University of Pretoria
Department of Economics Working Paper Series

Defining the Tax Revenue Base for Individuals in South African Economy Using a Micro-Simulating Tax Model

Niek Schoeman
University of Pretoria
Yolande van Heerden
University of Pretoria
Working Paper: 2010-25
October 2010

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 4125

DEFINING THE TAX REVENUE BASE FOR INDIVIDUALS IN THE SOUTH AFRICAN ECONOMY USING A MICRO-SIMULATING TAX MODEL

N.J. Schoeman[1*] and **Y. van Heerden**[2]

ABSTRACT

The use of micro-simulation tax modelling techniques is reasonably well-documented in a broad spectrum of literature in the field of public economics. This article is primarily concerned with assessing the revenue base for individuals by means of such a micro-simulation tax model, using the 2005/2006 Income and Expenditure survey. The challenge was to structure the model in such a way that differences in individual behaviour, the economic environment and the income levels of individuals be captured to reflect the true national economy. The model developed is an extension of the MS model framework as structured by Thompson and Schoeman (2006) as well as Wilkinson (2009). It is different though in the sense that StatsSa data is aligned with published data from the South African Revenue Services (SARS). Given the scarcity of data (limited surveys) this model is a static model assuming that the population characteristics do not change significantly over the period of the analysis and that it remains useful in the short term. The structured model applies a tax calculator to compute the tax liability for each individual under the 2005/2006 tax regulations and rules. The results based on IES data is then benchmarked against the latest available published SARS data in the bulletin Tax Statistics (2009) and the relevant data in the latest (2010) publication Budget Review from the National Treasury.

An analysis based on unadjusted data from the IES shows a substantial difference in tax liability compared to official tax figures published by SARS (R65 billion compared to the SARS figure of R101 billion for the 2005/06 IES survey[3] year). After benchmarking critical values and the imputation of missing data the numbers are now much closer (R105 billion compared to the SARS R101 billion). The analysis is concluded with some policy scenarios showing the impact of a change in marginal tax rates and the tax threshold. The results highlight the sensitivity of high income earners to changes in tax policy.

JEL: D31, H24

Key words: Micro-simulation, Tax revenue base, Personal income tax

---

[1]Professor in Economics, University of Pretoria.
[2] PhD student in the Department of Economics, University of Pretoria.
* Corresponding author
[3] The SARS aggregates are weighted averages of 2005 and 2006 data; since IES 2005/2006 covered the period 1 September 2005 to 31 August 2006, a weighted average of SARS data was constructed for comparison: 2005 (6 out of 12 months) and 2006 (6 out of 12 months).

1. INTRODUCTION

In the field of public economics micro-simulation (MS) models are used internationally for the empirical analysis of fiscal policy changes on revenue collection and expenditure, especially health care and retirement as well as other socio-economic expenditures (Buddelmeyer, Creedy & Kalb, 2007:3). It allows for individual characteristics such as the composition of the taxpaying population in terms of age, gender, income levels, etc. and is especially useful to simulate individual income and expenditure behaviour to policy changes that affect revenue (Citro & Hanushek, 1991:15). This is in contrast to macro models which are structured on an aggregate level without the detail information of individuals/households captured in the micro model (Štěpánková, 2002:36). Furthermore, static models should be distinguished from dynamic models. In a static MS model the demographic characteristics of a particular survey are kept unchanged whereas in a dynamic MS the demographic characteristics adjust over time (the data ages). In the domain of public economics, MS procedures, whether static or dynamic are useful to simulate the effect of a fiscal policy change on revenue and expenditure patterns within households. The procedure involves data validation, imputation of data, re-weighting and the up-dating of data to characterise the population as closely as possible (Redmond, Sutherland & Wilson, 1998:4).

The layout of the rest of the paper is as follows: Section 2 contains the structure of a MS model. Section 3 constructs an MS model for South Africa and gives the analytical framework. In Section 4 the MS model is validated against the actual SARS data. The impact of policy changes in the revenue base are presented in Section 5 and Section 6 concludes with some policy recommendations.

2. STRUCTURE OF A MICRO-SIMULATION TAX MODEL

According to Merz (1991:79) an MS tax model is a simulation model that applies rules and calculations to existing micro-data sets containing observations described by taxpayer characteristics. These micro units can be firms, individuals or families and are identified by characteristics such as age, gender, population group, income, expenditure, and education.

Figure 1 reflects the structure of an MS model. A salient feature is that more than one database can be used for the model. As part of the process of constructing an MS model, the quality of the survey data is evaluated and shortcomings are addressed by the imputation of missing variables. To bring the database to the current year of evaluation the database is aged. This re-weighting of the data base takes account of the population change. Next, the income and expenditure variables need to be up-rated to the current year of evaluation. For income, the average income per taxable income group is calculated and evaluated against the published data. For the expenditure data, the totals are up-rated to the National account totals for expenditure items, durable, non-durable and services to the current year of evaluation, assuming that individuals consume the same goods and services. Price inflation factors from the consumer price index have to be applied to the expenditure data as well. The population sample does not always contain all the variables needed for the simulation of the MS model. Individuals do not report their tax liabilities in surveys accurately, which usually gives an underestimation of tax liability. Therefore a tax calculator is used to compute the personal income tax (PIT) liability for each individual in the model. An individual's level of expenditure will then be derived using an expenditure function. The current economic, tax and social policies are simulated and the results can be used as a benchmark for further policy analysis of the micro units concerned. It is then important to evaluate the quality of the output of the MS model against the actual for a valid MS model. Lastly, the model should be documented for use and further development by others (Citro & Hanushek, 1991:2-4).

There are two types of MS models, namely static and dynamic models. Static MS models assume that the population characteristics do not change significantly over a period of time and are useful models in the short term. The survey data used in an MS model are conducted periodically and are usually dated when published. To forecast the database to the actual data in the year of investigation involves a static ageing process. The structure of the sample is required to be re-weighted while the sample size, age and gender stay the same. Each micro unit will now represent a different number (new weight) of individual units in the total population. Static models are easier to develop than behavioural models (Citro & Hanushek, 1991:3).

A dynamic MS model ages each micro unit based on probabilities. This provides a history of each individual, taking into account the whole life cycle from birth to death, for example, a child ages and is old enough to join the labour force, a person in the workforce ages to a pensioner, newborns enter the population etc. These changes will affect the characteristics of the survey sample size and also change the population which is important for tax, pension and social policy analyses (Merz, 1991:79-81).

3. CONSTRUCTING A MICRO-SIMULATION TAX MODEL FOR SOUTH AFRICA

3.1 Data

MS models are as reliable as the micro datasets on which these models are based (Lau, Yotopoulos, Chou & Lin, 1981:175). Data sources that are representative of the South African population show a high level of versatility in the way in which the data is utilised. Databases do have missing values as a result of non-responses due to refusal, non-usable information and disqualified answers. Incomplete surveys affect the quality of survey results and needs to be dealt with appropriately (Hérault, 2007:324).

