University of Pretoria
Department of Economics Working Paper Series

Using a Static Micro-Simulation Model to Evaluate the South African Income Tax System

K. L. Thompson and N. J. Schoeman
University of Pretoria
Working Paper: 2006-13
June 2006

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413
Fax: +27 12 362 5207
http://www.up.ac.za/up/web/en/academic/economics/index.html

Using a Static Micro-simulation Model to Evaluate the South African Income Tax System

KL Thompson and NJ Schoeman
University of Pretoria

March 2006

1 Introduction

The purpose of this study is to develop a static micro-simulation model (MSM) of personal income tax in South Africa. MSMs "can be useful both in creating theories and in examining effects of policy options" (Merz, 1991:77). MSMs are in particular very helpful tools for analysing the distributional effects of changes in policy variables on the population, particularly in the field of fiscal policies such as taxation, social security and welfare. "However, there are still many important areas of public policy to which micro-simulation has not yet been applied" (Brown and Harding, 2002:2). Few MSMs incorporate health, disability and aged-care benefits, or impacts of changes in non-fiscal and non-socio-economic variables such as interest rates.

A static micro-simulation model does not account for any changes in the population structures of the sample used. However, since this particular model estimates individual expenditure patterns, it can be used to evaluate the effects of personal income tax changes on individual expenditure, by estimating expenditure before and after a change in the tax policy. In this way, behavioural patterns are included in the model, although it is not a conclusively behavioural MSM.

As already mentioned, this micro-simulation model estimates individual expenditures. This is done using individual survey data from the 1999 Statistics South Africa October Household Survey (OHS). The primary data of interest in the OHS survey are gross income data pertaining to each individual, *Y*. Each individual's taxable income is then deduced using ratio's of medical aid and pension contributions from South African Revenue Services (SARS) IRP5 data:

$$Y_{taxable} = f(Y, medical\ aid\ contributions, pension\ contributions) \qquad \textbf{... 1}$$

The personal income tax paid by each individual is then calculated by applying the South African Department of Finance's 1999 tax policy to individual taxable incomes:

$$T = f(Y_{taxable}, tax\ structure) \qquad ... 2$$

Individual disposable incomes are then calculated by subtracting the personal income tax paid from the calculated taxable incomes:

$$Y_d = Y_{taxable} - T \qquad \text{... 3}$$

Using the calculated disposable income, savings data, and a variety of demographic indicators, individual expenditure data are then inferred using regression techniques:

$$C = f(Y_d, S, demographics) \qquad \text{... 4}$$

The MSM created in this study also serves the purpose of analysing the tax gap between actual taxes received by SARS and potential taxes calculated using Equation 3. This will be done by projecting the microeconomic MSM results to macroeconomic figures, and comparing actual data for variables such as taxable income, disposable income and consumption expenditure, with the estimated figures using the MSM results.

The outline of this paper is as follows: a review of micro-simulation modelling in Section 2 will highlight the types of MSMs in use, the procedures used in creating these models, and the strengths and weaknesses of MSMs; the data used in this particular MSM will be discussed in detail in Section 3; Section 4 will present the empirical results of this MSM; and the final conclusions and policy recommendations are presented in Section 5.

2 Review of Micro-simulation Modelling

The empirical methods used in this study, specifically micro-simulation modelling, and the types thereof are discussed in this section.

A MSM 'attempts to model and simulate the whole distribution of policy target variables, not only their mean values' (Klevmarken 1997:2). The main advantage of micro-simulation modelling over conventional econometric modelling is that it allows for heterogeneous behaviour, as not every economic subject necessarily behaves like the average subject (Klevmarken 1997:2).

Micro-simulation modelling simulates the behaviour of individual economic units, such as households, individuals, firms, government departments etc. The economic unit of interest in the present case is the tax-paying individual. A MSM draws on a population database, which can be cross-sectional or longitudinal (a panel). The database depicts various characteristics of each economic unit. In this case, the database used is a cross-section in the form of the 1999 October Household Survey.

The estimated effects obtained from micro-simulation modelling can be aggregated over all the units to obtain aggregate macroeconomic estimates.

2.1 Types of Micro-simulation Models

There are many classifications according to which MSMs can be described, in particular their behavioural nature (i.e. behavioural vs. non-behavioural models), their incorporation of time (i.e. static vs. dynamic models) and their integration of 'space'. These distinguishing characteristics are discussed in this section of this paper. Most econometric-based MSMs to date are non-spatial – they do not incorporate information on where the individuals being affected actually live, i.e. they tend to be based on national data. Using regional data allows for the prediction of spatial impacts on households and individuals, as well their consumer reactions to policy changes (Brown and Harding, 2002:4-5).

2.1.1 Dynamic vs. Static Models

A concept in micro-simulation modelling, known as aging the data, allows for changes in states or behaviours to occur. In this way, it is possible to distinguish between two types of MSMs (Klevmarken 1997:3): static models and dynamic models. Static models do not update the population structure endogenously, but any changes are accounted for by re-weighting of the data points. This is based on the fact that most micro-simulation databases are samples of a population, and thus each representative individual/unit is assigned a weight, with the sum of the weights being equal to the population. It is also possible to engage in static aging by changing each record based on projections of future periods.

Dynamic models include inbuilt mechanisms that allow for structural and compositional changes of the population over time, such as life-course redistribution, wealth accumulation, demographic behaviour, labour market mobility, and poverty and social exclusion transitions (O'Donoghue 2001). These individual changes in behaviours and conditions are simulated. In large dynamic models behavioural equations predict the probabilities of given states or events occurring. This is sufficient for a cross-section model, but for panels that would require updating the population for future time periods, Monte Carlo simulations are used to age the population.

In theory, dynamic models would obviously be more realistic and more representative of the population being analysed, but in reality, static MSMs are in high demand due to their inexpensiveness, their relative ease of development and their simplicity of use.

2.2.2 Behavioural vs. Non-Behavioural Models

It is possible for both dynamic and static MSMs to contain behavioural relations (Klevmarken 1997:12-14). Where dynamic modelling implies the incorporation of time into an MSM, behavioural modelling implies the incorporation of relations that model the behaviour of economic subjects at any point in time. For example, a transition probabilities matrix differentiated between region and gender will differentiate individuals' behaviour according to region and gender. Behavioural modelling can also be performed by estimating relations in a macro model (e.g. growth rates of macro variables) and then disaggregating these and feeding them into a MSM. In the case of tax MSMs, the incorporation of econometrically estimated labour supply functions[1] and commodity demand functions would satisfy the conditions of behavioural micro-simulation modelling.

[1] It should be noted that this is the area raising the greatest difficulty for modellers, due to a lack of survey information on those who are unemployed (although this is less of a problem in South Africa, which has extremely high unemployment rates, and a proportional representation of the unemployed in the OHS), and also due to the difficulty of estimating the

According to Creedy *et al* (2002:12), the majority of tax MSMs are non-behavioural – they do not allow for effects of tax changes on individuals' consumption and labour supply. The advantages of these models are obvious: they are simpler to develop and uphold, and they are accessible by a wider range of users. The analyses of these models usually incorporates graphs and tables of tax paid (or disposable income or expenditure, or whatever the variable of interest is) for various income groups or demographic groups.

Behavioural modelling serves three purposes according to Klevmarken (1997:14): missing data can be imputed; the population or sample's demographic characteristics can be updated or aged; and most importantly, the adjustments to policy changes can be captured. Creedy *et al* (2002:13) emphasise the need for behavioural tax MSMs, as tax policy changes invariably alter the consumption of various goods, the participation of individuals in the labour market, and the welfare of taxpayers.

Many existing behavioural tax MSMs are however restricted, due to the implicit assumption of exogeneity of factors such as household formation, marriage, births, deaths, retirement, and education decisions (Creedy *et al*, 2002:14). The degree of population heterogeneity in non-behavioural models is somewhat greater than in behavioural models, due to the fact that certain households tend to be excluded if they do not conform to underlying econometric assumptions (Creedy *et al*, 2002:15).

Martini and Trivellato (1997:95) have offered criticism of using behavioural responses in static MSMs: it is not improbable that policy changes have short-run effects only, after which inertial behaviour compels behaviour to return back to its original form; and, models' predictions tend to be clouded in uncertainty. They believe that behavioural parameters are much more relevant in dynamic MSMs, where the parameters' magnitudes are required before any change in policy is even considered.

---

fixed costs of supplying labour, as well as in estimating the constraints on labour supplied (Creedy *et a*l, 2002:16).

Klevmarken (1997:19-20) also differentiates between the following 'families' of behavioural models:

(1) *Models of transition between different states*: transitional probabilities model such as Markov models, probit, logit and hazard rate models. Probabilities are estimated conditional on individual characteristics, with stochastic deviations from the mean. Parsimony is encouraged.

(2) *Count data models*: model the number of occurrences of an event in a given time span, such as Poisson models. Probabilities are estimated conditional on individual characteristics, with stochastic deviations from the mean. Parsimony is encouraged.

(3) *Continuous data models*: conventional (non)linear regression models and equation systems. Deviations from the average are generated by adding stochastic disturbances to the model's systematic component. Parsimony is encouraged.

(4) *Random assignment schemes*: model structure is implied and never estimated. Population observations close to the unit to be estimated pass on data to this unit.

Over-simplification of the behavioural models in the first three cases may be at the detriment of the main strong point of micro-simulation modelling – the ability to allow individuals to behave differently.

The procedural methodology used when constructing a MSM is discussed next.

2.2 Procedures in Micro-simulation Modelling

A behavioural MSM, according to Creedy *et al* (2002:8), has three components:[2]

---

[2] Martini and Trivellato (1997:89) include as an initial component a baseline database "containing information on individual units … in particular socio-demographic characteristics … and economic information".

i. A static model consisting of accounting/arithmetic equations with which incomes and tax are calculated.
ii. A labour supply model which quantifies individuals' tastes for labour supply, income and leisure time.
iii. A mechanism to link points *i* and *ii*, i.e. to allocate the appropriate supply of labour in the presence of specific tax-benefit systems.

The above components illustrate the partial equilibrium nature of micro-simulation modelling: only the supply side of the labour market is dealt with, whilst the demand side of the real economy is accounted for by commodity demands. This naturally widens the scope for the linkage of MSMs with other general equilibrium models.

According to Merz (1991:94) the requirements and 'rules' laid out in Table 1 need to be met in order to develop an efficient and representative static MSM.

The choice of modelling approach in micro-simulation modelling depends, not surprisingly, on the intended use of the model (Klevmarken 1997:3-4). A MSM which will be used for forecasting and policy recommendations needs to be firmly based on empirically 'real' data, and can be known as an 'empirical model'. MSMs which are used to explore assumptions, but not inference, about economic agents and markets, usually are empirically weak with parameters assigned on an *ad hoc* basis, and can be termed 'abstract models'.

In this study, a static MSM with simple behavioural responses of the South African personal income tax system is constructed. This is done by estimating individual expenditures pertaining to a specific tax policy, and comparing these to actual expenditure data. At a later stage in later studies, the effects of a change in tax policy will be analysed by comparing individual expenditures before and after this tax policy change. This is illustrated in Figure 1.

A dynamic MSM model with behavioural relations, however, is a lot more complex, as illustrated in Figure 2.

A discussion of the strengths and weaknesses of micro-simulation modelling, although already alluded to previously, will follow.

