Adapting the South African National Income Dynamics Study for use as a base micro-data set for SAMOD

A Tax-benefit Micro-simulation Model

Gemma Wright[1] , Michael Noble[2] and Mulugeta Dinbabo[3]

UNIVERSITY of the
WESTERN CAPE

Contents

Acronyms

ABSTRACT

In this paper we present the main data issues that were confronted when the first wave of the National Income Dynamics Study (NIDS) 2008 was adapted for use as the micro-data underlying a recently developed static tax-benefit micro-simulation model for South Africa (SAMOD). NIDS is the first national household panel survey in South Africa. SAMOD V1.1, which is based on the EUROMOD platform, was initially underpinned by a dataset derived primarily from the South African Income and Expenditure Survey 2000 but additionally drew on data from several other national surveys. As the NIDS questionnaire is more comprehensive in scope than the IES by, for example, containing questions about intra-household relationships - it is better suited for determining eligibility for some of the existing social assistance arrangements such as the child support grant.

Based on the NIDS micro-dataset, SAMOD simulated plausible figures for eligibility for social assistance and income tax liability. However, indirect forms of taxation were not captured as well as in the IES-based version of SAMOD. As the NIDS questionnaire includes questions about receipt of grants, it was possible to identify the extent to which individuals estimated to be eligible for the grants using SAMOD overlapped with respondents' declared receipt, and how both sets of figures compared to the reported figures of the South African Social Security Agency. Similarly, simulated and reported tax figures were compared against figures from the National Treasury and South African Revenue Service. In a small case study it is demonstrated how SAMOD can be used to quantify the impact on poverty and child poverty of a hypothetical new social grant for low-income adults of working age.

Acknowledgements

This paper relates to a research project that that was funded by the Programme to Support Pro-Poor Policy Development (PSPPD), a partnership programme of the Presidency, the Republic of South Africa and the Delegation of the European Union. The contents of this paper are the sole responsibility of the authors and can in no way be taken to reflect the views of the Presidency (RSA) and the European Union.

The project was a collaboration between the Centre for the Analysis of South African Social Policy (CASASP) at the Oxford Institute for Social Policy (OISP) at the University of Oxford, and the Institute of Social Development (ISD) at the University of the Western Cape (UWC). Prof Pieter Le Roux, Dr Phakama Ntshongwana and Dr Kate Wilkinson are thanked for their contributions to the project and Professor Holly Sutherland and her team at the University of Essex are thanked for their advice on the project.

An earlier version of this paper was presented by Mulugeta Dinbabo at the International Microsimulation Association Conference on Microsimulation and Policy Design, Stockholm, Sweden, 8–10 June 2011, and by Dr Gemma Wright as part of the NATSEM 2011 Seminar Series, National Centre for Economic and Social Modelling, University of Canberra, Australia, 6 July 2011.
 SAMOD was developed using the EUROMOD framework. EUROMOD is a tax-benefit model for the European Union, supported by European Commission funding, developed by a large consortium of European researchers and co-ordinated by Holly Sutherland at the University of Essex.

Centre for the Analysis of South African Social Policy

UNIVERSITY of the WESTERN CAPE

1 INTRODUCTION AND BACKGROUND

This paper provides details about the way in which the first wave of the National Income Dynamics Study (NIDS 2008) was developed in order to become a source micro-dataset for a South African tax and benefit micro-simulation model (SAMOD). The project was undertaken by the Centre for the Analysis of South African Social Policy (CASASP) at the University of Oxford in collaboration with the Institute for Social Development at the University of the Western Cape (UWC).

This section introduces SAMOD, the tax and benefit micro-simulation model that was used by the project team. Section 2 outlines the various steps that were undertaken in order to prepare NIDS as a source micro-dataset for SAMOD. Section 3 presents the findings from tests that were undertaken to check the estimates arising from the NIDS-based model with reported tax and benefit figures and with estimates from an earlier version of SAMOD (SAMOD V1.1) which was underpinned by a different micro-dataset.Section 4 presents a small case study in order to demonstrate how SAMOD can be used to assess the impact of tax and benefit arrangements on poverty and inequality.

Micro-simulation modelling is a technique whereby a set of rules (in this instance, South Africa's tax and benefit arrangements) is applied to individuals as captured in a micro-dataset such as a social survey in order to simulate, at individual level, the impact of the rules on individuals and in aggregate for the population as a whole (see, for example, Mitton,Sutherland & Weeks, 2000; Zaidi, Harding & Williamsm, 2009).