The problem of missing values has been addressed through the imputation of missing values using the techniques explained by Peichl & Schaefer (2009:3). In the proposed model a tax calculator computes the tax liability for each individual under the 2005/2006 tax regulations and rules. The results based on IES data is then benchmarked against the latest available published SARS data in the bulletin Tax Statistics (2009) and the relevant data in the latest (2010) publication Budget Review from the National Treasury.

A problem encountered was that the IES and SARS databases have different base years (calendar versus fiscal year). Given the fact that this model is based on IES data, fiscal year data were reworked to calendar years.

A unique feature of this MS model is that the results have been benchmarked with the published SARS data on registered filers (Tax Statistics, 2009). Thus, the quality and reliability of the model could be evaluated by comparing the estimated total gross income of individuals from the survey data with the actual published data.

The following characteristic variables have been identified for each individual, namely gender, age group, education level, population group, settlement and household size - all available from the IES. Some individuals earn income from different sources and these incomes have been aggregated in order to calculate taxable income for each individual in the model. Table 1 indicates the income sources defined and used in the 2005/2006 IES. Gross income excluding imputed rent is divided into six main groups, namely income from work, income from capital, pensions and annuities, social insurance and grants and other income. Gross income mainly originates from employment income which comprises 82.1 per cent of the total gross income while approximately 17.1 per cent originates from the other main sources. Individuals receive most of their income as salary/wages which is 71.3 per cent of total gross income. Individuals with business income comprise 10.81 per cent of the total gross income.

The behaviour of individuals is explained through the categorical variables from the survey. The gender variable differentiates between males and females and it is important for the model to see if a particular gender group is under-represented or not to be included in the survey. Each individual in the household belongs to a population group namely: African/Black, Coloured, Indian/Asian and Whites. Education groups range from no schooling, primary schooling, secondary schooling, degrees and diplomas. Only qualifications already obtained are included. Diplomas and certificates only count if at least a six months course has been completed. Age is captured in complete years to the nearest complete number and categorised in five-year age groups. These age estimates have been applied during the benchmarking process. Household size is the number of persons living in the same dwelling including children. Settlement is where the dwelling unit is situated. Urban areas include cities and towns characterised by higher population densities, economic activity and infrastructure. Rural areas include farms and traditional areas characterised by low population densities, economic activity and infrastructure (Statistics South Africa, 2008:1-2).

The categorical data contains missing/unspecified values and values had to be imputed to improve the quality of the data set. These missing values resulted from non-responses, refusal, non-useable information and disqualified answers.

For the categorical variables in the IES survey containing missing/non-response data, a frequency table for each variable was obtained to determine the distribution of the missing values. When computing values for the missing categorical variables the frequency distribution of the original responses remained unchanged. This methodology is available in the SAS programme known as RANUNI (Uniform Random Number Generator). The algorithm is briefly as follows:

In equation 1 R_i is the i^{th} random number, a is the multiplier and c the percentage increase.

$$R_{i+1} = (aR_i + c)(\bmod m) \quad i = 0,1,2,\ldots\ldots \qquad (1)$$

The RANUNI function then generates a random number using a generator developed by Lehmer (1951) from a uniform (0, m) distribution and turns it into (0,1) by dividing by m. The number in parentheses is the seed/random number of the random number generator. If the seed is adjusted to a non-zero number, the same random numbers are being generated, every time the program is activated (Fan, Felsovalyi, Sivo & Keenan, 2002:26).

Table 2 shows that the gender distribution before imputation reflects male responses at 47.1 per cent, female responses at 52.8 per cent and non-responses amount to 0.1 per cent. Using the RANUNI statistical method a missing value is replaced by a female response when the RANUNI is less than 52.8 or alternatively to a male response should the RANUNI be less than 47.1 per cent. It is also evident that the female and male distribution before and after the imputation has only deviated slightly between males and females, which is statistically correct. Thus, the male ratio only increased from 47.1 to 47.17 per cent and the female ratio increased from 52.8 per cent to 52.83 per cent.

In Table 3 the population distribution before imputation is as follows: Africans 78.5 per cent, Coloureds 13.6 per cent, Indian/Asian 1.6 per cent and Whites 6.2 per cent. The non-response number amounts to 0.1 per cent for the total population in the survey. The distribution between the population groups has only deviated slightly. For example, the ratio for African/Black increases only marginally from 78.5 to 78.6 per cent.

In Table 4 the age distribution before imputation, shows an aggregate distribution of 33 per cent for age group 0 to 14 years. For age group 15 to 24 years, 25 to 44, 45 to 64, 65 years and older, an aggregate distribution is 21.3, 25.1, 14.5 and 5.8, respectively. The non-response number amounts to 0.23 per cent for the total age group in the survey. The distribution between the age groups after imputation only adjusted slightly. For example, the aggregate ratio for age group 15 to 24 years increases from 21.34 to 21.38 per cent and age group 65 years and older from 5.81 to 5.82 per cent.

The distribution of the education groups before and after imputation can be seen in Table 5. For the no schooling group the distribution is 20.7 per cent. Primary schooling (Grade R-Grade 12) has an aggregate distribution of 73.8 per cent, NIC or diploma group an aggregate distribution of 3.6 per cent and a degree an aggregate distribution of 1.3 per cent. After imputation the distribution between the education groups has only changed slightly. For example, the aggregate ratio for no schooling increases from 20.7 to 20.8 per cent and the primary group increases from 73.8 to 74.2 per cent.

3.2 Comparing IES average income of the taxpayers

The average income of the individuals is compared to the average income per taxable income group of the published SARS data set for 2005/2006 in Table 6. Thus the SARS data serves as the benchmark data based on the principle that it is supposed to be a reflection of the full taxpayer base of the South African economy. It is evident from the Table that average income per taxable income group in the two different databases is comparatively close to each other indicating that the IES database can be used for the simulation of the MS model.

3.3 Analytical framework

The structure of the analytical framework is outlined in Table 7. Gross income is the result of the income specification in Table 1. From Gross income tax expenditures are deducted (exclusions, allowances and deductions) which then provides taxable income. Tax liability is calculated according to the tax rates and rebates in the tax tables of 2005/2006 (Budget review, 2006). By deducting rebates from the gross tax liability, net tax liability is derived.

In this model the net tax liability for each individual (i) is calculated using the IES dataset. Tax expenditure ($t\exp_i$) data is not accurately recorded in the IES and therefore the SARS published data in the Tax Statistics 2009 have been used as a proxy to calculate a ratio for tax expenditures to be applied to each individual gross income group. An average tax expenditure ratio (τ_{exp}) is derived from taxable income ($tbinc_i$) and gross income (y_i) per taxable income group in equation (2).

$$\tau_{exp} = \frac{y_i - tbinc_i}{y_i} \tag{2}$$

Equation (2) is applied to the IES and SARS dataset and in Table 8 the calculated tax expenditure ratio for each taxable income group can be seen.