Table 1: Requirement profile for a static MSM

i. Initial data preparation and construction
  a. Microdata processing
  b. Macrodata processing
  c. Modifications of initial data
  d. Statistical methods for matching (merging microdata)
  e. Construction of initial file
  f. Extraction of subfiles

ii. Module construction
  a. Construction of micro modules
  b. Construction of macro modules
  c. Econometric and statistical methods for hypotheses testing and formulation

iii. Modifications of module parameters – model operations
  a. Scenario formulation
  b. Parameter changes
  c. Module changes
  d. Handling of module sequence
  e. Linkage micro to macro or other models
  f. Testing

iv. Adjustment of microdata
  a. Demographic adjustment: ‘static aging’
  b. Economic aging
  c. Stochastic changes and ‘alignment’
  d. Sensitivity analyses and changing aggregate control data
  e. Statistical adjustment methods

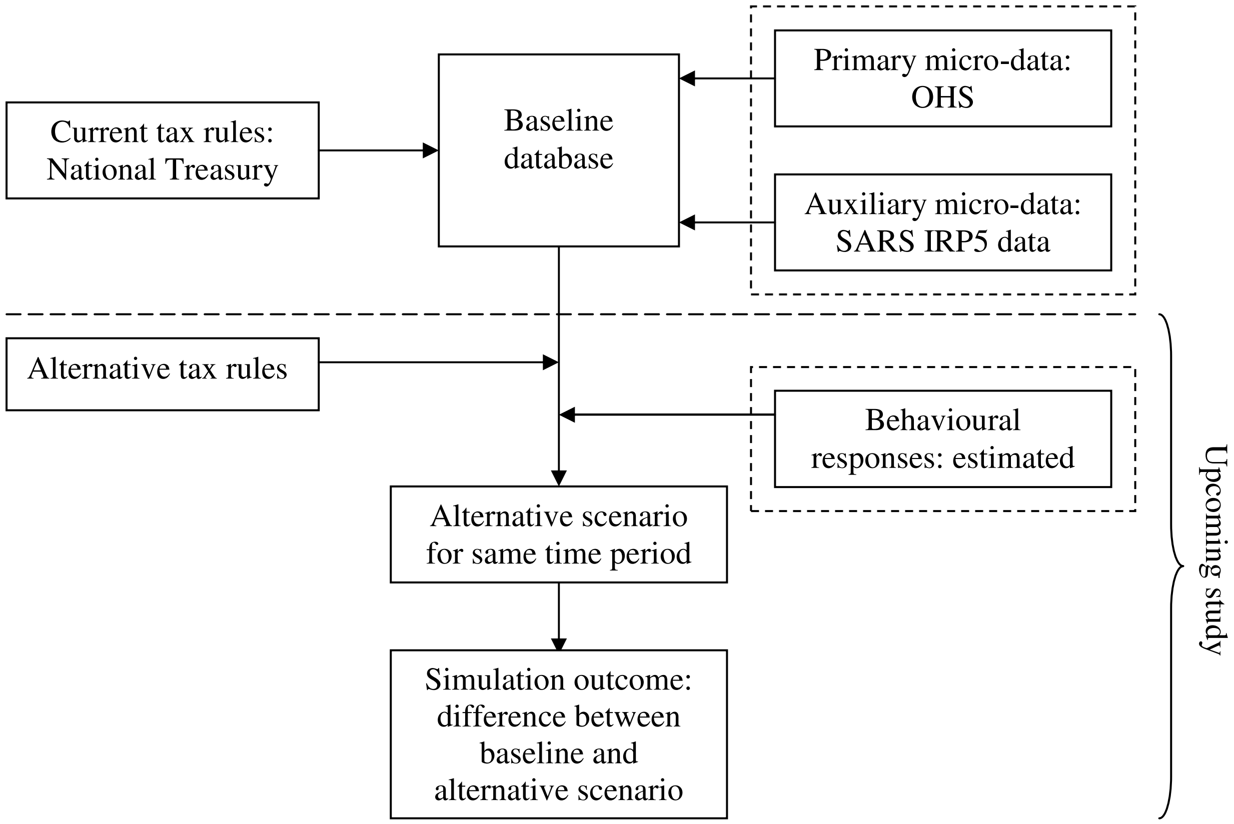
v. Evaluation of simulation
  a. Results of single simulation
  b. Results of several simulation runs
  c. Statistical methods for data analyses

vi. Efficiency in processing

vii. Ease of use

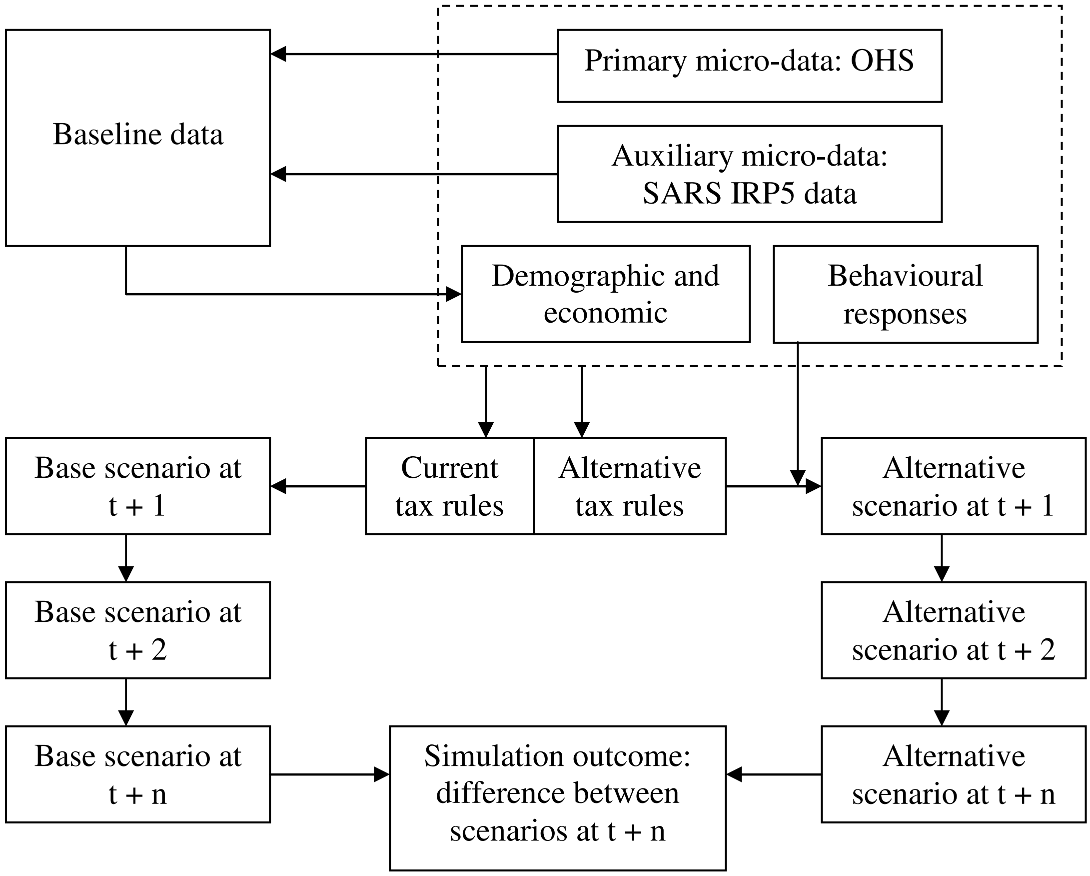
Source: Merz (1991:94)

Figure 1: Structure of the static behavioural MSM and data sources used in this study (micro-data in dotted box)



Source: Martini and Trivellato (1997:91)

Figure 2: Structure of a hypothetical dynamic behavioural MSM and data sources (micro-data in dotted box)



Source: Martini and Trivellato (1997:92)

2.3 Strengths and Weaknesses of micro-simulation models

As already mentioned, a major advantage of micro-simulation modelling is its ability to estimate behavioural relationships at a micro level, between individual economic units, and then to aggregate these to reflect macro relationships.

Micro-simulation modelling also permits distributional analysis, that is, examining the costs and benefits pertaining to various groups over a population.

Since micro-simulation modelling captures the original data's covariances, it retains significant interactions that aggregative and cell based methodologies tend to ignore.

In the field of taxation and policy analysis, the need to use models of some type is inescapable. Invariably, the model used is a simplification of reality, and these simplifications tend to be 'forced into the open', thereby enabling modellers to recognise the limitations of their modelling (Creedy *et al*, 2002:6). This is especially true for the case of independent researchers, who are obliged to publish their full results, unlike government modellers. It is for this reason that Creedy *et al* (2002:7) strongly support the use of several models, thereby benefiting from various strategies, and hopefully diminishing the variety and spread of simplified assumptions. They also warn against the use of a simple model as if it were a reflection of the real world, making strong, and sometimes permanent, policy recommendations on the basis of unrealistic models.

Since MSMs are not general equilibrium models, and represent only one side of, usually, one market, they may be deemed incomplete in their social and economic representations. According to Creedy *et al* (2002:7), the ideal MSM would be "a life-cycle, overlapping generations, dynamic general equilibrium open economy model … with endogenous choices regarding the education, occupational choice, labour supply, household formation, consumption and saving behaviour of all individuals". Furthermore, this ideal model should act like a 'black box' to the majority of its users, with results being generated simply and routinely. Since this might be regarded as an impossible feat, their aforementioned advice that as many different models as possible be consulted, seems all the more sensible.

A major shortcoming of micro-simulation modelling which is compensated for in macro models is the lack of endogenous feedback systems. It is for this reason that it is suggested that micro-simulation effects be incorporated into a macro model.

Lastly, a severe setback in micro-simulation modelling pertains not to the models themselves, but rather to data inadequacies. Micro data is cumbersome, time-consuming, and thus expensive, to work with. There are generally long lags between data collection (usually in the form of a survey) and the release of said data. This is in contrast to macro data which can be obtained on a quarterly, or even monthly, basis, with short lags between collection and dissemination. One of the implications of these data inadequacies is the incorporation of calibrated parameters, rather than estimated parameters. Creedy *et al* (2002:8) warn that this exercise should be kept to a minimum.

In the case of micro-simulation modelling of tax and transfer systems, the data inadequacies are particularly troubling, as accurate income data and accurate tax data are unlikely. This can be compensated for by combining data from governmental revenue agencies with survey data – an exercise that requires extreme caution in the combination process.

A discussion of the data used in this study ensues.

3 Data

The source data for this study is the 1999 October Household Survey (OHS) conducted by Statistics South Africa (Stats SA). The OHS questionnaire consists of 7 sections, containing specific information as follows:

i. Sections 1 and 4: Personal Details, e.g. age, education, employment, etc.
ii. Section 2: Births and Children
iii. Section 3: Employment Information
iv. Section 5: Migrant Workers
v. Section 6: Household Details
vi. Section 7: Farming Details and Practices

Data used in this study is from sections 1, 3, 4 and 6 from the OHS, as well as from the South African Revenue Services (SARS) filer information, and is shown below in Table 2. Characteristics pertaining to 30 921 individuals are used in this study.