South African examples of tax and benefit micro-simulation research include Adelzadeh (2005 and 2007), Chitiga, Cokburn, Decoluwe, Fofana and Mabugu, (2010), Haarman (2000), Herault (2005), Samson, Babson, Haarman C, Haarman D, Khathi, Mac Quene & Van Niekerk (2002), Lee, Ndlede, Macquene, Van Niekerk, Ganahi, Harigaya & Abrahams, (2002 and 2004), Wilkinson (2009a) and Woolard, Simkins, Oosthuizen & Woolards, (2005).

The micro-simulation model that was used for this project is called SAMOD, which was developed from the EUROMOD platform – please see the box below for further details.

Whilst SAMOD V1.1 was successfully constructed, it was not ideal to have a base micro-dataset that was derived from so many different survey datasets. Not only did it mean that the preparation of the source micro-data was a more complex exercise than it might otherwise have been, but it in fact cannot be repeated as a process because the IES and LFS datasets can no longer be linked. Given these issues, the aim of the project was to assess the extent to which NIDS 2008 could be used as a source micro-dataset for SAMOD and, if possible, to produce an updated version of the model which could be referred to as "SAMOD V1.2". If successful, this would result in a more transparent process of preparing the source micro-data which could be repeated over time as successive waves of NIDS become available. There is of course the concern that NIDS may cease to be representative of the population over time. However, steps may be taken to retain representativeness by refreshing the sample when this becomes an issue (Woolard, Leibbrandt and De Villiers, 2010).

As NIDS 2008 is nationally representative, the aggregate national impact of tax and benefits on the income distribution can be calculated. In addition to simulating existing tax and benefit rules, new tax and benefit rules can be simulated and, using appropriate weights, their cost and financial impact on individuals and the population at large can be estimated.

The following section explains the methodological processes that were undertaken to prepare NIDS as the source micro-data set for SAMOD.

Box 1 Relationship of SAMOD to EUROMOD

SAMOD Version 1.1 was produced by the Centre for the Analysis of South African Social Policy (CASASP) at the University of Oxford for the South African Government Department of Social Development between 2007 and 2010. The development of SAMOD V1.1 was funded by the UK Department for International Development Southern Africa as part of its Strengthening Analytical Capacity for Evidence-based Decision-making (SACED) Programme. SAMOD V1.1 was developed from the EUROMOD Version D17 platform, which was made available to the research team by the EUROMOD team at the University of Essex. EUROMOD is a tax and benefit micro-simulation model for the European Union (EU) (Sutherland, 2001; Sutherland, Figari, Lelkes, Levy, Leitz, Mantovani & Paulus, 2008). EUROMOD was developed, and continues to be developed, in order to enable the measurement of the effects of taxes and benefits on household incomes and work incentives for the population of each country in the EU and across the EU (see <www.iser.essex.ac.uk/research/euromod>, for further details about EUROMOD including links to country reports and working papers).SAMOD is jointly owned by the University of Oxford, the Department of Social Development of the Republic of South Africa, and the University of Essex.

As with EUROMOD, SAMOD's user interface is in EXCEL and has been designed to maximise the flexibility of the model for the user. For example, the user can define and modify income and tax unit concepts, as well as modify or eliminate existing tax and benefit arrangements or introduce new policies.

SAMOD is a static tax and benefit micro-simulation model. This means that the original incomes and characteristics of the population in the micro-dataset are held constant. It is possible to measure the "next day" impact on poverty and inequality of specified changes to policies. SAMOD does not itself estimate behavioural responses to policy changes although it can be used to provide information on incomes if relevant behaviour (such as being in paid work) changed.[4]

The intention behind the development of SAMOD was that it could be used in-house by the South African government as well as within academia in order to increase the number of civil servants and researchers able to use micro-simulation techniques in South Africa who could work on this or other models.

SAMOD V1.1 was completed in 2010, and contains tax and benefit rules (or "systems") for 2007, 2008, 2009 and 2010, and a user manual was produced (Wilkinson, Wright & Noble 2009). SAMOD V1.1 is underpinned by a source micro-dataset relating to 2007 which was constructed from a number of different datasets including the Income and Expenditure Survey (IES) 2000 and the IES 2006/7, the Labour Force Survey (LFS) 2000 and LFS 2006, and the 2007 Community Survey 2007. Wilkinson (2009a) provides a detailed account of how SAMOD V1.1 was constructed, but in summary the LFS 2000 data was linked to the IES 2000 data which was then up-weighted using more recent data, including 2007 population estimates produced by the Actuarial Society of South Africa, to a 2007 time point.