These ratios by taxable income group (equation 2) are then applied to each individual IES gross income group in equation (3) to calculate individual tax expenditures.

$$t\exp_i = y_i * \tau_{exp} \tag{3}$$

Taxable income is defined as gross income less tax expenditures:

$$tbinc_i = y_i - tax\exp_i \tag{4}$$

The tax liability for each individual is calculated in equation (5) by applying tax codes determined annually by Parliament (Table 9) to taxable income:

$$pit_i = f(tbinc_i : \tau_{structure}) \tag{5}$$

Disposable income (yd_i) in equation (6) is the difference between gross income and tax liability:

$$yd_i = y_i - pit_i \quad . \tag{6}$$

Table 9 contains the rates of normal tax payable by natural persons in respect of the year of assessment ending 28 February 2006. In terms of the above rates of taxation, a person who has a taxable income up to R80 000 pays tax at a fixed rate of 18%. Income up to R300 000 is taxed at increasing rates. All income in excess of R300 000 is taxed at 40%.

Thus, the model calculates tax liability given the existing tax codes which can be changed for policy simulation purposes. It should also be mentioned that this procedure is of course a static method and that behavioural changes are not discounted for. However, it allows for policy simulations with regard to thresholds, marginal tax rates, allowances and income bands according to the six income categories. Obviously the impact of tax policy changes is much broader than only the static effects and therefore, in the next phase of the model development, the model will be adjusted to include dynamic behaviour of individuals or households. In order to forecast tax liability over the medium term the micro model is aligned to a macro-econometric model to estimate the tax liability over the short and long run.

As a cross-check on the model results the tax liability is also estimated by using elasticities on an aggregate/macro level. Again, macro data is used such as national income as a proxy of taxable income. Also, the variables such as gross domestic product (gdp), wages (w) and employment (n) are used to adjust income levels of individuals over the forecasting period. For forecasting purposes, a simple income function was estimated using an Ordinary Least Square Regression[4] procedure with data from SARB. A detailed description of the data is provided in Table 10. A dummy variable was incorporated to account for the structural break in the data series caused by major tax reforms in 2001 (see Nyamongo & Schoeman, 2007:482).

The appropriateness of the estimation techniques has to be valued against the availability of data. This study used 29 observations, and the number of methods that would be feasible was therefore limited. Differencing the series once, the ADF unit root test confirms stationarity of the series, (all series are I(0)).

Personal Income Tax (PIT) is positively correlated with tax ratios and real economic growth. The function used to estimate PIT takes the form:

$$ltax = f(\overset{+}{lgdp}, \overset{+}{ltaxratio})$$ [5]

In Table 11 the signs and magnitudes of the variables in the long-run equation do conform to *a priori* expectations. The probabilities of the t-statistics of the explanatory variables are all statistically significant at one per cent level of significance. The Adjusted R^2 value indicates a

[4] The Engle-Granger two-step estimation technique has potential defects in the sense that it assumes that there is one cointegrating vector. It also carries forward an error made in the first step to the second step of estimation. Since the number of variables (n = 3) in the model is greater than two, there can be more than one cointegrating relationship.

[5] See Table 10

good fit with 99.8 per cent of the variation in PIT being explained by the equation. It is expected that an increase in output (lgdp), tax ratio (ltaxratio) will increase tax (tax). A one per cent increase in output would lead to a 1.1 per cent increase in PIT, and a one per cent increase in the tax ratio would lead to a 0.8 per cent increase in PIT.

To test if there is cointegration[6] between the variables the hypothesis test for cointegration is as follows:

H_0: no cointegration
H_1: cointegration

In Table 12 the variables are cointegrated at a 10% level of significance, as the ADF statistic -3.93 is smaller than the calculated MacKinnon (MacKinnon, 1991) critical value of -3.6783, thereby rejecting the null hypothesis.

The Error Correction model incorporating the short-run effects on tax corrects the stochastic residuals from the long-run cointegrating regression. The results are shown in Table 13.

All the variables included in the ECM were originally I(1). Differencing them once transformed them into I(0) series. The error correction coefficient is negative and statistically different from zero. The Adjusted R^2 value indicates that 65 per cent of the variation in taxes is being explained by the ECM.

All the diagnostic tests were performed on the ECM, with the following results in Table 14. Thus given the diagnostic results at a one percentage level of significance, it is reasonable to conclude that the residuals do satisfy the assumptions of the classical normal linear regression model.

In order to address the problem of non-stationarity of the time series in the cointegration equation the coefficients and t-values had to be adjusted using the error correction model (ECM). Thus the t-statistics are also not suitable for inference because of their inaccuracy and biasedness. The ECM is used via its residuals to adjust the long-run coefficients and their corresponding t-

[6] Cointegration involves combining economic data series (although I(1)) through a linear combination (cointegrating vector) into a single series, which is itself stationary. The ECM separates the long-run relationship between the variables from the short run responses. The Engle-Granger (Engle & Granger, 1987) two-step procedure and the error correction paradigms were adopted, despite their potential defects. This technique entails the determination of the long-term cointegration relationship through testing for stationarity of the residuals using Augmented Dickey-Fuller (ADF) tests. Any non-stationarity is then corrected for by means of a short-term error correction model (ECM).

statistics. The residuals from the ECM are then regressed on the variables included in the long-run equation multiplied by the negative coefficient of the residuals from the cointegrating equation retrieved from the ECM. The results of the Engle_Yoo regression can be seen in Table 15.

In Table 16 the adjusted coefficients are statistically highly significant as their respective t-statistics are all larger than 1.96 in absolute value terms. These new adjusted t-statistics, can be used for statistical inference. A one per cent increase in output would lead to a 1.08 per cent increase in PIT, and a one per cent increase in the tax ratio would lead to a 0.67 per cent increase in taxes.

Estimated PIT is modelled and then compared to the actual available data in Figure 3. To forecast PIT values from 2010 onwards, growth rates of the explanatory variables can be calculated as the year-on-year percentage changes observed during the last 5 years.

4. MODEL VALIDATION

After tax liability has been simulated with the MS model, the results from the model are compared with the reported IES and published SARS data. All data sets have been adjusted to the survey year 1 September 2005 to 31 August 2006. In Table 17 it is interesting to note that when an analysis based on unadjusted data from the IES is drawn, there is a substantial difference in tax liability compared to official tax figures published by SARS (R65 billion compared to the SARS figure of R101 billion for the 2005/06 survey year). After benchmarking critical values and the imputation of missing categorical data the numbers are now much closer (R105 billion compared to the SARS R101 billion). The MS model only calculates tax liability and forward payments affect the actual amount collected (R133 billion). The gap between the simulated and actual taxable income is about 13 per cent, meaning that the MS model is comparatively close to the actual reported values published by SARS.