Table 2: Data description

Variable	**Question used in questionnaire**	**Comment**
Individual income	*What is …'s total salary/pay at the main job? Including overtime, allowances and bonus, before any tax or deductions. Is this weekly, monthly or annually?*	The respondent had the opportunity to give an exact amount or to highlight an income category. The data used was continuous.
Deductions	*n/a*	This was extracted from SARS IRP5 and filer data. The deductions included were medical aid fund contributions and pension fund contributions.[3]
Personal income tax paid	*n/a*	This is calculated using the 1998/1999 Budget Review's tax brackets.
Expenditure		Derived from household expenditure: by dividing household expenditure by the number of adults in the home.
Savings	*n/a*	
Marital status	*What is …'s present marital status?*	People married in a civil or traditional ceremony are regarded as married, whilst widow(er)s, divorcees and those living together are regarded as being unmarried.
Number of people living in household	*n/a*	The number of people, including children and babies, that spend at least four nights per week in the house.
Gender	*Is … male or female?*	
Age	*Age in completed years*	
Race	*What population group does … belong to?*	The respondent chose between African/Black, Coloured, Indian/Asian, White or Other.
Education	*What is the highest level of education that … has completed?*	The respondent could choose between 23 categories. For the purposes of simplicity, these have been combined into: primary school; some high school but not matric; matric; an NTC I, II or III; university degree or diploma (under- or postgraduate).
Location i.e. rural or urban	*n/a*	The choice of an urban or rural area was dependent on the sampling of the survey.
Province	*n/a*	This was also dependent on the sampling technique of the survey.

3 For predetermined income categories, the percentage of deductions (*D*) to total taxable income (X_T) is used to approximate the percentage of deductions to total income (*X*). This then enables the subtraction of deductions (*d*) from total income for each individual (*x*) in the OHS sample.

The OHS was drawn by a two-stage sampling procedure conducted by Stats SA. "A sample of 30 000 households was drawn in 3 000 enumerator areas (EAs) (that is 10 households per enumerator area). A two-stage sampling procedure was applied and the sample was stratified, clustered and selected to meet the requirements of probability sampling. The sample was based on the 1996 Population Census enumerator areas and the estimated number of households from the 1996 Population Census. The sample was explicitly stratified by province and area type (urban/rural). Within each explicit stratum the EAs were stratified by simply arranging them in geographical order by District Council, Magisterial District and, within the magisterial district, by average household income (for formal urban areas or hostels) or EA. The allocated number of EAs was systematically selected with probability proportional to size in each stratum."

The inclusion probability (p_1) of an EA was based on the number of EAs in the sample in the i-th stratum (m_i) (where stratum is the District Council in a province), the number of persons residing in the selected EA (A_i), and the total number of persons in the population of the i-th stratum (sA_i):

$$p_1 = \frac{m \cdot A_i}{sA_i} \qquad \text{... 5}$$

"The measure of size was the estimated number of households in each EA. A systematic sample of 10 households was drawn". The inclusion probability of a household (p_2) was based on the number of households in the selected EA and on the fact that ten households per EA were systematically selected:

$$p_1 = \frac{10}{no.of\ households\ in\ selected\ EA} \qquad \text{... 6}$$

The implications of using survey data have been widely documented, and the conflicting view points are presented below.

3.1 Advantages and Limitations of using Survey Data

The fact that survey data presents information on an individual level makes it extremely powerful as a policy analysis tool, since the micro effects of changes can be scrutinised. This is as opposed to aggregate data.

According to Martini and Trivellato (1997:85), the questions modellers should ask themselves when analysing data available for policy analysis are the following:

i. What is the *interaction* between the policy and social conditions – are there any gaps in the policy's coverage?
ii. What are the *incentive* effects of the policy change – are work and savings decisions being effected?
iii. What are the *distributional* impacts of the policy change – who wins and who loses?

These questions form the basis of the development of the base data for the MSM. A requirement of the base data used in micro-simulation modelling is that it be a representative sample of individual units (Martini and Trivellato, 1997:93). In other words, it "must be able to reproduce reasonably well all the relevant characteristics of the population that are affected by the range of policies that the model should simulate". Martini and Trivellato (1997:94) go on to say that it is virtually impossible to find a primary data set containing all the information needed for simulating public policies, and thus auxiliary data sets are usually required to complement the primary information.

Martini and Trivellato (1997:98) advance repeated household surveys as the ideal database for micro-simulation modelling, as the sample size is large, the content covered in a household survey is extensive, and (ideally) recent data is made available periodically. However, there are various troubling issues in household surveys that hamper these 'ideal characteristics':[4]

4 These issues are proposed by and extracted from Martini and Trivellato (1997:98-100).

i. *Survey design problems*:
    a. *Representativeness*: under-coverage and incorrect population definitions lead to a lack of representativeness.
    b. *Multiple units of analysis*: inconsistency in units of measurement, particularly when integrating data bases, may arise, e.g. households vs. individual taxpayers.

ii. *Content problems*:
    a. *Inclusion of relevant variables*: the survey data is unlikely to hold all relevant variables required in the model, hence the need for auxiliary data sources.
    b. *Variable quality*: the included variables may be inadequately detailed due to inadequate scales of measurement (e.g. income by category, rather than continuously) or inadequate reference periods.

iii. *Data quality problems*:
    a. *Non-responses*: these are obviously troublesome, as they may have been important inclusions.
    b. *Response errors*: these are, too, troublesome, especially as they are often undetectable.

iv. *Dissemination problems*:
    a. Accessibility: this might be hampered by confidentiality concerns, especially where income data is involved.
    b. *Timeliness of release*: this will obviously play a major role in longitudinal and dynamic MSMs.

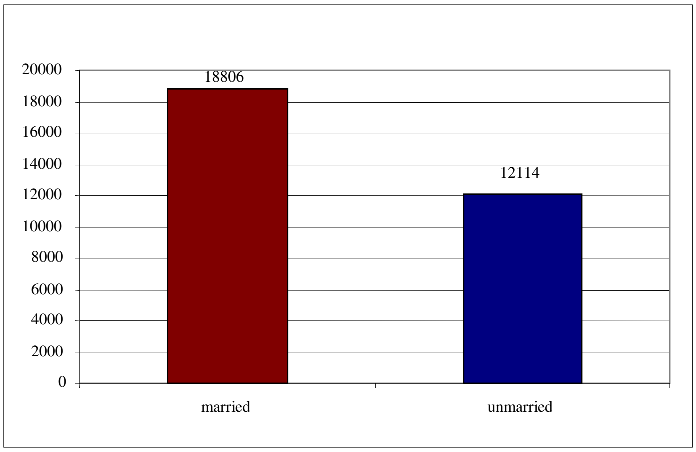
A comprehensive analysis of the 1999 OHS data used in the survey follows.

3.2 Data Analysis

3.2.1 1999 October Household Survey Representativeness

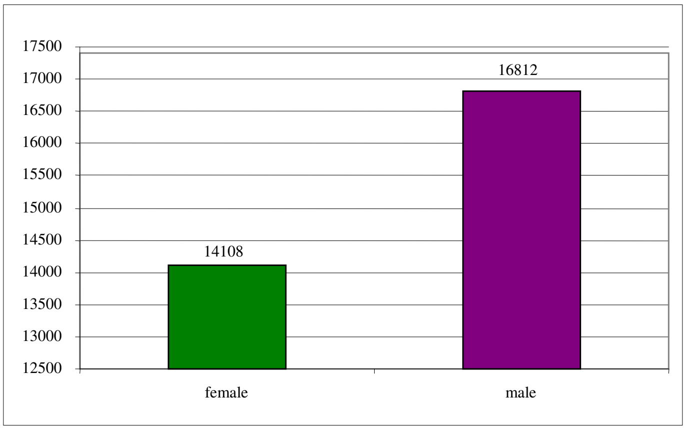
To obtain an indication of the representativeness of the 1999 OHS data, some of the main indicators are analysed graphically in Figures 3 through 10.

Figure 3: Marital status of the respondents of the OHS survey



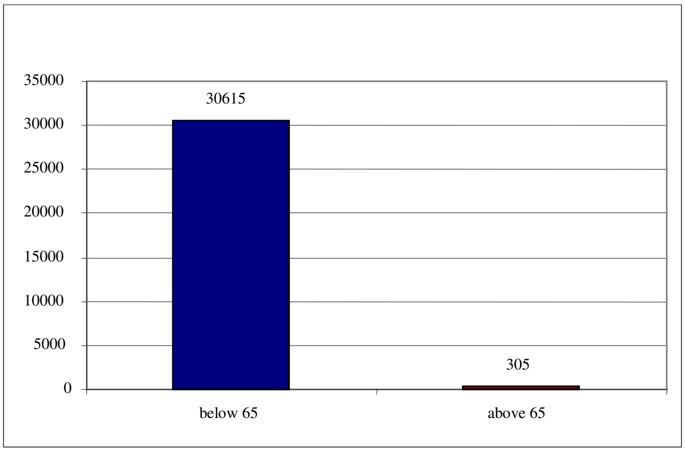
The respondents appear to be fairly evenly split between those who are married, and those who are unmarried, with a slight predilection towards those who are married.

Figure 4: Gender profile of the respondents of the OHS survey



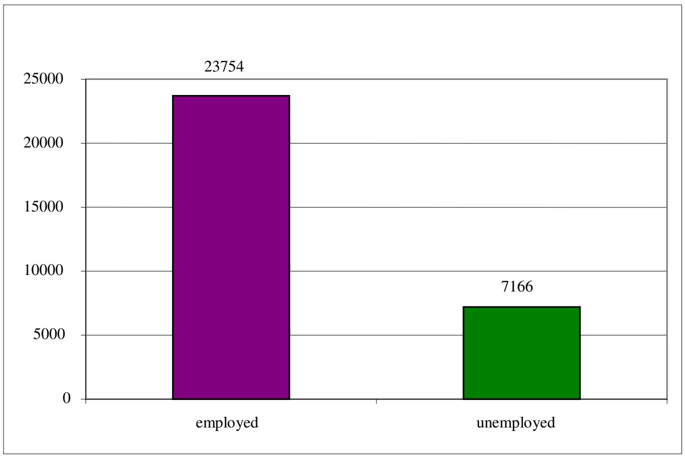
The survey seems to under-represent females, with the overwhelming majority of respondents being male. This is a worrisome fact that should be kept in mind, as it will likely have an effect on the empirical outcomes of the model.

Figure 5: Age profile of the respondents of the OHS survey



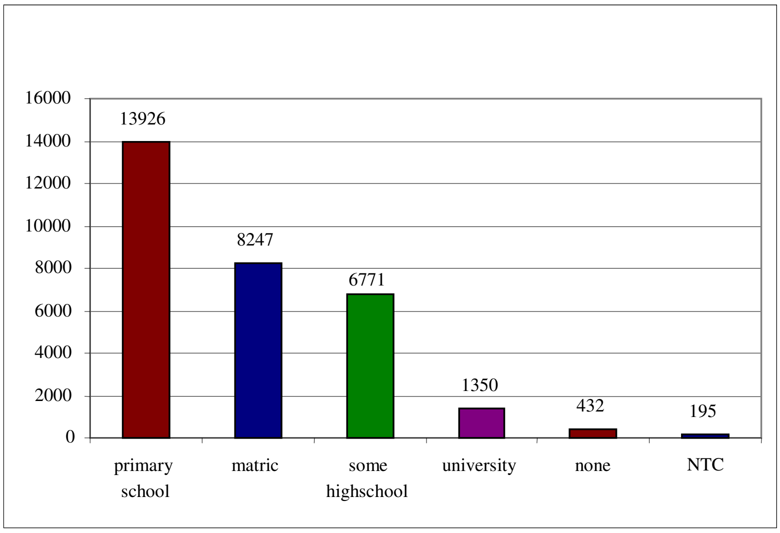
With regards to the age structure of the survey, pensioners are severely under-represented. This will have definite repercussions, should this simple model be extended to incorporate a pension and transfer system in the future.