While SAMOD V1.1 was successfully constructed, it was not ideal to have a base micro-dataset that was derived from so many different survey datasets. Not only did it mean that the preparation of the source micro-data was a more complex exercise than it might otherwise have been, but it in fact cannot be repeated as a process because the IES and LFS datasets can no longer be linked. Given these issues, the aim of the project was to assess the extent to which NIDS 2008 could be used as a source micro-dataset for SAMOD and, if possible, to produce an updated version of the model which could be referred to as "SAMOD V1.2". If successful, this would result in a more transparent process of preparing the source micro-data which could be repeated over time as successive waves of NIDS become available. There is, of course, the concern that NIDS may cease to be representative of the population over time. However, steps may be taken to retain representativeness by refreshing the sample when this becomes an issue (Woolard *et al*., 2010).

As NIDS 2008 is nationally representative, the aggregate national impact of tax and benefits on the income distribution can be calculated. In addition to its simulating existing tax and benefit rules, new tax and benefit rules can be simulated and, using appropriate weights, their cost and financial impact on individuals and the population at large can be estimated.

The following section explains the methodological processes that were undertaken to prepare NIDS as the source micro-dataset for SAMOD.

4 However, some inference on behavioural changes may be inferred using, for example, changes in marginal effective tax rates after a policy reform.

2 METHODOLOGY

The project began by exploring the NIDS dataset, questionnaires, metadata (Saldru, 2009), technical papers and discussion papers (available at http://www.nids.uct.ac.za) and recent analysis of NIDS (Leibbrandt *et al*., 2010; Smit, 2010). The NIDS data comprise seven different files which had to be combined: the adult, child, proxy, derived individual, household, derived household, and household rosterfiles (Saldru, 2009). A single flat file was created comprising a case for each adult, child or proxy, so that each row contained the data relating to that particular individual. The relevant individual derived data, household data, househ old derived data, and household roster data were then appended to the end of each row. This resulted in a total of 28,255 cases or individuals at the outset, and the ability to access derived or household information for each individual within a single flat file.

Using the NIDS questionnaires and metadata, variables were identified that could be used to generate each input variable for SAMOD; when alternatives where identified, the different options were compared and tested. In some instances a single variable in NIDS could be used as the required input variable; in other instances it was necessary to use a number of NIDS variables in order to construct a particular input variable. The policies that were simulated using input variables from the NIDS micro-dataset are listed in Figure 1 below. The intention was that the SAMOD policies should reflect the tax and benefit rules in existence in 2008 so that grant eligibility and tax liability can be simulated as accurately as possible (Republic of South Africa, 2004; 2008; Black Sash, 2010). The tax and benefit policies are discussed in more detailbelow.

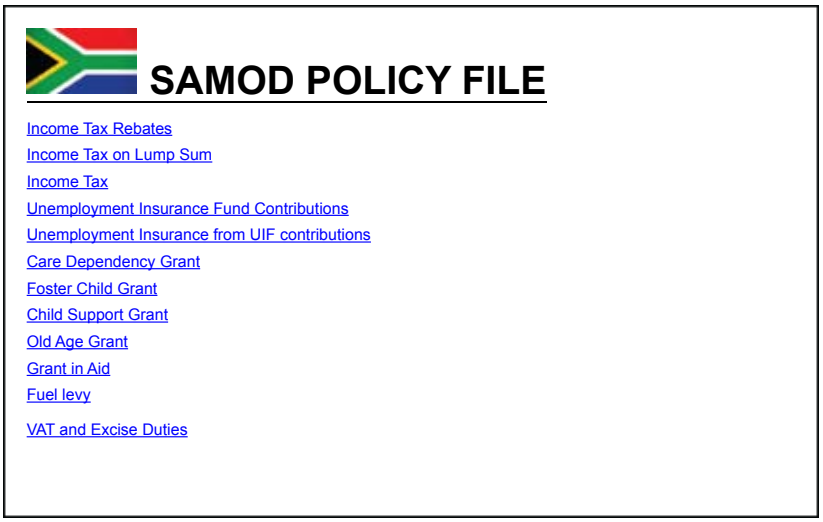
A number of policies could not be simulated within SAMOD due to lack of data in NIDS. This was also the case when the IES was the underpinning dataset (see Wilkinson, 2009a).These include War Veteran's Grant, Social relief of distress, compensation for occupational injuries and diseases, compensation for road traffic accidents, motor vehicle tax, road toll fees, private medical and retirement schemes, corporate taxation, capital gains tax and transfer duties. More generally, a number of government subsidised services exist which impact on people's incomes but are not included in the model. These include free electricity allowances, free water allowances, no-fee schools and total or partial school fee exemptions, free school meals, municipal indigent grants, early childhood development subsidies, and free primary health care. Entitlement sometimes varies by area and quality often does so. No clear definition exists of the social wage; however, these items would probably fall within its scope and would potentially have an impact on poverty if taken into account (May, 2004; Meth, 2008).