Comparing the results of the analysis based on both the adjusted IES survey data and the SARS published data indicate that the difference is actually only marginal. The results are firstly compared on a gender base, followed by a comparison of different income groups.

Table 18 shows that in the case of gender, the SARS data account for the taxable income of females to be approximately R159,3 billion compared to the model calculation based on IES data of R169,4 billion. In terms of the number of taxpayers, the two sets of data is actually very close (1,627 million compared to 1,646 million). The average taxable income based on the IES data is slightly higher than in the SARS data but the tax liability is lower. The reason is that more of the females fall into the lower taxable income groups as will be indicated in the next paragraph. As far as males are concerned, the Table shows that the IES data reflect more males than the SARS data with both taxable income and tax liability also higher. In this case the reason is that in the case of the former data set (IES) more males are included that do not reflect on the official SARS database. As a result the total adjusted tax liability as calculated and based on IES data exceeds the official SARS number by more than R3.6 billion. Males account for almost 72 per cent of total tax liability and comprise 58 per cent of the registered taxpayers earning 66 per cent of total taxable income. Males are about 56 per cent of the employed labour market (StatsSa, 2008:8). The data indicate the uneven distribution of taxable income between males and females a (difference of 32 per cent) despite the fact that the numbers only differ by 16 per cent. As a result the tax liability of males is about 75% of the total tax liability.

Table 19 shows the number of taxpayers, taxable income and tax assessed by age group. In the case of the first age group (<18), the numbers included in the SARS data is close to the survey data, but the taxable income in the SARS data is about 40 per cent less than the IES data (R757 million compared to R1 292 million). This is interesting but not unexpected given the fact that a number of youngsters are listed on the SARS data, based on the fact that although they are not formally employed, they have earned taxable income from inheritances, grants, and/or other sources of income which are not reflected in the IES data. In the case of all the other age groups (except the category 35-44 and >65) the IES taxable income and tax liability data is less than the SARS data, but the values are very close.

For the age group under 34 the number of taxpayers as a percentage of total taxpayers is around 26 per cent and contributing about 21 per cent of total tax liability. The age group 34-44 comprises 30 per cent of total taxpayers and contributes 33 per cent. The age group above 44 comprises 43 per cent of total taxpayers and is responsible for 46 per cent of total tax liability. As far as the age group >65 is concerned, the IES data comprises only about 30 per cent of the SARS number of taxpayers and clearly does not reflect the full number of taxpayers. This is probably due to the fact that the coverage of the IES did not include individuals in institutions such as hospitals and old-age homes. Therefore, the age group >65 is under represented in the IES dataset. Regarding the age groups in between, the differences are not large and the tax liability therefore also not substantial.

5. THE IMPACT OF POLICY CHANGES ON THE REVENUE BASE

As indicated at the start of this paper an MS model is a useful tool to simulate different policy scenarios. It is important to ensure that the MS model is valid by comparing it to actual data before policy scenarios can be tested. A change in the tax structure results in a change in an individual's estimated tax liability which directly influences the individual's disposable income. The model allows users to change the parameters underlying the tax structure (marginal tax rates, rebates and threshold levels) to simulate changes in tax policy. In the text to follow, three different policy scenarios have been tested.

Scenario 1: A one per cent increase in the marginal tax rates of the first, second and third taxable income groups, from 18%, 25% and 30%, respectively.

Table 20 shows the result of the first tax reform with the difference between the before and after tax liability effect in Column 5. A tax change in the first income group's marginal rate influences all groups because the primary and secondary rebate is a calculated value of the marginal rate of the first income group multiplied by the threshold. An increase in the marginal rate of the first income group will increase taxes paid by individuals in the first income group by an average of R148.70. The second income group's tax liability decreases by an average of R123.19 due to the fact that the implicit increase in the rebate exceeds the increase in the tax burden. The third income group's tax liability decreases by an average of R110,90. This trend is replicated in all the rest of the income stratifications for similar reasons. As far as the other tax brackets are concerned, tax liability declines based on the fact that the rebate implicitly increases. The rest of the tax brackets average tax liability decrease because of the increase in rebates.

Thus, comparing the actual and simulated values after such a tax reform shows that the lowest income group will be hardest hit while the higher income groups will actually gain from it. In total, tax liability decreases by R119 million and therefore an increase in the marginal tax rate of the first three income groups reduces the size of the tax base.

Scenario 2: Increase in the marginal tax rates of one per cent in the highest two income groups.

Table 21 shows that when the marginal tax rates of the last two income groups are increased by one per cent from the current 38 and 40 per cent respectively, the average taxes paid by the fifth group increases by an average of R281.24 while the last group's tax liability increases by an average of R2,635.72. The latter is substantial clearly indicating the progressiveness of the South African tax system with the highest income groups proportionally more affected by relatively small changes in tax policy. Obviously the other income groups are not affected by this policy change in a static setup of the model. In total, tax liability increases by R799 million, therefore an increase in the higher income group's marginal tax rate increases the size of the tax base.

Scenario 3: The tax threshold for individuals is increased by 10 per cent.

Table 22 shows that if the tax threshold is increased by 10 per cent, the tax liability of all income groups will decrease by a fairly fixed amount except in the case of the lower income group. In the case of the latter fewer individuals are included as taxpayers with such a lowering of the threshold and therefore the average decline in revenue received from this group is less than that of the other groups, the reason being that the latter now receives a lesser amount in rebates. Increasing the threshold allows significant real tax relief for the lower income earners, the first two income group's tax liability is reduced by 6.7 and 5.2 per cent respectively. In total tax liability decreases by R2.7 billion (2.6 per cent), therefore an increase in the threshold reduces the size of the tax base.

5. CONCLUSIONS AND RECOMMENDATIONS

This paper describes the initial phases of constructing an MS model. The model is different from other MS models in the sense that it calibrates the IES survey data with published official tax data of the South African Revenue Services.

Due to missing values and unreported data, data had to be imputed for the simulation of the MS model. In this first phase of model development, data manipulation accounted for the brunt of the research efforts. In this regard the following procedures are notable. For the missing categorical variables in the IES survey, a frequency table for each variable was obtained to determine the distribution of the missing values. The RANUNI statistical method was used to impute the missing values. The frequency distribution of the original responses remained unchanged. More than one source of income is aggregated and linked to one individual. Data on tax liability are not accurately recorded in the survey, therefore a tax calculator is used to calculate tax liability for each individual. The IES and SARS databases have different base years (calendar versus fiscal year). The MS model is based on IES data, therefore fiscal year data were reworked to calendar years (1 September to 31 August).