Figure 6: Employment status of the respondents of the OHS survey



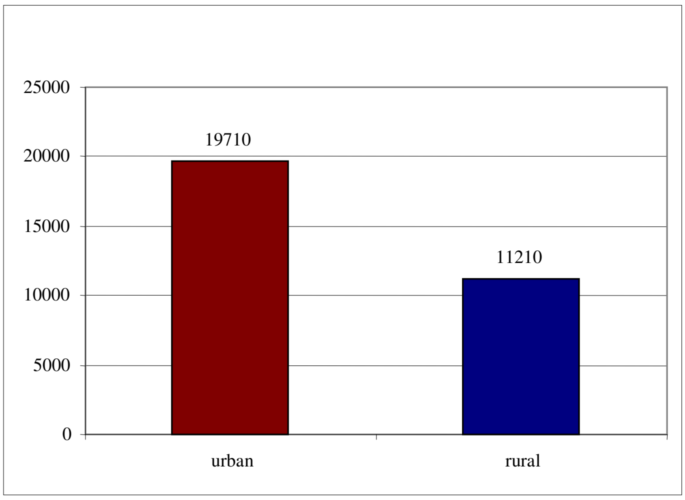
The majority of respondents seem to be employed, with an apparent unemployment rate of approximately 23 per cent – this is obviously in with the official statistics of 23.2 per cent unemployment, as these statistics are derived from the OHS.

Figure 7: Educational qualifications of the respondents of the OHS survey



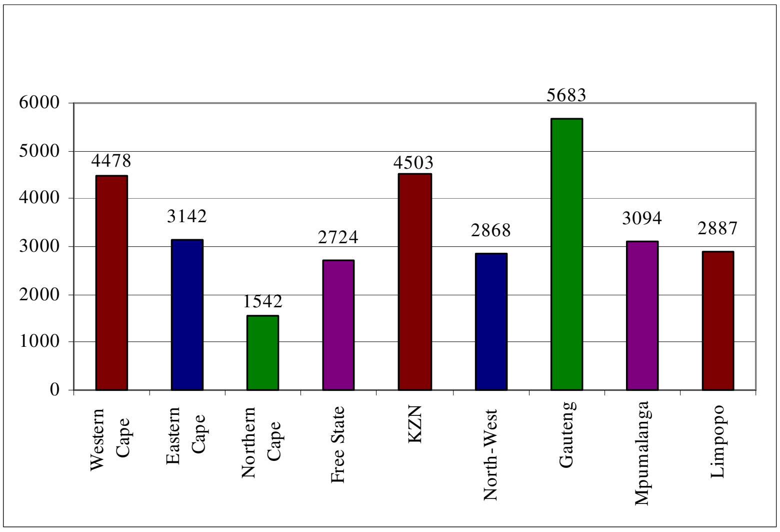
It is clear that the majority of respondents have primary school education only, although very few have none at all. Surprisingly, there are more respondents with a matric qualification than those with a lower level of high school qualification. There are very few respondents with the artisan qualification of an NTC diploma.

Figure 8: Rural-urban status of the respondents of the OHS survey



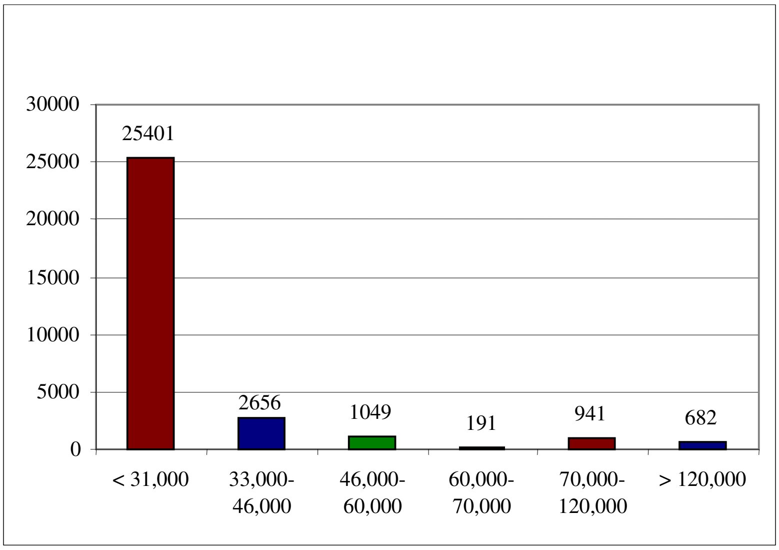
The majority of respondents are urban-dwellers. It is debatable whether this is in line with demographic expectations. This spread is dependent on Statistics South Africa's sampling procedure.

Figure 9: Provincial distribution of the respondents of the OHS survey



The provincial spread of the respondents is very representative of the population sizes of these provinces. Gauteng, the Western Cape and Kwazulu-Natal are the most populous provinces, followed by the Eastern Cape, Mpumalanga and Limpopo respectively. The least populous province is the Northern Cape, followed by the Free State and the North-West respectively. This spread is of course dependent on Statistics South Africa's sampling procedure.

Figure 10: Distribution of respondents' incomes



The income spread of the respondents is relatively uneven. Individuals earning less than R31,000 make up the majority of the respondents, whilst those earning no income number as many as 8,686! The income group containing those individuals earning between R60,000 and R70,000 seems to be grossly misrepresented compared to those groups bordering it. This may have implications for the modelling procedure.

3.2.2 Statistical Analysis of the Data

Due to the wide range of variables available that can be used as explanatory of expenditure, it is necessary to perform some correlation analysis to determine if these variables are in fact correlated with expenditure. Table 3 shows the results of Pearson correlation tests between expenditure and a range of potential explanatory variables. Due to the extremely large sample size in this study, it is quite unlikely that high correlation coefficients will be retrieved. Thus, in deciding to include a regressor or not, it is better to analyse the statistical significance of the correlation coefficient rather than the size of the coefficient. The p-values in Table 3 show the probability of obtaining a larger correlation coefficient, so thus a small p-value indicates a statistically significant and acceptably large correlation coefficient.

Table 3: Pearson correlation coefficients between expenditure and the variables listed

	race	marital status	gender	Education	age_2[5]	age_1[6]
ρ	-0.26495	0.11412	-0.01118	0.49659	0.01387	-0.00393
p-value	<0.0001	<0.0001	0.0956	<0.0001	0.0387	0.5574
	province	*urban/rural location*	*household size*	*Savings*	*disposable income*	
ρ	0.01347	0.23272	-0.22716	0.09534	0.11373	
p-value	0.0446	<0.0001	<0.0001	<0.0001	<0.0001	

From the results in Table 3 it would seem that *age_1* may not be a good predictor of expenditure, and should not be included in the model. Variables that are 'suspicious' are *gender*, *age_2* and *province*. To verify the results obtained in Table 3, partial correlation coefficients are analysed. These are correlation coefficients between two variables taken in isolation, i.e. where the effects of all other variables are removed. The Pearson partial correlation coefficients are shown in Table 4.

[5] *Age_2* is a dummy variable where *age_2 = 1* refers to individuals 65 years old and younger, whilst *age_2 = 0* refers to individuals older than 65 years.

[6] *Age_1* is a continuous variable of each respondent's actual age in years.

Table 4: Partial Pearson correlation coefficients between expenditure and the variables listed

	race	marital status	gender	education	age_2[7]	age_1[8]
ρ_p p-value	-0.15471 <0.0001	0.06072 <0.0001	0.09073 <0.0001	0.20801 <0.0001	-0.01614 <0.0161	-0.02695 <0.0001
	province	*urban/rural location*	*household size*	*savings*	*disposable income*	
ρ_p p-value	-0.05575 <0.0001	-0.00139 0.8357	-0.18150 <0.0001	-0.41055 <0.0001	0.41184 <0.0001	

In contrast to Table 3 the results from Table 4 show that *age_1* and province are in fact correlated with expenditure, whilst *age_2* is less significantly so. Urban/rural location is not correlated with expenditure according to Table 4, which is also in contrast with Table 3. Clearly, the correlation test results are quite conflicting in some cases. For this reason, none of the variables above can convincingly be excluded from a regression against expenditure.

It is of course always important to note that the significant correlation coefficients above do not necessarily imply causality. It is for this reason that certain variables available for this study are not used, since it would be nonsensical to relate them against expenditure, even if they were to have significant correlation coefficients.

Bearing in mind that this is a behavioural study of individual conduct in the face of tax policy changes, the movement of individuals between tax brackets needs to be incorporated into the model. Before this can be done, though, it is important to verify that the variables in question do exhibit significant bias between the tax brackets. Each qualitative variable[9] is analysed along with a tax bracket dummy variable, and a χ^2 test for independence between the two variables is conducted. The results of these analyses are shown in frequency Tables 11 through 17 in Appendix A.

---

7 *Age_2* is a dummy variable where *age_2 = 1* refers to individuals 65 years old and younger, whilst *age_2 = 0* refers to individuals older than 65 years.

8 *Age_1* is a continuous variable of each respondent's actual age in years.

9 The nature of χ^2-test does not allow for the analysis of continuous variables.

The tax bracket variable is in fact a set of dummy variables based upon the 1998/1999 personal income tax policy of the South African Department of Finance (1998:C2). This tax policy is shown in Table 5.

Table 5: 1998/1999 Personal Income Tax Policy

Taxable Income Group	Tax Paid
Less than R31,000	0.19% on every R1
R31,001 - R46,000	R5,890 plus 0.3% on the amount greater than R31,000
R46,001 - R60,000	R10,390 plus 0.39% on the amount greater than R50,000
R60,001 - R70,000	R15,850 plus 0.43% on the amount greater than R60,000
R70,001 - R120,000	R20,150 plus 0.44% on the amount greater than R70,000
More than R120,001	R42,150 plus 0.45% on the amount greater than R120,000
Primary Rebate: 65 years and younger	R3,515
Older than 65 years	R6,175

Source: Department of Finance (1998:C2)

The dummy variable representing these tax brackets is called *duminc*, where *duminc_1 = 1* when earning less than or equal to R31,000 p.a.; *duminc_2 = 1* when earning between R31,001 and R46,000 p.a.; *duminc_3 = 1* when earning between R46,001 and R60,000 p.a.; *duminc_4 = 1* when earning between R60,001 and R70,000 p.a.; *duminc_5 = 1* when earning between R70,001 and R120,000 p.a., and when all are zero, the individual earns in excess of R120,000.

Using the results of the χ^2-tests conducted in Tables 11 through 17 in Appendix A, we can conclude that all the dependent categorical variables other than age can be multiplied by the tax bracket dummy variable so as to include their interaction effects in the model. A description of each of the variables to be used in the MSM is presented in Table 6.

The next section illustrates the various empirical results obtained from the MSM used in this study.

Table 6: Description of data used in regression.

Variable	***Abbreviation***	***Description***
Race & Interaction between race & tax brackets	*race_1, _2, _3* & *drace11 – drace52*	A set of dummy variables describing race as either African (*race_1*), Coloured (*race_2*), Asian/Indian/Other (*race_3*) or White.
Educational qualification & Interaction between education & tax brackets	*educ_1 - _5* & *deduc11 – deduc55*	A set of dummy variables describing qualification obtained as either primary school level (educ_1), high school level (but without matric) (educ_2), matric (educ_3), NTC I, II or III (educ_4), university degree/diploma (educ_5), or none.
Age & Interaction between age & tax brackets	*age* & *dage1 – dage5*	A continuous variable of each respondent's actual age in years, of which the natural logarithm has been taken.
Province & Interaction between province & tax brackets	*prov_1 - _8* & *dprov11 – dprov58*	A set of dummy variables describing the province each respondent lives i.e. the Western Cape (prov_1), Eastern Cape (prov_2), Northern Cape (prov_3), Free State (prov_4), KZN (prov_5), North-West (prov_6), Gauteng (prov_7), Mpumalanga (prov_8) and the Northern Province.
Rural/urban location & Interaction between location & tax brackets	*loc* & *dloc1 – dloc5*	A dummy variable where *loc = 1* means the respondent is from an urban area.
Household size & Interaction between household size & tax brackets	*numhh* & *dnumhh1 – dnumhh5*	A continuous variable detailing the number of people (including children and babies) spending at least four nights per week in the respondent's household, of which the natural logarithm has been taken.
Savings & Interaction between savings & tax brackets	*save* & *dsave1 – dsave5*	A continuous series of the savings of each respondent, of which the natural logarithm has been taken.
Disposable income & Interaction between income & tax brackets	*dispinc* & *ddispinc1 – ddispinc5*	A continuous series calculated as taxable income[10] less personal income tax paid,[11] of which the natural logarithm has been taken.