The requisite demography variables (for example, age, population group, sex) were created, as well as a composite employment status variable. Benefit and tax input variables were constructed, in order to identify reported recipients of social grants and payment of income tax, so that these figures could be compared with simulated eligibility for grants and income tax liability. A range of income variables were constructed which were needed in order to simulate the means tests for the grants and income tax liability. Expenditure variables were constructed in order to simulate value-added tax (VAT), excise duty and fuel levy.

As well as ensuring that the NIDS-based input variables in the micro-dataset would be recognised by the SAMOD "engine", it was also necessary to ensure that they were correctly located in the SAMOD user-interface which comprises a set of linked Excel worksheets. For example, the new NIDS-specific variables were incorporated into the variable description file, and a NIDS-specific income list was also created–income lists determine which variables are taken into account in different income concepts,for example, employment income, taxable income, income taken into account when means-testing, items that are liable for VAT- or are VAT-exempt or zero-rated. The resultant NIDS 2008 system comprises the tax and benefit rules for 2008 which have been expressed in such a way that they can be implemented using data available within NIDS.

In summary, with the exception of the expenditure data, it proved possible to create the same set of variables for SAMOD using NIDS variables as had been constructed previously for SAMOD V1.1 using a combination of the IES, LFS and Community Survey data (see Wright, Noble, Dinbabo, Neshongwana, Wilkinson & Le Roux, (2011) for further details about the methodological processes that were undertaken). The next section explores the extent to which the NIDS micro-dataset enabled the policies to be simulated adequately.

Figure 1a: List of SAMOD policies



3 TESTING THE RESULTS

Before a model can be used to simulate new policies, it is first necessary to be able to simulate existing policies for the relevant data time point, in this instance 2008. As well as preparing the input data, a main part of the project entailed testing the output data, having run the model, to ensure that the simulations were plausible and, where necessary and possible, making adjustments to the way in which the input variables were constructed.

The NIDS fieldwork began in early February 2008, with the initial round concluding in July 2008 and the second phase taking place after that point (Woolard *et al*., 2010). As far as possible, tax and benefit rules relating to the middle of 2008 (June/July 2008) were applied. However, dowing to the extended timeframe of the fieldwork and a number of changes to the benefit rules during the fieldwork period, there is no optimal time point with which to make direct comparisons with reported tax and benefit data. In any event, it proved difficult to obtain South African Social Security Agency (SASSA) figures for mid-2008; the earliest time point after the fieldwork for which figures could be obtained was the end of February 2009 (SASSA, 2009). Reported tax data relating to 2007/08 and 2008/09 were used (National Treasury and SARS, 2008; 2009).

As well as comparing simulated totals of grant eligibility against SASSA's actual figures and simulated income tax figures against the National Treasury's actual tax take, it was additionally possible to compare the simulated figures against the weighted reported receipt of grants and payment of income tax based on the actual responses of the NIDS survey respondents. Comparisons are also made in this section with the simulated grant and tax figures using the IES-based micro-dataset.

Social Assistance

Table 1 provides figures for social assistance. Column B shows the number of recipients of each grant as reported by SASSA (SASSA, 2009). Column C shows the number of people identified as receiving the grants in NIDS, in other words, NIDS respondents who reported receipt of grants, using post-stratified weights. Column D shows the number of people estimated by SAMOD to be eligible for each grant, using the NIDS dataset. Column E shows the estimated take-up rate for each grant, which is the number of actual recipients divided by the number estimated by SAMOD to be eligible.

So, for example, SASSA reported 8.65m recipients of the child support grant (CSG). This is very close to the number of reported recipients of CSG in NIDS (8.36m using the variable "grcurtyp" and the post-stratified weights). Using SAMOD it was estimated that there were 10.15m eligible children, and so the take-up rate for CSG would be 85%.

Similarly, for the old age grant (OAG), there were 2.39m recipients according to SASSA and 2.43m according to NIDS. Using SAMOD it was estimated that 2.66m people were eligible for OAG, resulting in a take-up rate of 90%.