After the manipulation of the IES dataset an MS model is developed and the model data results are compared to the actual SARS data. The results positively indicate that the MS model is a valid model and can be used to simulate different tax policy scenarios. The scenarios reveal that those in the highest tax bracket will be hardest hit with an average increase in tax liability of approximately R2 635 per annum. Given the fact that individuals with a business income have the option to register as companies, this scenario should be carefully considered since the top marginal rate is already 12 % higher than the corporate rate. The result might simply be a reduction in the tax base. The scenarios concluded that for those in the lower income brackets tax liability decreases on average by R30.5 per annum. For an individual earning on average more than R180 000, tax liability decreases with R966 per annum. Therefore, policy changes to those income groups should be carefully considered. These scenarios highlight the progressiveness of the South African tax codes.

From a theoretical perspective, the proposed study will make a valuable contribution to policy makers and academics that will use the MS model as a tool to simulate the effect of new tax policies before they are applied.

Further studies will be conducted to estimate consumption functions for individuals and to link the MS model results to a macro Computable General Equilibrium (CGE) model capturing aggregate wage and price effects which strengthens the results from the MS model.

REFERENCES

Buddelmeyer, H., Creedy, J. & Kalb, G. 2007. *Tax policy - Design and Behavioural Microsimulation Modelling.* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing Limited.

Budget Review. 2006. Pretoria. Department of Finance.

Budget Review. 2010. Pretoria. Department of Finance.

Citro, C.F. & Hanushek A. 1991. *Improving Information for Social Policy Decisions The Uses of Microsimulation Modeling.* Washington, D.C.: National Academy Press.

Davies, J.B. 2009. Combining Microsimulation with CGE and Macro Modelling for Distributional Analysis in Developing and Transition Countries. *International Journal of Microsimulation,* 2(1):49-65. [Online] Available from: http://www.microsimulation.org/IJM/V2_1/IJM_2_1_4.pdf [Accessed: 2010-03-30].

Deaton, A. 1997. *The Analysis of Household surveys, A microeconometric Approache to Development Policy.* The Johns Hopkins University press, Baltimore, Maryland.

Dickey, D.A. and Fuller, W.A. 1981. Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, 49(4), pp 1057-1072.

Engle, R.F. and Granger C.W.J. 1987. Co-integration and error correction: Representation, estimation, and testing, *Econometrica*, 55(2), pp 251-276.

EViews 7 User's Guide II. 2010. United States of America: Quantitative Micro Software, LLC.

Fan, X. & Felsövályi, A. 2002. SAS for Monte Carlo studies: a guide for quantitative research. *SAS Institute* [Online] Available from: [Accessed: 2010-05-13].

Gujarati, D. 2006. *Basic Econometrics*, 4TH Edition. National Centre for Social and Economic Modelling. West Point Military Academy McGraw-Hill Higher Education

Harris, R.I.D. 1995. *Using cointegration analysis in econometric modelling*, Prentice Hall.

Hérault, N. 2006. Building and Linking a Microsimulation Model to a CGE Model for South Africa. *South African Journal of Economics*, 74(1):34-58.

Hérault, N. 2007. Trade Liberalisation, Poverty and Inequality in South Africa: A Computable General Equilibrium-Microsimulation Analysis. *The Economic Record,* 83(262):317-328.

Lau, L.J., Yotopoulos, P.A., Chou, E.C. & Lin, W. 1981. The microeconomics of distribution: A simulation of the farm economy. *Journal of Policy Modelling,* 3(2):175-206. [Online] Available from: http://www.sciencedirect.com/science/article/B6V82-45GNMFC-3S/2/deefbb47a22b852e71eac48397252222 [Accessed: 2010-03-19].

MacKinnon, J.G. 1991. *Critical values for cointegration tests. In long run economic relationships*, London, Oxford University Press, pp 267-276.

Mabugu, R. & Chitiga, M. 2007. *South Africa Trade Liberalization and Poverty in a Dynamic Microsimulation CGE Model.* Working Paper Series. University of Pretoria, Department of Economics. [Online] Available from: http://web.up.ac.za/UserFiles/WP_2007_18.pdf [Accessed: 2010-03-19].

Merz, J. 1991. Microsimulation - A survey of principles, developments and applications. *International Journal of Forecasting,* 7(1):77-104. [Online] Available from: http://www.sciencedirect.com/science/article/B6V92-45PMMPV-B/2/4203c554f9a3c65c0b37d73b93b92935 [Accessed: 2010-03-19].

Nyamongo, M.E. & Schoeman, N.J. 2007. Tax reform and the progressivity of personal income tax in South Africa. *South African Journal of Economics*, 75(3): 482. **[Online] Available from:** [http://www.blackwellpublishing.com/journal.asp?ref=0038-2280&site=1] **[Accessed: 2010-10-9].**

Peichl, A. & Schaefer, T. 2009. FiFoSiM - An Integrated Tax Benefit Microsimulation and CGE Model for Germany. *International Journal of Microsimulation,* 2(1):1-15. [Online] Available from: http://www.microsimulation.org/IJM/V2_1/IJM_2_1_1.pdf [Accessed: 2010-03-19].

Redmond, G., Sutherland, H. and Wilson, M. 1998. *The arithmetic of tax and social security reform A user's guide to microsimulation methods.* Cambridge University Press.

SARB (South African Reserve Bank). *Various issues of the Quarterly Bulletin*, Pretoria.

Statistics South Africa. 2008. *Income and Expenditure of Households 2005/2006: Statistical release P01001*. [Online] Available from: http://www.statssa.gov.za/publications/statsdownload.asp?PPN=P0100&SCH=4108 [Accessed: 2010-01-03].

Štěpánková, P. 2002. Using Microsimulation Models for Assessing the Redistribution Function of a Tax-Benefit System. *Finance a úvěr',* 52(1):36-50.

Thompson, K.L. & Schoeman, N.J. 2006. *Using a Static Micro-Simulation Model to Evaluate the South African Income Tax System*. Working Paper Series. University of Pretoria, Department of Economics. [Online] Available from:http://www.up.ac.za/up/web/en/academic/economics/index.html [Accessed: 2010-03-19].

Tax Statistics. 2009. Pretoria. National Treasury and the South African Revenue Service.

Vaqar, A. & Cathal, O. 2007. CGE-Microsimulation Modelling: A Survey. MPRA Paper No. 93072010.05/12. [Online] Available from: http://mpra.ub.uni-muenchen.de/9307/ [Accessed: 2010-05-10].

Wagenhals, G. 2004. *Tax-benefit microsimulation models for Germany: A Survey.* Discussion paper. University of Hohenheim, Institute of National Economy Apprenticeship, 235. [Online] Available from: http://www.uni-hohenheim.de/RePEc/hoh/papers/235.pdf [Accessed: 2010-03-29].