[10] Calculated as gross income less medical aid and pension fund contributions, which were derived from ratio's obtained from SARS 2003 filer data.

[11] Calculated using the 1998/99 tax ratio's and rebates applied to taxable income (the ratio's can be seen in Table 5).

4 Empirical Results

4.1 Micro-simulation Regression Model Results

The objective of the modelling process is to obtain individual expenditure as a function of disposable income, savings and a range of demographic variables. This section will concentrate on two sets of results: on the actual regression results of the MSM; and on the results of expanding the MSM to macroeconomic levels.

As previously mentioned, the fact that the correlation coefficients between expenditure and various variables are significant does not necessarily imply causality or economic significance. For this reason, a regression is run on all of the variables which are significantly correlated with expenditure, and which seem to intuitively have a reasonable economic relationship with expenditure. Also included in the regression are the effects of income groups (proxied by tax brackets) on all of the independent variables. This is to satisfy the requirement of behavioural micro-simulation modelling, i.e. the movement of individuals between different tax brackets as a result of a policy change is included in the model. As shown in Appendix A the interactions between tax brackets and all of the independent variables are significant. These interactions are included by multiplying each regressor by a set of dummy variables encompassing the 1998/99 tax brackets laid out by the Department of Finance (1998:C2) shown in Table 5.[12]

As previously mentioned, there are some modifications that need to be made to the data obtained from the OHS. Firstly, those individuals who stated that their annual incomes are zero are excluded from the model. This is because, even though they do make expenditures, they do not pay income tax, and thus their behaviour before and after an income tax policy change will be the same. Secondly, the natural logarithm of all continuous data is taken.

The final results of the regression are shown in Table 18 in Appendix B.

---

[12] It should be noted that the correlation that exists between disposable income and the tax bracket dummies is the reason for not including the dummy variables themselves as a separate regressor.

The regression results of Table 18 are remarkably robust when evaluated economically and statistically (an adjusted R^2 value of 65 per cent for survey data is especially satisfactory). Although there are many coefficients that are regarded as conventionally statistically insignificant, they are included either because they are but a few of many of the same group of dummy variables, or they too important economically to exclude from the model, such as province.

Merz (1991:98) admits that the validation of micro-simulation results is a very demanding task, which is not done often in the available literature. It is not correct, by principle, to compare the impacts of social policy to the initial database used in the creation of the database. However, this is usually done in practice, and is what we will do in this study. The coefficients of all the variables are now briefly discussed. It is important to bear in mind when analysing the interaction variables that the lay out of the income bracket dummy variable is as follows:

duminc_1 = 1 when less than R31,000
duminc_2 = 1 when between R31,001 and R46,000
duminc_3 = 1 when between R46,001 and R60,000
duminc_4 = 1 when between R60,001 and R70,000
duminc_5 = 1 when between R70,001 and R120,000
all = *0* when more than R120,000

Due to the fact that the race dummy variable has Whites as the baseline. The coefficients are then analysed in relation to this. Thus, Africans spend less than Whites, and they spend less than Coloureds and Asians/Indians/other respectively. In relation to Whites that earn more than R120,000, expenditure by

Since the benchmark for the education dummy variable is that the individual has no educational qualification, which improvements from primary school education in *educ_1* through to university education in *educ_5*, the coefficients are interpreted relative to this. All levels of education result in higher expenditure than those individuals who have none – an economically sound result. Also a significant *a priori* result is the fact that this expenditure increases with higher levels of education.

A positive coefficient on the age variable is reasonable, since individuals' incomes tend to increase with age, and so thus their expenditures will increase with age. This is not an illogical result, since there are only 245 individuals out of 22,234 in the sample who are older than 65 years (it is expected that a negative age coefficient would apply only in the case of pensioners).

Although many of the province variables are not statistically significant, it was decided that they should remain in the model for purely economic reasons, and the coefficients prove their value! The baseline province in this case is Limpopo, and all others are analysed in relation to this. Gauteng (prov_7) has the largest expenditure per individual higher than Limpopo, followed by the Western Cape and then the Northern Cape (the latter being a slightly puzzling result). Kwazulu-Natal and the North-West have the next highest expenditures per individual above Limpopo respectively, trailed by the Eastern Cape and Mpumalanga respectively. The only province to have lower individual expenditure levels than Limpopo is the Free State.

The location dummy variable is given the value of 1 if the respondent lives in an urban area, and zero if they live in a rural area. Thus, the baseline is urban areas. A positive coefficient on this variable implies that people living in urban areas spend more than in rural areas, correct according to *a priori* expectations.

Above completes the analysis of the qualitative independent variables. The results of the continuous independent variables are discussed next. The negative coefficient on the variable pertaining to the number of people in a household can be explained rationally: the larger the size of the household, the less an individual in that household will have to spend him/herself.

The negative coefficient on savings is of course as expected: the more an individual spends, the less income he/she has remaining to save.

The positive coefficient on disposable income is of course also correct according to economic theory: the more income available, the higher the expenditure.

The next section will discuss the process and results of expanding the microeconomic results of the regression to macroeconomic levels.

4.2 Macroeconomic Results

Apart from the estimation of individual expenditures, the MSM in this model is used to project individual data for disposable income, taxable income and expenditure to a macroeconomic level. The national results of these projections for the year of 1999 (this being the year the OHS was conducted) compared to the actual figures reported in the South African Reserve Bank Quarterly Bulletin (SARB QB) are shown in Table 7.

Table 7: Macroeconomic results of MSM

Variable	1999 – SARB QB	1999 – MSM
Taxable income	N/A	R 437, 711, 919, 450
Disposable income[13]	R 521, 149, 000, 000	R 320, 862, 665, 410
Private consumption expenditure[14]	R 581, 101, 000, 000	R 340, 510, 314, 060

It is evident that the MSM has underestimated the figures of disposable income and private consumption expenditure. Unfortunately it is not possible to find national data on taxable income against which the MSM results can be compared. Whether looking at published national figures, or looking at the MSM results, it is evident that South Africans are over-extending themselves, and spending more than what their income allows them to.

It is also possible, using the MSM, to present the above data on taxable income, disposable income and private expenditure stratified according to income group[15] and to province. These results are shown in Tables 8 through 10.

13 SARB QB code 6246.
14 SARB QB code 6235.
15 The income groups used will be the same as those used by SARS.

Table 8: MSM results of taxable income according to income group and province

Income group:	*< 0*	*R 1 – 20,000*	*R 20,001 – 30,000*	*R 30,001 – 40,000*
Western Cape (WC)	-	7,805,504,415	5,625,803,178	3,678,937,255
Eastern Cape (EC)	-	3,472,281,320	2,358,892,437	2,247,412,822
Northern Cape (NC)	-	975,461,239	638,648,234	515,786,277
Free State (FS)	-	3,706,290,072	1,954,997,644	994,491,114
Kwazulu-Natal (KZN)	-	8,482,711,344	5,515,268,898	3,847,498,924
North-West (NW)	-	3,932,797,632	2,669,203,704	1,706,351,774
Gauteng (GP)	-	10,530,717,734	10,842,199,207	8,528,315,495
Mpumalanga (MP)	-	3,694,501,336	1,858,009,203	1,287,737,048
Northern Province (NP)	-	3,008,109,278	2,214,626,783	2,135,516,967
Total (RSA)	**-**	45,608,374,371	33,677,649,289	24,942,047,676

	R 40,001 – 50,000	*R 50,001 – 60,000*	*R 60,001 –70,000*	*R 70,001 – 80,000*	*R 80,001 – 90,000*
WC	3,269,815,183	4,055,702,102	3,898,875,665	3,550,231,089	2,994,022,093
EC	2,201,034,837	2,115,285,594	2,044,137,665	2,063,644,201	1,952,670,677
NC	514,843,098	519,474,235	467,793,729	439,613,442	391,900,557
FS	1,282,780,419	1,641,019,070	1,717,808,742	1,319,926,020	1,121,642,511
KZN	4,217,085,809	3,487,046,308	3,709,086,410	3,255,842,365	1,954,633,054
NW	1,262,188,455	1,467,396,633	1,444,416,267	1,070,006,553	876,938,750
GP	8,493,562,559	8,908,024,513	8,077,858,707	6,995,787,139	6,183,012,307
MP	1,204,967,718	1,530,920,503	1,276,797,843	1,169,292,413	967,817,113
NP	2,454,071,255	2,450,638,363	1,804,078,727	1,370,004,701	900,564,804
RSA	24,900,349,334	26,175,507,320	24,440,853,755	21,234,347,922	17,343,201,866

	R 90,001 – 100,000	*R 100,001 – 110,000*	*R 110,001 – 120,000*	*R 120,001 – 130,000*
WC	477,425,364	2,760,691,564	2,758,039,186	186,581,926
EC	123,431,936	808,605,850	680,889,131	108,258,767
NC	68,738,567	529,454,281	534,126,811	21,993,845
FS	196,041,190	501,647,856	634,835,917	43,313,059
KZN	745,235,012	1,227,527,458	1,369,149,439	205,721,179
NW	290,226,116	365,964,159	376,357,275	251,937,728
GP	1,279,194,545	5,950,237,422	6,681,400,991	356,298,035
MP	474,741,135	669,335,251	732,847,605	219,395,608
NP	412,578,355	754,374,986	614,834,275	157,332,353
RSA	4,067,612,220	13,567,838,827	14,382,480,630	1,550,832,500

	R 130,001 – 140,000	*R 140,001 – 150,000*	*R 150,001 – 200,000*	*R 200,001 – 300,000*
WC	106,908,150	54,586,061	3,104,763,595	4,494,770,683
EC	222,135,645	169,063,685	1,327,169,049	1,183,943,833
NC	77,253,124	78,934,115	446,506,536	596,535,856
FS	96,629,626	102,181,445	908,705,106	1,032,101,246
KZN	213,382,758	75,147,811	2,626,222,605	1,585,120,350
NW	271,954,518	141,872,634	1,011,816,473	954,845,336
GP	548,971,796	566,573,224	8,304,220,851	9,600,166,864
MP	242,853,321	211,413,861	861,952,749	1,250,616,715
NP	252,815,292	155,346,243	1,229,213,280	451,287,342
RSA	2,032,904,231	1,555,119,079	19,820,570,245	21,149,388,224

	R 300,001 – 400,000	*R 400,001 – 500,000*	*R 500,001 – 750,000*	*R 750,001 – 1,000,000*
WC	2,469,258,599	192,945,179	477,515,882	1,606,380,022
EC	1,105,042,702	180,094,301	262,185,939	707,656,792
NC	379,217,902	77,459,219	182,015,889	167,386,950
FS	357,574,779	443,725,151	188,372,926	605,079,594
KZN	1,479,796,139	223,124,776	1,302,567,265	1,298,666,396
NW	547,568,109	164,524,494	425,271,354	527,546,358
GP	10,159,044,442	279,411,325	644,674,000	2,596,352,345
MP	942,764,292	282,524,022	823,070,566	526,954,383
NP	262,578,372	527,137,806	508,409,348	687,424,021
RSA	17,702,845,336	2,370,946,272	4,814,083,168	8,723,446,862

	R 1,000,001 – 2,000,000	*R 2,000,001 – 5,000,000*	*> R 5,000,000*	***Total***
WC	727,184,143	4,612,918,699	20,048,575,819	78,957,435,851
EC	1,621,085,298	1,061,674,252	6,766,434,064	34,783,030,796
NC	958,994,004	1,161,321,412	2,353,193,562	12,096,652,884
FS	1,000,977,468	1,073,050,750	8,817,923,469	29,741,115,175
KZN	2,974,957,670	1,619,262,771	12,417,517,413	63,832,572,153
NW	1,541,853,914	862,503,386	5,352,832,268	27,516,373,892
GP	2,035,574,268	5,889,599,568	4,165,755,608	127,616,952,945
MP	2,116,834,003	2,029,634,175	5,004,263,380	29,379,244,244
NP	444,753,267	4,419,872,233	6,572,973,459	33,788,541,511
RSA	13,422,214,036	22,729,837,247	71,499,469,041	**437,711,919,450**

Evident in Table 8 is the fact that the income group of those earning more than R5,000,000 per annum has the highest taxable income (R 71,499,469,041). Gauteng is the province with the highest taxable income (R 127,616,952,945).

