010) The Evolution and Impact of Social Security in South Africa. MOBILISING EUROPEAN RESEARCH FOR DEVELOPMENT POLICIES.

Woolard, I. and Wilson, R. (2004) Developing a National Skills Forecasting Tool for South Africa, *Paper commissioned by EEPR, Human Sciences Research Council*.

Zagier, D. (1983) Inequalities for the Gini coefficient of composite populations, *Journal of Mathematical Economics,* 12(2), pp. 103-118.

Zeller, R. A. and Carmines, E. G. (1980) *Measurement in the social sciences: The link between theory and data*, Cambridge Univ Pr.

ANNEXURE A

Table 20:Income deciles

Table 20 shows the decile ranges as defined by Statistics South Africa. There thresholds were used to create the dependency index per income decile based on total household income.

Income Deciles	1	2	3	4	5	6	7	8	9	10
Income/annum	R 7 239	R 11 378	R 15 252	R 20 203	R 26 290	R 36 048	R 53 387	R 90 472	R 180 212	>

Table 21:Tax Brackets

The tax brackets and rebate structure used for this analysis was that of the 2008 SARS revision (www.sars.gov.za). Although the data from the IES was the 2005/2006 data series, it is more appropriate to consider the tax structure as of 2008 for better comparison to future and current trends. It is likely due to the lower income tax threshold, that income tax would be paid across sooner than income decile 7 if the past structure was used. Individual contributions per economically active household member were considered, to get actual tax paid. This was then calculated as a function of total income to get the average household tax rate.

Table 2.1: Personal Income Tax (PIT) brackets, 2002/03 and 2008/09

	2002/03		2008/09		Marginal PIT rates	Cumulative percentage increase
Rand						
Brackets	–	40 000	–	122 000	18%	205.0%
	40 001	80 000	122 001	195 000	25%	143.8%
	80 001	110 000	195 001	270 000	30%	145.5%
	110 001	170 000	270 001	380 000	35%	123.5%
	170 001	240 000	380 001	490 000	38%	104.2%
	240 001	and over	490 001	and over	40%	
Rebates						
Primary		4 860		8 280		70.4%
Secondary		3 000		5 040		68.0%
Tax thresholds						
Below age 65		27 000		46 000		70.4%
Age 65 and over		42 640		74 000		73.5%

Table 22:Free services per household

Free subsidised services are the service fees as of the 2007 where the minimum thresholds were adjusted. This results in an annual fee per household per year, that was applied to all households below income decile 7 that had access to an electrical point in their homes.

Service	Price/unit	Annual subsidised units	Annual value of Service
Water	R 3.24	72	R 233.28
Electricity	R 0.65	600	R 390.00

Table 23:VAT paid per income decile

The average consumption tax paid was calculated by allocating weightings to the average basket purchases (Stats SA IES 2008) for each group, and then multiplying these weightings to get the average consumption tax paid per household.

Percentage of consumption by category by income decile:

Category	1	2	3	4	5	6	7	8	9	10
food + beverage	36%	36%	34%	33%	31%	28%	25%	20%	13%	7%
Fuels, energy	21%	19%	19%	18%	19%	18%	19%	23%	27%	25%
Health	2%	2%	2%	2%	2%	2%	2%	2%	2%	2%
Education	1%	1%	1%	1%	1%	2%	2%	3%	3%	3%
Clothing	10%	9%	9%	9%	9%	8%	8%	7%	5%	3%
Total	69%	67%	64%	63%	61%	58%	55%	53%	51%	39%

Percentage of category exempt from VAT per income decile:

Category	1	2	3	4	5	6	7	8	9	10
food + beverage	80%	80%	70%	60%	50%	50%	40%	30%	30%	20%
Fuels, energy	80%	80%	80%	70%	60%	50%	40%	20%	10%	5%
Health	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Education	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Clothing	50%	50%	50%	40%	30%	20%	10%	0%	0%	0
Total	53%	52%	46%	39%	32%	28%	22%	15%	12%	7%
Income decile	1	2	3	4	5	6	7	8	9	10
VAT paid per decil	6.6%	6.8%	7.6%	8.5%	9.5%	10.0%	10.9%	12.0%	12.4%	13.0%

Table 24: Representivity of sample data compared to the NiDS 2009 dataset

Table 24 shows the representivity of the data in this study compared to the NiDS wave data series published in 2009. As can be seen there is superficial likeness between the two data sets, which shows a likely agreed representivity between these two independent studies.

	Black		Coloured		Indian/Asian		White	
	Sample	NiDS	Sample	NiDS	Sample	NiDS	Sample	NiDS
Income decile 1	94%	97%	6%	2%	0%	0%	0%	1%
Income decile 10	24%	28%	11%	7%	6%	9%	58%	57%