Wilkinson, K. 2009. Adapting EUROMOD for use in a developing country - the case of South Africa and SAMOD. No EM5/09, EUROMOD Working Papers. [Online] Available from: http://www.iser.essex.ac.uk/publications/working-papers/euromod/em5-09.pdf [Accessed: 2010-05-10]

Table 1: Sources of gross income

Source	IES 2005/2006 R (billion)		%
Income from work:	690.8		82.1%
Salaries and wages		599.9	
Self-employment and business income		90.9	
Income from capital:	10.8		1.3%
Interest received		4.2	
Dividends		1.6	
Rent income		4.9	
Royalties		0.1	
Private pensions and annuities:	24.3		2.9%
Pensions from previous employment		19.8	
Annuities from own investment		4.5	
Social insurance and grants:	56.8		6.8%
Old age and war pensions		25.3	
Disability grants		10.4	
Family and other allowances		20	
UIF, Workmen's Compensation		1.1	
Other income:	58.3		6.9%
Alimony, palimony and other allowances		11.1	
Other income from individuals		3.9	
Benefits, donations and gifts, and cash labola		3.7	
Tax refunds received		1.7	
Other : Letting of fixed property, Annuities, Hobbies, Gratuities, Income from gambling		37.9	
Gross income (Excluding Imputed rent on owned) dwelling)	841		100%

Source: Statistics South Africa (2008:9)

Table 2: Gender distribution

Gender	Distribution before imputation	Distribution after imputation
Male	47.1%	47.17%
Female	52.8%	52.83%
Non-response	0.1%	
Total	**100%**	**100%**

Source: Authors calculation in SAS 9.2

Table 3: Population distribution

Population	Distribution before imputation	Distribution after imputation
African/Black	78.5%	78.60%
Coloured	13.6%	13.64%
Indian/Asian	1.6%	1.56%
White	6.2%	6.20%
Non-response	0.1%	
Total	**100%**	**100%**

Source: Authors calculation in SAS 9.2

Table 4: Age group distribution

Age (years)	Distribution before imputation	Distribution after imputation
0 -4	10.25%	10.27%
5 -9	10.78%	10.81%
10 – 14	12.04%	12.07%
15 - 19	11.73%	11.75%
20 - 24	9.61%	9.63%
25 - 29	7.11%	7.12%
30 - 34	6.47%	6.48%
35 - 39	5.97%	5.99%
40 – 44	5.56%	5.58%
45 - 49	4.59%	4.60%
50 - 54	4.09%	4.10%
55 - 59	3.18%	3.19%
60 - 64	2.58%	2.59%
65 - 69	2.15%	2.15%
70 - 74	1.52%	1.52%
75 -79	1.11%	1.11%
80 -84	0.53%	0.53%
> 85	0.51%	0.51%
Non-response	0.23%	
Total	**100%**	**100%**

Source: Authors calculation in SAS 9.2.

Table 5: Highest level of education distribution

Level of education	Distribution before imputation	Distribution after imputation
No schooling	20.67	20.81
Grade R	3.25	3.26
Grade 1	3.19	3.21
Grade 2	3.71	3.73
Grade 3	4.27	4.30
Grade 4	4.61	4.63
Grade 5	4.91	4.94
Grade 6	5.46	5.49
Grade 7	6.94	6.99
Grade 8	7.33	7.37
Grade 9	6.39	6.42
Grade 10	7.21	7.24
Grade 11	6.03	6.06
Grade 12	10.52	10.57
NTC I	0.09	0.09
NTC II	0.08	0.08
NTC III	0.26	0.26
Diploma/certificate with less than Grade	0.20	0.20
Diploma with less than Grade 12	0.26	0.26

Certificate with Grade 12	0.74	0.75
Diploma with Grade 12	2.00	2.01
Bachelors Degree	0.64	0.65
Bachelors Degree and Diploma	0.27	0.27
Honours Degree	0.25	0.25
Masters/ Doctorate Degree	0.16	0.16
Non-response	0.60	
Total	**100**	**100**

Source: Authors calculation in SAS 9.2

Table 6: Average Gross income

Taxable income group	SARS	IES	Scale factor
R0 – R80 000	R48,064.02	R56,301.17	0.854
R80 001 – R130 000	R103,643.88	R102,704.15	1.009
R130 001 – R180 000	R150,840.37	R153,224.20	0.984
R180 001 – R230 000	R202,659.59	R202,365.21	1.001
R230 001 – R300 000	R261,202.39	R260,355.87	1.003
R300 001 and above	R558,246.47	R571,723.34	0.976

Source: Authors calculation in SAS 9.2

Table 7: Calculation of personal income tax base

Gross income
***Less*, Tax expenditures**
Exclusions
Entertainment expenses
Telephone allowances
Computer allowances
Current pension fund contributions
Provident fund contributions
Medical fund contributions
Donations
Other allowance and expenses
Taxable income
Calculate tax with tax tables
Gross Tax liability
***Less,* Rebates**
Net Tax liability

Source: Tax Statistics (2009:52)

Table 8: Tax expenditure factor

Taxable income group	Tax expenditure factor
R0 – R80 000	0.1085
R80 001 – R130 000	0.0458
R130 001 – R180 000	0.0348
R180 001 – R230 000	0.0184
R230 001 – R300 000	0.0087
R300 001 and above	0.0175

Source: Authors calculation in SAS 9.2

Table 9: Personal Income Tax structure 2005/2006

Taxable income Brackets	Marginal Rates of Tax	
R0 – R80 000		18% of each R1
R80 001 – R130 000	R14 400 +	25% of the amount above R80 000
R130 001 – R180 000	R26 900 +	30% of the amount above R130 000
R180 001 – R230 000	R41 900 +	35% of the amount above R180 000
R230 001 – R300 000	R59 400 +	38% of the amount above R230 000
R300 001 and above	R86 000 +	40% of the amount above R300 000
Primary Rebate: R6 300		
Secondary Rebate: R10 800		
Tax thresholds for Below 65 years: R35 000		
Tax thresholds for 65 year and older: R60 000		

Source: South African National Treasury

Table 10: Selected variables used for estimating personal income tax revenue

Abbreviation	Description	Transformation used
gdpn	gross domestic product at market prices	R millions current prices
pitrev	personal income tax as % of total revenue	Percentage
revgdp	total revenue as a percentage of gdp	Percentage
tax	personal income tax	pitrev/100*revgdp*100*gdpn
coe	compensation of employees	R millions current prices coe = n*w
propinc	property income	R millions current prices
tbinc	taxable income	coe + propinc
taxratio	tax ratio	tax(-1)/tbinc(-1)*100
n	total employment	R millions current prices
w	wages	R millions current prices
dum	Structural break from 2000	