Overall, the weighted counts of self-reported recipients in NIDS (column C) are very close to SASSA's figures (column B) for each of the grants. The weighted count of self-reported recipients in NIDS amounts to 97% of SASSA actuals for CSG, 65% for FCG, 62% for the care dependency grant (CDG)[5] , 102% for OAG and 100ct% for DG. Receipt of GIA does not appear to have been asked about in NIDS.

5 CSG, FCG and CDG were measured using grcurtyp in the child questionnaire.It should be recalled that FCG and CDG can be claimed up to 18, and FCG can be extended beyond 18 if the child is still in school; these older children will not have been captured in the child questionnaire which is restricted to people aged 0–14 inclusive, which would go some way to explaining the lower take-up rates reported here for FCG and CDG in terms of NIDS self-reported figures as a percentage of SASSA actuals. Reported receipt of OAG was measured using incgovpen and incgovpen_v, and receipt of DG was measured using incdis and incdis_v. The data about grant receipt in the individual derived file generally compared less favourably against SASSA actuals than the non-derived data. The individual derived data file captured the following proportions of SASSA's actual figures: 65% for CSG, 52% for FCG, 62% for CDG, 106% for OAG and 100% for DG.

Some of the policy rules were more difficult to simulate than others using the data available in NIDS. For the FCG, children are eligible only if they have been placed in the care of foster parents by a court order; however this criterion could not be measured using the NIDS data. In practice, some primary caregivers of children whose parents have both died ("double orphans") also receive FCG as a source of financial support (Hall, 2010; Meintjes, Budlender, Giese & Johnson, 2003). As a proxy, children were identified as potentially eligible if their primary caregiver was defined in NIDS as a foster parent (43 cases in the child questionnaire) or if both of their parents were dead.On this basis, 700 200 children were estimated to be eligible for FCG, resulting in a take-up rate of 69%. However, at best this is only an approximation: some children with live parents will be eligible for FCG if they have had a court order to remove them from their parents for their own safety, and some children with dead parents will be in receipt of CSG and not FCG.

It also proved difficult to simulate eligibility for the disability-related grants precisely in SAMOD as the eligibility criteria include a health test by a medical professional. Apart from the age thresholds and means tests, eligibility therefore mainly had to be determined in advance of running the model, drawing from a number of different questions in NIDS. Adults were identified as potentially eligible for the disability grant if the main reason given for unavailability for employment in the last four weeks was given as "I am sick/disabled" (noemex==3). Adults were also captured if they described their health as poor (hldes==5) and additionally reported that they had one or more illnesses or disabilities (physically handicapped, HIV/Aids, Epilepsy/fits, Emphysema, Alzheimer's disease, other) or that they currently had a stroke or TB. Proxies were included if the reason for conducting a proxy survey was given as "permanently too unwell or disabled" (resprox ==6), or if the current activity of the proxy was "long term sick or disabled" (emp==7). Based on these criteria and the age thresholds and means tests, using SAMOD, it was estimated that 1.5m people were eligible for the DG, compared to the 1.3m SASSA recipients, resulting in an estimated take-up rate of 86%.

Different questions were used to identify disabled children for whom Care Dependency Grant (CDG) could be claimed. Questions D10_1 to D10_3 in the child questionnaire were used in order to determine whether the child has a serious illness or disability. However, this is a problematic question as the exact wording is 'has the child had any serious illnesses or disabilities?'. In theory, therefore, the child could have had the illnesses or (non-permanent) disabilities in the past and since recovered from them. A requirement was therefore added in that the child had been ill for at least 3 days in the last 30 days which reduced the number of children captured to a fifth of the original figure. As the child survey only covers 0–14 year olds, children with a serious illness or disability aged 15–17 inclusive were captured from the other questionnaires using the same code as for the adults. On this basis, using SAMODit was estimated that 165 600 people would be eligible for CDG, resulting in a take-up rate of 64%.

Grant-in-aid was not asked about directly in NIDS. However, there were questions available with which to estimate probable candidates for this grant. People were identified who reported that they could not dress, or bath, or eat, or go to the toilet by themselves, or could neither do light housework nor climb a flight of stairs by themselves. Based on this information, SAMOD captured 149 463 eligible people, resulting in a take-up rate of just 30%.