Table 9: MSM results of disposable income according to income group and province

Income group:	*< 0*	*R 1 – 20,000*	*R 20,001 – 30,000*	*R 30,001 – 40,000*
Western Cape (WC)	-	7,805,504,415	5,625,803,178	2,824,362,923
Eastern Cape (EC)	-	3,472,281,320	2,358,892,437	1,718,899,170
Northern Cape (NC)	-	975,461,239	638,648,234	397,048,425
Free State (FS)	-	3,706,290,072	1,954,997,644	765,492,663
Kwazulu-Natal (KZN)	-	8,482,711,344	5,515,268,898	3,027,684,287
North-West (NW)	-	3,932,797,632	2,669,203,704	1,322,251,475
Gauteng (GP)	-	10,530,717,734	10,842,199,207	6,564,884,388
Mpumalanga (MP)	-	3,694,501,336	1,858,009,203	992,103,416
Northern Province (NP)	-	3,008,109,278	2,214,626,783	1,647,372,527
Total (RSA)	-	45,608,374,371	33,677,649,289	19,260,099,275

	R 40,001 – 50,000	*R 50,001 – 60,000*	*R 60,001 –70,000*	*R 70,001 – 80,000*	*R 80,001 – 90,000*
WC	2,444,211,476	2,997,499,937	2,901,830,325	2,696,353,835	2,262,011,281
EC	1,645,833,354	1,563,043,573	1,521,534,733	1,564,224,120	1,475,856,307
NC	385,069,169	383,874,851	347,906,532	333,774,729	295,884,018
FS	959,038,702	1,213,388,144	1,290,085,991	1,004,578,560	846,490,532
KZN	3,156,512,909	2,578,214,244	2,787,508,904	2,485,347,503	1,476,645,184
NW	943,764,011	1,085,397,531	1,085,347,554	814,529,778	661,299,330
GP	6,350,973,712	6,583,105,421	6,012,565,919	5,316,567,948	4,665,811,677
MP	901,526,603	1,132,441,974	957,838,803	890,372,050	728,806,343
NP	1,835,846,266	1,813,700,573	1,358,041,598	1,046,262,514	678,903,971
RSA	18,622,776,203	19,350,666,249	18,262,660,359	16,152,011,037	13,091,708,641

	R 90,001 – 100,000	*R 100,001 – 110,000*	*R 110,001 – 120,000*	*R 120,001 – 130,000*
WC	346,065,856	1,961,382,527	1,955,236,444	130,582,919
EC	89,558,003	575,124,207	482,601,319	75,426,491
NC	49,991,359	376,036,007	378,792,800	15,424,752
FS	142,657,260	356,590,986	449,304,031	30,226,632
KZN	541,700,004	872,042,310	970,071,610	142,541,597
NW	210,182,888	260,208,085	266,450,555	175,283,117
GP	929,296,501	4,226,333,656	4,733,959,548	247,926,934
MP	344,280,414	476,494,307	518,366,123	152,894,146
NP	299,690,838	537,213,181	435,676,819	97,498,394
RSA	2,953,423,123	9,641,425,265	10,190,459,249	1,067,804,983

	R 130,001 – 140,000	*R 140,001 – 150,000*	*R 150,001 – 200,000*	*R 200,001 – 300,000*
WC	73,325,600	37,330,672	2,088,055,305	2,905,992,833
EC	141,282,994	109,796,122	892,791,017	765,075,640
NC	45,063,137	53,916,280	297,716,231	385,382,974
FS	66,355,334	69,630,907	610,886,852	667,437,718
KZN	147,061,928	51,392,575	1,761,132,347	1,021,873,867
NW	187,093,024	97,024,783	676,891,293	618,303,429
GP	376,688,890	387,073,733	5,585,832,204	6,209,802,409
MP	166,734,197	144,276,459	578,904,333	808,666,230
NP	173,469,221	106,239,203	822,468,631	294,731,015
RSA	1,377,074,326	1,056,680,733	13,314,678,214	13,677,266,113

	R 300,001 – 400,000	*R 400,001 – 500,000*	*R 500,001 – 750,000*	*R 750,001 – 1,000,000*
WC	1,557,343,593	118,862,293	290,867,390	857,413,798
EC	695,796,519	112,293,516	159,906,666	387,955,522
NC	239,879,995	48,248,616	84,745,620	100,406,699
FS	226,411,711	276,378,926	115,595,743	326,095,096
KZN	934,870,483	138,985,960	793,783,375	771,460,059
NW	347,713,035	101,448,913	258,706,798	318,026,300
GP	6,403,956,307	172,292,779	394,379,341	1,505,561,176
MP	594,534,559	174,470,881	498,594,826	316,240,491
NP	165,489,198	325,530,541	256,400,805	357,657,890
RSA	11,165,995,399	1,468,512,425	2,852,980,564	4,940,817,032

	R 1,000,001 – 2,000,000	R 2,000,001 – 5,000,000	> R 5,000,000	**Total**
WC	428,720,892	2,696,355,113	11,649,036,242	56,654,148,848
EC	902,420,912	623,609,888	3,928,518,911	25,262,722,743
NC	569,505,597	680,210,956	1,223,330,148	8,306,318,367
FS	593,245,824	628,708,881	5,133,413,951	21,433,302,163
KZN	1,789,837,516	948,745,997	7,737,627,174	48,133,020,075
NW	912,349,389	506,276,670	3,351,292,917	20,801,842,210
GP	1,211,335,707	3,456,904,592	2,426,372,463	95,134,542,246
MP	1,254,281,094	1,070,836,906	2,913,079,709	21,168,254,405
NP	264,553,766	2,583,316,509	3,645,714,833	23,968,514,354
RSA	7,926,250,698	13,194,965,514	42,008,386,349	**320,862,665,410**

From Table 9 it is found that the income group with the highest disposable income is those earning R1 to R20,000 (R 45,608,374,371), and again, Gauteng has the highest taxable income (R 95,134,542,246). Although the highest income bracket had the highest taxable income, they didn't have the highest disposable income, obviously due to the fact that these individuals pay large amounts of income tax (actually totalling R 29,491,082,692 according to the MSM), whilst the lowest income fall below the tax threshold and pay nothing.

Table 10: MSM results of consumption expenditure according to income group and province

Income group:	*< 0*	*R 1 – 20,000*	*R 20,001 – 30,000*	*R 30,001 – 40,000*
Western Cape (WC)	-	14,059,107,942	6,511,832,058	7,573,386,779
Eastern Cape (EC)	-	6,448,891,553	2,394,288,063	4,700,252,002
Northern Cape (NC)	-	1,845,083,491	703,727,427	981,087,885
Free State (FS)	-	7,995,167,461	2,725,692,906	2,214,912,205
Kwazulu-Natal (KZN)	-	15,541,884,538	5,497,928,731	5,533,061,620
North-West (NW)	-	5,983,044,219	2,527,917,737	2,226,168,589
Gauteng (GP)	-	17,484,343,710	11,236,776,325	12,937,189,340
Mpumalanga (MP)	-	6,004,234,815	1,786,821,284	1,604,150,117
Northern Province (NP)	-	4,903,873,443	1,984,274,647	2,728,083,890
Total (RSA)	-	80,265,631,172	35,369,259,179	40,498,292,427

	R 40,001 – 50,000	*R 50,001 – 60,000*	*R 60,001 –70,000*	*R 70,001 – 80,000*	*R 80,001 – 90,000*
WC	3,543,520,827	1,829,122,720	5,113,809,513	904,017,556	4,798,629,161
EC	1,957,114,192	557,572,332	2,174,369,534	131,292,987	2,715,485,107
NC	507,746,861	196,473,719	417,100,284	151,873,280	463,314,977
FS	1,313,321,236	992,160,841	1,946,461,321	375,224,159	1,433,642,549
KZN	3,305,578,761	1,497,672,890	2,871,382,514	2,223,166,752	2,848,339,500
NW	1,026,952,982	785,852,352	967,264,543	397,297,936	744,788,704
GP	7,609,209,093	4,243,491,531	7,510,933,311	2,874,601,609	8,748,306,409
MP	877,965,678	1,012,454,273	890,538,698	426,831,150	1,407,253,794
NP	1,598,183,140	1,704,628,038	892,833,359	956,463,553	767,843,532
RSA	21,739,592,770	12,819,428,696	22,784,693,076	8,440,768,982	23,927,603,732

	R 90,001 – 100,000	*R 100,001 – 110,000*	*R 110,001 – 120,000*	*R 120,001 – 130,000*
WC	242,955,711	259,012,256	1,968,878,800	63,852,938
EC	31,623,396	100,266,877	363,273,240	25,231,139
NC	43,551,222	18,668,422	300,661,703	10,179,075
FS	139,202,919	59,098,680	398,287,744	33,716,371
KZN	583,806,534	103,846,677	819,877,030	97,543,490
NW	104,300,926	21,532,282	121,041,630	49,129,244
GP	930,359,340	323,898,131	4,391,497,419	151,250,991
MP	216,926,480	121,996,608	449,129,925	116,051,725
NP	142,989,801	83,708,211	340,293,199	35,866,501
RSA	2,331,415,403	1,092,028,144	9,152,940,689	546,954,973

	R 130,001 – 140,000	*R 140,001 – 150,000*	*R 150,001 – 200,000*	*R 200,001 – 300,000*
WC	6,414,606,078	5,457,358	6,314,277,484	2,074,268,869
EC	1,589,803,566	90,988,073	1,589,803,566	370,882,358
NC	689,670,390	94,707,848	594,962,542	156,290,166
FS	1,189,671,966	33,588,074	1,123,477,587	301,844,557
KZN	2,755,520,367	15,412,807	2,667,243,722	229,381,119
NW	820,931,676	61,547,811	650,169,669	125,020,445
GP	16,678,174,198	559,692,389	15,849,135,087	4,475,275,202
MP	1,684,846,388	68,976,504	1,525,917,759	418,679,877
NP	1,159,107,421	102,701,512	965,419,145	80,786,339
RSA	32,982,332,052	942,084,304	31,280,406,561	8,232,428,933