Source: South African Reserve Bank quarterly bulletin, various issues

Table 11: Elasticities for the long run cointegration equation

Dependent variable: ltax	
Variables	**Coefficient**
lgdp	1.056676
ltaxratio	0.775893
dum	-0.202260
c	-5.140545

Source: EViews 7

Table 12: Testing stationarity of the cointegrating residual

Series	Model	Lags	τ
Res_lr	Constant, no trend	0	-3.933546

Source: EViews 7

Table 13: Regression output of the Error Correction Model for Tax

Variable	Coefficient
Res_lr(-1)	-0.664922
d(lgdp)	0.996696
d(ltaxratio)	0.695768

Source: EViews 7

Table 14: Diagnostic tests

Layout	Test	Test statistic	p-value	Conclusion
Normality	*Jarque Bera*	JB = 5.7	0.1	Normally distributed
Serial Correlation	*Ljung-Box Q*	$LB_Q = 9.2$	0.7	No serial correlation
	Breusch-Godfrey	$nR^2 = 0.3$	0.9	No serial correlation
Heteroscedasticity	*ARCH LM*	$nR^2 = 0.5$	0.5	No heteroscedasticity
	White (no cross)	$nR^2 = 3.1$	0.4	No heteroscedasticity
Specification	*Ramsey RESET*	LR = 0.006	0.9	Indicative of stability

Source: EViews 7

Table 15: Results of the Engle-Yoo regression

Variable	Coefficient
0.664922*ldgp	0.019123
0.664922*ltaxrtio	-0.100894

Source: EViews 7

Table 16: Adjusted long-run coefficients and t-statistics

Dependent variable: res_rt-1

Variable	Adjusted Coefficients	Standard Error	Adjusted t-Statistic
lgdp	1.075799	0.20604	52.21311
ltaxratio	0.674999	0.10828	6.233829

Source: EViews 7

Table 17: Comparison of IES, MS model and SARS for the survey year 2005/2006

Data	Gross Income	Taxable Income	Tax Liability
IES Survey data	R841,000,000,000 *(Total population)*	n/a	R64,700,000,000
MS Model	R751,219,462,881 *(Total population)* R 562,879,714,067 *(only taxpayers income)*	R 538,277,608,711	R 105,576,969,787
SARS Tax Statistics	n/a	R 469,228,000,000	R101,956,000,000 *(Assessed)* R133,111,847,500 *(Actual collected)*

Source: Statistics South Africa (2008:12), Tax Statistics (2009:15, 36)

Table 18: Number of taxpayers, taxable income and tax assessed by gender group

	SARS 2005/2006			IES 2005/2006		
Gender	**Number of taxpayers**	**Taxable income (R million)**	**Tax Liability (R million)**	**Number of taxpayers**	**Taxable income (R million)**	**Tax Liability (R million)**
Female	1,627,729	159,279	28,373	1,646,601	169,368	27,131
Male	2,204,264	309,949	73,583	2,724,890	368,901	78,446
Total	**3,831,993**	**469,228**	**101,956**	4,371,491	538,269	105,577

Source: Tax Statistics (2009:45), Authors calculation in SAS 9.2

Table 19: Number of taxpayers, taxable income and tax assessed by age group

	SARS 2005/2006			IES 2005/2006		
Age	**Number of taxpayers**	**Taxable income (R million)**	**Tax Liability (R million)**	**Number of taxpayers**	**Taxable income (R million)**	**Tax Liability (R million)**
< 18	19,234	757	105	21,264	1,292	106
18 - 24	121,021	7,935	1,052	190,590	16,722	2,408
25 - 34	880,952	102,176	19,775	1,289,655	128,770	21,251
35 - 44	1,132,337	154,340	33,997	1,251,802	152,036	28,785
45 - 54	875,526	122,218	28,714	1,007,126	144,487	31,643
55 - 64	502,066	60,565	14,626	516,865	82,943	19,657
> 65	300,858	21,238	3,689	94,189	12,019	1,727
Total	**3,831,993**	**469,228**	**101,956**	**4,371,491**	**538,269**	**105,577**

Source: Tax Statistics (2009:45), Authors calculation in SAS 9.2

Table 20: Tax reform 1: Average values per income group

Taxable income group (R)	**Average Gross income (R)**	**Average Taxable Income (R)**	**Average Tax Liability before reform**	**Average Tax Liability after reform**	**Average Tax Liability (R)**	**Growth in Tax liability (%)**
	1	**2**	**3**	**4**	**5 (4-3)**	**6 (4/3)**
0 – 80 000	56,301.17	50,191.95	2,676.60	2,825.3	148.70	5.6
80 001 –130 000	102,704.1	97,999.20	12,446.30	12,323.1	-123.19	-1.0
130 001 –180 000	153,224.2	147,887.0	25,826.38	25,715.5	-110.90	-0.4
180 001 –230 000	202,365.2	198,641.5	41,990.29	41,667.7	-322.63	-0.8
230 001 –300 000	260,355.8	258,101.8	63,642.33	63,284.7	-357.62	-0.6
>300 001	571,723.3	561,736.7	184,261.2	183,903.7	-357.49	-0.2

Source: Authors calculation in SAS 9.2, by using a MS model

Table 21: Tax reform 2: Average values per income group

Taxable income group (R)	**Average Gross income (R)**	**Average Taxable Income (R)**	**Average Tax Liability before reform**	**Average Tax Liability after reform**	**Average Tax Liability (R)**	**Growth in Tax liability (%)**
	1	**2**	**3**	**4**	**5 (4-3)**	**6 (4/3)**
0 – 80 000	56,301.17	50,191.95	2,676.60	2,676.6	0	0
80 001 –130 000	102,704.1	97,999.20	12,446.30	12,446.3	0	0
130 001 –180 000	153,224.2	147,887.0	25,826.38	25,826.38	0	0
180 001 –230 000	202,365.2	198,641.5	41,990.29	41,990.29	0	0
230 001 –300 000	260,355.8	258,101.8	63,642.33	63,923.57	281.2	0.44
>300 001	571,723.3	561,736.7	184,261.2	186,896.9	2,635.7	1.43