	R 300,001 – 400,000	*R 400,001 – 500,000*	*R 500,001 – 750,000*	*R 750,001 – 1,000,000*
WC	838,276,479	38,895,530	36,710,318	20,191,809
EC	175,229,482	9,917,026	13,544,126	8,153,694
NC	48,791,373	1,190,077	8,499,219	3,356,886
FS	19,163,058	8,739,370	11,378,056	7,243,972
KZN	374,639,301	23,275,164	62,234,710	15,185,809
NW	65,373,516	5,161,837	32,683,703	13,427,642
GP	3,198,072,115	29,455,298	179,868,062	38,593,622
MP	158,186,623	16,275,800	77,352,111	31,145,237
NP	27,510,058	104,249,827	21,731,723	6,018,918
RSA	4,905,242,005	237,159,929	413,771,084	47,929,765

	R 1,000,001 – 2,000,000	*R 2,000,001 – 5,000,000*	*> R 5,000,000*	**Total**
WC	15,796,525	67,070,276	117,617,388	62,811,294,377
EC	83,137,132	6,957,463	35,592,671	25,573,669,580
NC	76,206,204	15,610,909	10,275,654	7,339,029,614
FS	52,647,334	9,613,643	149,795,467	22,534,051,478
KZN	154,838,358	171,989,553	66,289,401	47,460,099,345
NW	27,735,845	5,795,122	23,384,494	16,786,522,905
GP	114,015,020	231,443,069	258,807,144	120,054,388,412
MP	179,948,000	31,112,491	37,352,833	19,144,148,169
NP	83,863,113	90,406,906	26,273,905	18,807,110,181
RSA	550,212,041	598,886,941	563,572,832	**340,510,314,060**

Of particular interest in Table 10 is the fact that the income with the largest expenditure amount (R 80,265,631,172) is the group of people earning between R1 and R20,000 per annum – the most impoverished people who earn between nothing and the minimum wage. The reason why this figure makes sense, though, is because this group of people make up a large proportion of the population – South Africa's Gini coefficient in 1995 was 59.33 (World Bank), indicating high inequality in the income distribution, with a large percentage of the population earning a small percentage of the national income.

According to Table 10, the province with the largest expenditure amount (R 120,054,388,412) is Gauteng. This is obvious, as Gauteng is often regarded as the 'economic powerhouse' of South Africa.

The next section will state final conclusions and policy recommendations.

5 Conclusion and Policy Analysis

The purpose of this study was to develop a static micro-simulation model (MSM) of personal income tax in South Africa. MSMs are useful tools for analysing the distributional effects of changes in policy variables on the population, particularly in the field of fiscal policies.

This micro-simulation model estimates individual expenditures using individual survey data from the 1999 OHS. The primary data of interest in the OHS survey are gross income data pertaining to each individual, Y. Each individual's taxable income is then deduced using ratio's of medical aid and pension contributions from SARS IRP5 data. The personal income tax paid by each individual is then calculated by applying the South African Department of Finance's 1999 tax policy to individual taxable incomes. Individual disposable incomes are then calculated by subtracting the personal income tax paid from the calculated taxable incomes. Using the calculated disposable income, savings data, and a variety of demographic indicators, individual expenditure data are then inferred using regression techniques.

The MSM created in this study also serves the purpose of analysing the tax gap between actual taxes received by SARS and potential taxes calculated. This was done by projecting the microeconomic MSM results to macroeconomic figures, and comparing actual data for variables such as taxable income, disposable income and consumption expenditure, with the estimated figures using the MSM results. The MSM results were found to be an underestimation of national data. It is also possible to use the MSM results to stratify consumption expenditure, disposable income and taxable income according to province and income group. Of particular interest in these results is the fact that the group of lowest income earners (R1 to R20,000) have the largest disposable income and spend the largest amount. The highest income earners however have the highest taxable income.

The results of this study can be used widely to compare published figures with the figures in this study derived from survey data – arguably more reliable than national figures?

References

Brown, L. and Harding, A. 2002. Social modelling and public policy: application of microsimulation modelling in Australia. *Journal of Artificial Societies and Social Stimulation*, 5(4). [Online] Available from: http://jasss.soc.surrey.ac.uk/5/4/6.html [Accessed: 2005-10-28].

Creedy, J., Duncan, A.S., Harris, M. and Scutella, R. 2002. *Microsimulation modelling of taxation and the labour market: the Melbourne Institute tax and transfer simulator*. Cheltenham, UK: Edward Elgar Publishing.

Klevmarken, N.A. 1997. Behavioural modeling in micro simulation models. A survey. Unpublished.

Martini, A. and Trivellato, U. 1997. The role of survey data in microsimulation models for social policy analysis. *Labour*, 11(1)83-112.

Merz, J. 1991. Microsimulation – a survey of principles, developments and applications. *International Journal of Forecasting*, 7(77-104).

O'Donoghue, C. 2001. Introduction to the special issue on dynamic micro-simulation modelling. *Brazilian Electronic Journal of Economics*, 4(2). [Online] Available from: http://www.beje.decon.ufpe.br/v4n2/intro.htm [Accessed: 2005-07-26].

Republic of South Africa. Department of Finance. 1998. *Budget Review*. Pretoria: Government Printer.

South African Reserve Bank. *Quarterly Bulletin*, various issues. Pretoria: South African Reserve Bank.

Statistics South Africa. 1999. *October Household Survey 1999: Metadata General notes*. [Online] Available from:http://www.statssa.gov.za/publications/publicationsearch.asp?PN=fqtrx&PM=&PY=rzzz&PS=1.

World Bank. Various years. World Bank Development Indicators.

Appendix A: χ^2 Test for Independence

The null hypothesis in this case is that the variable in question is independent from tax brackets, i.e. the variable does not demonstrate trends that are caused by being within a specific tax bracket.

The format of the individual cells in the frequency table is as follows: the top figure is the frequency of individuals in that tax bracket holding that certain characteristic (e.g. white or married or female etc.); the second figure is the corresponding percentage for those individuals; the third figure shows the row frequency (e.g. in Table 12 of those individuals who are married, 70.91% earn less than R31,000); and the fourth figure in each cell shows the column percentage (e.g. in Table 12 of those individuals who earn less than R31,000, 64.82% are married).

Table 11: Frequency table and χ^2 test of race against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
White	744 3.35 26.22 4.45	694 3.12 24.45 26.13	421 1.89 14.83 40.13	62 0.28 2.18 32.46	502 2.26 17.69 53.35	415 1.87 14.62 60.85	2,838 12.76
African	12,871 57.89 84.09 77.00	1,433 6.45 9.36 53.95	436 1.96 2.85 41.56	98 0.44 0.64 51.31	28 1.27 1.84 29.97	186 0.84 1.22 27.27	15,306 68.84
Coloured	2,777 12.49 80.19 16.61	399 1.79 11.52 15.02	129 0.58 3.73 12.30	20 0.09 0.58 10.47	99 0.45 2.86 10.52	39 0.18 6.51 5.72	3,463 15.58
Indian/Asian & other	323 1.45 51.52 1.93	130 0.58 20.73 4.89	63 0.28 10.05 6.01	11 0.05 1.75 5.76	58 0.26 9.25 6.16	42 0.19 6.70 6.16	627 2.82
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (20)	= 5422.52						
p-value	< 0.0001						

Considering the null hypothesis that race is independent of tax brackets, the χ^2-test above conclusively shows that there is in fact a significant statistical matching between race and income distribution.

Table 12: Frequency table and χ^2 test of marital status against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
Unmarried	5,880 26.45 84.54 35.18	620 2.79 8.91 23.34	188 0.85 2.70 17.92	41 0.18 0.59 21.47	137 0.62 1.97 14.56	89 0.40 1.28 13.05	6,955 31.28
Married	10,835 48.73 70.91 64.82	2,036 9.16 13.33 76.66	861 3.87 5.64 82.08	150 0.67 0.98 78.53	804 3.62 5.26 85.44	593 2.67 3.88 86.95	15,279 68.72
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (5)	= 519.449						
p-value	< 0.0001						

Under the null hypothesis that marital status is independent of tax brackets, the statistically significant χ^2-statistic above shows that marital status does have trends dependent on the income distribution of the respondents.

Table 13: Frequency table and χ^2 test of gender against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
Male	9,196 41.36 72.57 55.02	1,553 6.98 12.26 58.47	601 2.70 4.74 57.29	122 0.55 0.96 63.87	646 2.91 5.10 68.65	554 2.49 4.37 81.23	12,672 56.99
Female	7,519 33.82 78.63 4.98	1,103 4.96 11.54 41.53	448 2.01 4.69 42.71	69 0.31 0.72 36.13	295 1.33 3.09 31.35	128 0.58 1.34 18.77	9,562 43.01
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (5)	= 248.382						
p-value	< 0.0001						

The null hypothesis that gender is independent of tax brackets is rejected by the significant χ^2-test above, which implies that there is in fact a significant statistical matching between gender and income distribution.

Table 14: Frequency table and χ^2 test of educational qualification against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
None	281 1.26 87.54 1.68	23 0.1 7.17 0.87	3 0.01 0.93 0.29	0 0 0 0	5 0.02 1.56 0.53	9 0.04 2.8 1.32	321 1.44
Primary school	9,742 43.82 93.51 58.28	40 1.89 4.03 15.81	91 0.41 0.87 8.67	16 0.07 0.15 8.38	68 0.31 0.65 7.23	81 0.36 0.78 11.88	10,418 46.86
High school (not matric)	3,599 16.19 81.63 21.53	515 2.32 11.68 19.39	122 0.55 2.77 11.63	22 0.1 0.5 11.52	85 0.38 1.93 9.03	66 0.3 1.5 9.68	4,409 19.83
Matric	2,866 12.89 49.65 17.15	1,408 6.33 24.39 53.01	631 2.84 10.93 60.15	109 0.49 1.89 57.07	469 2.11 8.13 49.84	289 1.3 5.01 42.38	5,772 25.96
NTC diploma	39 0.18 25.49 0.23	50 0.22 32.68 1.88	22 0.1 14.38 2.1	6 0.03 3.92 3.14	28 0.13 18.3 2.98	8 0.04 5.23 1.17	153 0.69
University degree / diploma	188 0.85 16.19 1.12	240 1.08 20.67 9.04	180 0.81 15.5 17.16	38 0.17 3.27 19.9	286 1.29 24.63 30.39	229 1.03 19.72 33.58	1,161 5.22
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (25)	= 7545.06						
p-value	< 0.0001						

Given the statistically significant χ^2-statistic above, it is possible to reject the null hypothesis that educational qualification is independent of tax brackets, in other words educational qualification exhibits trends according to income distribution.

Table 15: Frequency table and χ^2 test of age structure against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
Older than 65 years	198 0.89 80.82 1.18	21 0.09 8.57 0.79	12 0.05 4.9 1.14	1 0 0.41 0.52	5 0.02 2.04 0.53	8 0.04 3.27 1.17	245 1.1
65 years and younger	16,517 74.29 75.11 98.82	2,635 11.85 11.98 99.21	1,037 4.66 4.72 98.86	190 0.85 0.86 99.48	936 4.21 4.26 99.47	674 3.03 3.07 98.83	21,989 98.9
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (5)	= 6.8548						
p-value	= 0.2317						

Testing the null hypothesis that age group (either retirement age or below) is independent of tax brackets, the χ^2-statistic above shows that there is no significant statistical relationship between age group and income distribution – age group does not exhibit trends that are caused by being situated in a specific income group.