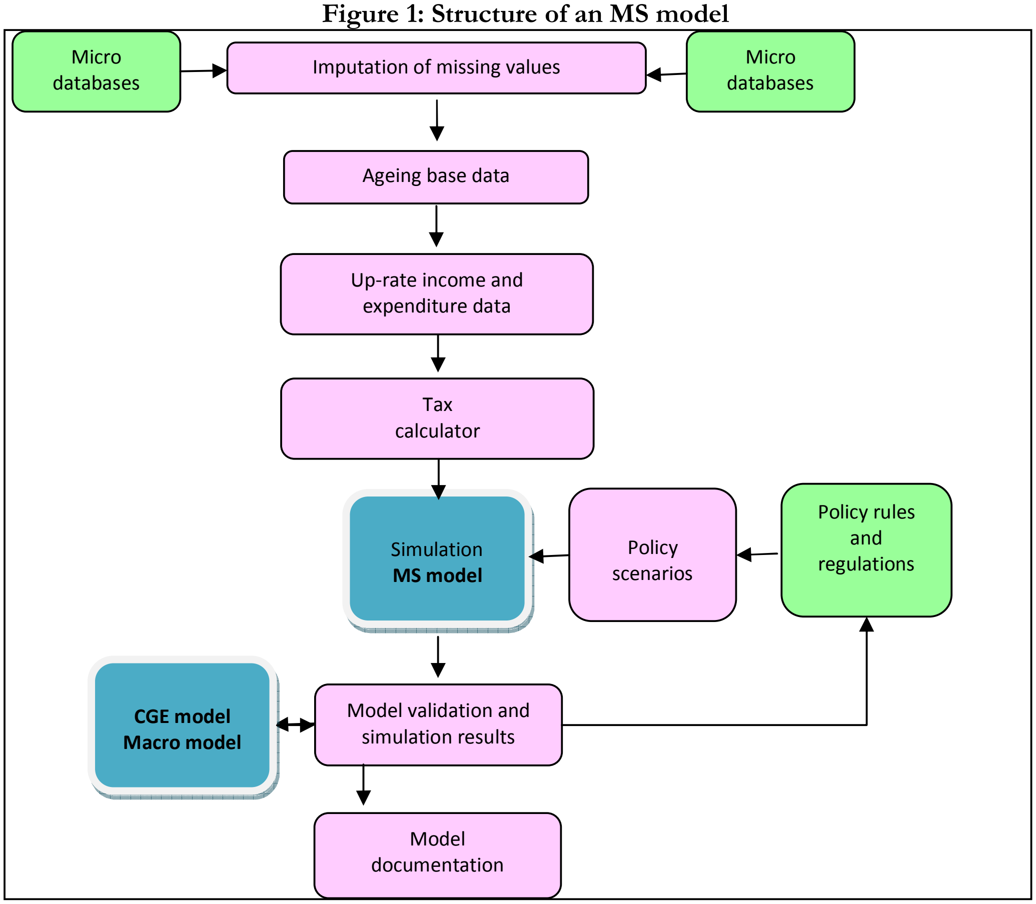
Source: Authors calculation in SAS 9.2, by using a MS model.

Table 22: Tax reform 3: Average values per income group

Taxable income group (R)	Average Gross income (R)	Average Taxable Income (R)	Average Tax Liability before reform	Average Tax Liability after reform	Average Tax Liability (R)	Growth in Tax liability (%)
	1	2	3	4	5 (4-3)	6 (4/3)
0 – 80 000	56,301.17	50,191.95	2,676.60	2,497.40	-179.20	-6.7
80 001 –130 000	102,704.1	97,999.20	12,446.30	11,800.03	-646.27	-5.2
130 001 –180 000	153,224.2	147,887.0	25,826.38	25,180.69	-645.69	-2.5
180 001 –230 000	202,365.2	198,641.5	41,990.29	41,346.37	-643.92	-1.5
230 001 –300 000	260,355.8	258,101.8	63,642.33	62,998.62	-643.71	-1.0
>300 001	571,723.3	561,736.7	184,261.2	183,617.7	-643.48	-0.4

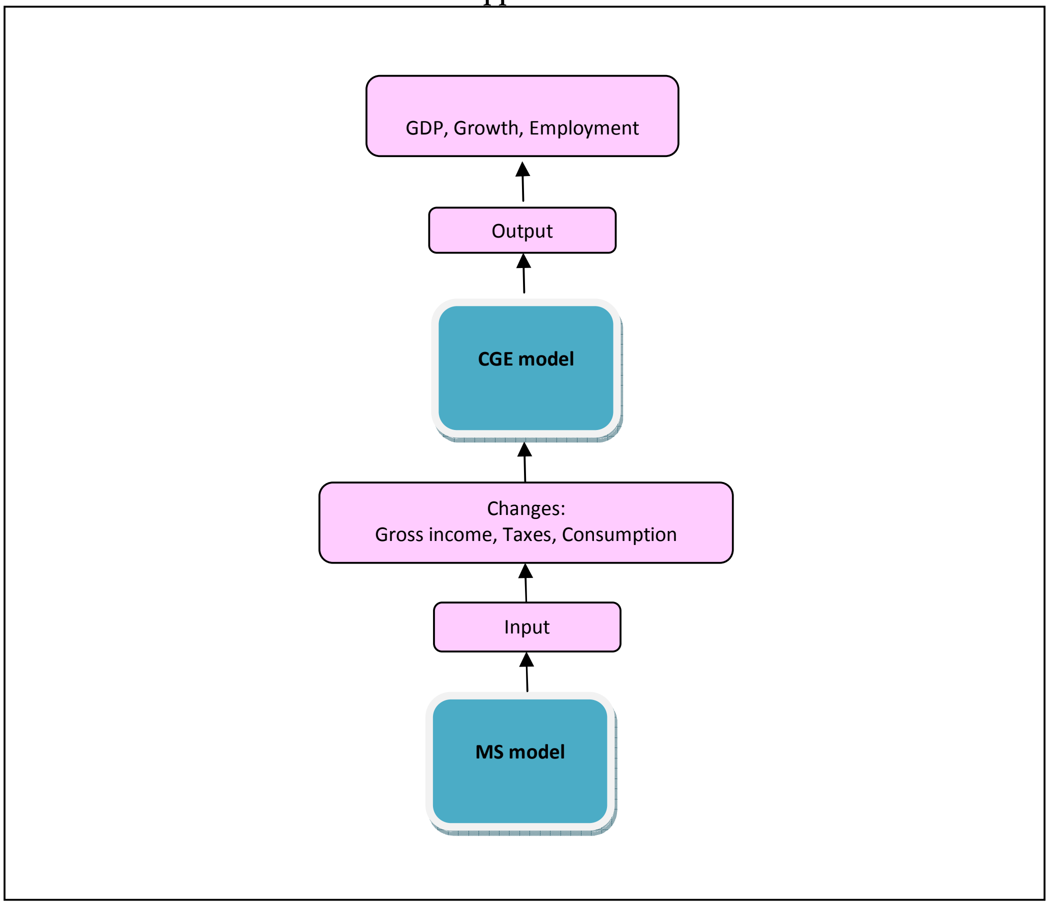
Source: Authors calculation in SAS 9.2, by using a MS model

Figure 1: Structure of an MS model



Source: Citro and Hanushek (1991:2-4).

Figure 2: MS model and Computable General Equilibrium (CGE) model bottom up approach



Source: Peichl, (2008:12).

Figure 3: Fitted and actual plot of PIT