Table 16: Frequency table and χ^2 test of urban/rural location against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
Rural	7,180 32.29 88.62 42.96	471 2.12 5.81 17.73	174 0.78 2.15 16.59	41 0.18 0.51 21.47	99 0.45 1.22 10.52	137 0.62 1.69 20.09	8,102 36.44
Urban	9,535 42.88 67.47 57.04	2,185 9.83 15.46 82.27	875 3.94 6.19 83.41	150 0.67 1.06 78.53	842 3.79 5.96 89.48	545 2.45 3.86 79.91	14,132 63.56
Total	16,715 75.18	2,656 11.95	1,049 4.72	191 0.86	941 4.23	682 3.07	22,234 100
χ^2 (5)	= 1256.33						
p-value	< 0.0001						

Under the null hypothesis that urban/rural location is independent of tax brackets, the statistically significant χ^2-statistic above shows that there is in fact dependence of location on income distribution trends.

Table 17: Frequency table and χ^2 test of province against tax brackets

	Less than R31,000	R31,001 – R46,000	R46,001 – R60,000	R60,001 – R70,000	R70,001 – R120,000	More than R120,000	***Total***
Limpopo	1,389	279	118	26	66	39	1,917
	6.25	1.25	0.53	0.12	0.3	0.18	8.62
	72.46	14.55	6.16	1.36	3.44	2.03	
	8.31	10.5	11.25	13.61	7.01	5.72	
Western Cape	2,748	428	176	22	185	124	3,683
	12.36	1.92	0.79	0.1	0.83	0.56	16.56
	74.61	11.62	4.78	0.6	5.02	3.37	
	16.44	16.11	16.78	11.52	19.66	18.18	
Eastern Cape	1,456	248	81	10	76	40	1,911
	6.55	1.12	0.36	0.04	0.34	0.18	8.59
	76.1	12.98	4.24	0.52	3.98	2.09	
	5.47	9.34	7.72	5.24	8.08	5.87	
Northern Cape	914	132	48	5	60	42	1,201
	4.11	0.59	0.22	0.02	0.27	0.19	5.4
	76.1	10.99	4	0.42	5	3.5	
	5.47	4.97	4.58	2.62	6.38	6.16	
Free State	1,709	148	79	22	61	42	2,061
	7.69	0.67	0.36	0.1	0.27	0.19	9.27
	82.92	7.18	3.83	1.07	2.96	2.04	
	10.22	5.57	7.53	11.52	6.48	6.16	
Kwazulu – Natal	2,421	335	111	35	90	58	3,050
	10.89	1.51	0.5	0.16	0.4	0.26	13.72
	79.28	10.98	3.64	1.15	2.95	1.9	
	14.48	12.61	10.58	18.32	9.56	8.5	
North-West	1,678	202	78	20	47	47	2,072
	7.55	0.91	0.35	0.09	0.21	0.21	9.32
	80.98	9.75	3.76	0.97	2.27	2.27	
	10.04	7.61	7.44	10.47	4.99	6.89	
Gauteng	2,586	701	264	32	273	225	4,081
	11.63	3.15	1.19	0.14	1.23	1.01	18.35
	63.37	17.18	6.47	0.78	6.69	5.51	
	15.47	26.39	25.17	16.75	29.01	32.99	
Mpumalanga	1,814	183	94	19	83	65	2,258
	8.16	0.82	0.42	0.09	0.37	0.29	10.16
	80.34	8.1	4.16	0.84	3.68	2.88	
	10.85	6.89	8.96	9.95	8.82	9.53	
Total	16,715	2,656	1,049	191	941	682	22,234
	75.18	11.95	4.72	0.86	4.23	3.07	100
χ^2 ***(40)***	= 603.934						
p-value	< 0.0001						

Given the statistically significant χ^2-test statistic above, it is possible to state that there is in fact dependence of province on trends in income distribution.

Appendix B: Micro-simulation Model Regression Results

Table 18: Regression results of MSM

Variable	Parameter estimate	Standard error	t-statistic	p-value
race_1	-0.608	0.075	-8.14	<0.0001
race_2	-0.569	0.13	-4.43	<0.0001
race_3	-0.005	0.11	-0.04	0.0001
drace11	-0.042	0.08	-0.60	0.5962
drace12	0.141	0.13	1.07	0.2859
drace13	-0.286	0.12	-2.39	0.017
drace21	0.185	0.08	2.26	0.024
drace22	0.415	0.14	3.01	0.0026
drace23	-0.157	0.13	-1.21	0.2277
drace31	0.129	0.09	1.42	0.1559
drace32	0.490	0.15	3.26	0.0011
drace33	-0.223	0.15	-1.52	0.1281
drace41	0.285	0.16	1.82	0.0680
drace42	0.417	0.26	1.61	0.1064
drace43	-0.039	0.28	-0.14	0.8906
drace51	0.155	0.09	1.67	0.0953
drace52	0.279	0.15	1.84	0.0662
drace53	-0.030	0.15	-0.20	0.8422
educ_1	0.763	0.23	3.36	0.0008
educ_2	1.069	0.23	4.67	<0.0001
educ_3	1.888	0.21	8.81	<0.0001
educ_4	1.838	0.32	5.78	<0.0001
educ_5	2.205	0.22	10.17	<0.0001
deduc11	-0.720	0.23	-3.13	0.0017
deduc12	-0.896	0.23	-3.86	0.0001
deduc13	-1.609	0.22	-7.40	<0.0001
deduc14	-1.465	0.34	-4.31	<0.0001
deduc15	-1.694	0.23	-7.52	<0.0001
deduc21	-1.340	0.27	-4.99	<0.0001
deduc22	-1.404	0.27	-5.20	<0.0001
deduc23	-2.047	0.26	-7.97	<0.0001
deduc24	-2.055	0.36	-5.70	<0.0001
deduc25	-2.239	0.26	-8.55	<0.0001
deduc31	-1.228	0.45	-2.71	0.0068
deduc32	-1.206	0.45	-2.66	0.0078
deduc33	-1.790	0.44	-4.04	<0.0001
deduc34	-1.831	0.52	-3.51	0.0005
deduc35	-2.056	0.45	-4.60	<0.0001
deduc41	-125.499	19.46	-6.45	<0.0001
deduc42	-125.453	19.42	-6.46	<0.0001
deduc43	-126.206	19.43	-6.49	<0.0001
deduc44	-126.235	19.44	-6.49	<0.000
deduc45	-126.367	19.42	-6.51	<0.0001

Variable	Parameter estimate	Standard error	t-statistic	p-value
deduc51	-1.395	0.39	-3.58	0.0003
deduc52	-1.219	0.39	-3.15	0.0017
deduc53	-1.736	0.37	-4.63	<0.0001
deduc54	-1.653	0.46	-3.60	0.0003
deduc55	-1.971	0.38	-5.24	<0.0001
age	0.545	0.08	6.49	<0.0001
dage1	-0.532	0.08	-6.27	<0.0001
dage2	-0.697	0.10	-7.08	<0.0001
dage3	-0.624	0.12	-5.22	<0.0001
dage4	-0.512	0.23	-2.18	0.0289
dage5	-0.574	0.12	-4.68	<0.0001
prov_1	0.192	0.13	1.49	0.1372
prov_2	0.059	0.15	0.39	0.6982
prov_3	0.187	0.15	1.22	0.2206
prov_4	-0.217	0.15	-1.44	0.1499
prov_5	0.169	0.14	1.22	0.2234
prov_6	0.130	0.15	0.88	0.3780
prov_7	0.201	0.12	1.70	0.0889
prov_8	0.031	0.14	0.23	0.8191
dprov11	-0.103	0.13	-0.78	0.4336
dprov12	-0.25	0.15	-0.16	0.8691
dprov13	-0.238	0.16	-1.53	0.1264
dprov14	0.411	0.15	2.70	0.0070
dprov15	-0.074	0.14	-0.53	0.5977
dprov16	-0.115	0.15	-0.78	0.4382
dprov17	-0.248	0.12	-2.06	0.0395
dprov18	0.021	0.14	0.15	0.8805
dprov21	-0.148	0.14	-1.03	0.3015
dprov22	0.003	0.16	0.02	0.9859
dprov23	-0.171	0.17	-1.00	0.3161
dprov24	0.310	0.17	1.87	0.0620
dprov25	-0.118	0.15	-0.79	0.4287
dprov26	-0.138	0.16	-0.87	0.3867
dprov27	-0.243	0.13	-1.89	0.0583
dprov28	-0.105	0.15	-0.70	0.4865
dprov31	-0.007	0.16	-0.04	0.9670
dprov32	0.038	0.18	0.21	0.8361
dprov33	-0.213	0.20	-1.09	0.2755
dprov34	0.351	0.18	1.95	0.0515
dprov35	-0.002	0.17	-0.01	0.9928
dprov36	-0.047	0.18	-0.27	0.7887
dprov37	-0.143	0.14	-1.01	0.3138
dprov38	-0.032	0.16	-0.19	0.8461
dprov41	-0.109	0.28	-0.39	0.6972
dprov42	-0.562	0.30	-1.89	0.0591
dprov43	0.192	0.40	0.48	0.6342
dprov44	0.273	0.26	1.04	0.2972
dprov45	0.156	0.24	0.66	0.5096

Variable	Parameter estimate	Standard error	t-statistic	p-value
dprov46	0.090	0.25	0.36	0.7225
dprov47	-0.097	0.23	-0.41	0.6798
dprov48	0.022	0.25	0.09	0.9302
dprov51	0.045	0.168	0.27	0.7908
dprov52	0.053	0.19	0.28	0.7826
dprov53	-0.037	0.20	-0.19	0.8511
dprov54	0.396	0.20	2.02	0.0430
dprov55	0.081	0.18	0.45	0.6532
dprov56	-0.264	0.20	-1.35	0.1785
dprov57	-0.057	0.15	-0.37	0.7099
dprov58	0.235	0.18	1.32	0.1853
loc	0.351	0.07	4.69	<0.0001
dloc1	-0.306	0.08	-4.04	<0.0001
dloc2	-0.241	0.08	-2.88	0.0039
dloc3	-0.217	0.10	-2.27	0.0235
dloc4	-0.316	0.16	-1.94	0.0528
dloc5	-0.221	0.11	-2.08	0.0379
numhh	-0.304	0.05	-6.48	<0.0001
dnumhh1	-0.018	0.05	-0.38	0.7044
dnumhh2	-0.055	0.05	-1.04	0.2967
dnumhh3	0.108	0.06	1.74	0.0824
dnumhh4	0.180	0.11	1.63	0.1023
dnumhh5	-0.017	0.07	-0.26	0.7942
save	-7.949	0.94	-8.44	<0.0001
dsave1	4.984	0.94	5.29	<0.0001
dsave2	-2.621	0.99	-2.63	0.0084
dsave3	-8.393	1.20	-6.97	<0.0001
dsave4	-22.583	2.56	-8.82	<0.0001
dsave5	-2.996	1.16	-2.58	0.0097
dispinc	8.528	0.986	8.65	<0.0001
ddispinc1	-4.950	0.99	-5.02	<0.0001
ddispinc2	2.890	1.04	2.78	0.0054
ddispinc3	8.818	1.25	7.06	<0.0001
ddispinc4	34.860	3.77	9.24	<0.0001
ddispinc5	3.394	1.21	2.80	0.0051
intercept	0.788	0.11	7.49	<0.0001

F = 328.31 | $\overline{R}^2 = 0.6516$