Interpreting the differences in Columns B, C and D in Table 1

In addition to the time point issue referred to at the beginning of Section 3, there are many reasons why the figures in columns B, C and D in Table 1 might differ. The SASSA figures in column B will contain eligible recipients and an unquantifiable number of ineligible recipients (sometimes referred to as an error of inclusion, vertical inefficiency, or a Type 2 error). The SASSA figures will not include eligible non-recipients (sometimes referred to as an error of exclusion, or horizontal inefficiency, or a Type 1 error). SASSA's figures also rely on the ability of the SOCPEN system to capture all live cases in payment, and the ability of SASSA to correctly target and administer the social grants.

The NIDS figures in column C will contain eligible recipients but are less likely to contain ineligible recipients than in Column B. Any people who are purposefully misusing the system would be unlikely to report their receipt in a survey, so ineligible recipients are more likely to comprise people who are unaware of the fact that they are ineligible (this could be caused, for example, by an increase of income at the time of the survey which took the person above the means test threshold). Again, Column C will not include eligible non-recipients. The figures in Column C will also depend on the extent to which grant recipients were included in the survey sample, the extent to which the post-stratified weights adequately represent the national profile of grant recipients, and the extent to which NIDS respondents accurately reported grant receipt.

The simulated figures in Column D should contain all eligible recipients of social grants. However, these figures depend on the accuracy of the income and relationship data in the NIDS dataset, the adequacy of

the survey weights, and the ability to translate the grant eligibility rules into SAMOD commands in the policy files.

In order to explore this issue in more detail, analysis can be undertaken to unpick the data in Columns C and D of Table 1in order to explore the extent to which there was overlap at an individual level between reported receipt by the NIDS respondent and simulated eligibility by SAMOD. As an example, it is possible to ask: what proportion of the NIDS respondents who reported OAG receipt were identified using SAMOD as eligible for OAG? The results are presented in Table 2 below. Thus, 85% of those who reported receipt of OAG were identified as eligible for OAG by SAMOD. If the weighted NIDS dataset perfectly captures a nationally representative sample of OAG recipients, and perfectly captures income and relationship data, and if SAMOD perfectly simulates the OAG policy rules (none of which is likely to be the case, for the reasons outlined above) this means that 15% of the reported OAG recipients are ineligible. Table 2 also shows that 99% of those who did not report in NIDS that they are in receipt of OAG were estimated using SAMOD to be ineligible for OAG.[6]

Conversely, one can ask: what proportion of the people identified as eligible for OAG using SAMOD actually reported in NIDS that they are in receipt of OAG? The results are presented in Table 3 below.

Of those identified as eligible for OAG by SAMOD, 76% reported receipt of OAG. Again, if NIDS perfectly captures receipt of OAG and if SAMOD perfectly simulates eligibility (again unlikely to be the case), this means that 24% of those eligible for OAG are not yet receiving it, resulting in a take-up rate of 76%, which is rather lower than the take-up rate presented in Table 1. Of those identified as ineligible for OAG by SAMOD, over 99% did not report that they were receiving OAG.

This is a useful exercise as it helps to shed some light on the extent to which one can rely on different sources of information about grant receipt and eligibility. By comparing survey data, administrative data, and simulations, the various discrepancies can be explored in detail and over time. The findings presented here do not enable one to draw any particular conclusions, but they do highlight the need to scrutinise such issues further. In the UK, for example, social assistance take up rates are calculated by the government, using administrative data to measure actual receipt and survey data to measure eligibility, owing to the difficulty in accurately capturing social assistance receipt in surveys (DWP, 2010).

Table 1: Social assistance

Grant A	Actual recipients (SASSA) B	Self-reported recipients (NIDS) C	SAMOD-simulated eligibles (using NIDS) D	Estimated take-up rates (Actuals/ SAMOD-simulated eligibles) % E
Child Support Grant	8 651 959	8 357 511	10 153 273	85.2
Foster Child Grant	482 148	312 457	700 207	68.9
Care Dependency Grant	106 401	65 597	165 608	64.2
Old Age Grant	2 392 984	2 432 649	2 657 899	90.0
Disability Grant	1 292 571	1 287 787	1 497 840	86.3
Grant-in-aid	45 177	Not asked about	149 463	30.2

Table 2: Comparing reported receipt of Old Age Grant in NIDS with simulated eligibility

Reported receipt of OAG?	% ineligible (SAMOD)	% eligible (SAMOD)	Row total
Yes	15.0	85.0	100
No	98.6	1.4	100

6 This table relates to the total population (all ages) in order to include recipients who may fall outside the age thresholds. Therefore, the finding that an estimated 1.4% of those who did not report receipt of OAG are estimated using SAMOD to be eligible for OAG does not conflict with the findings in Table 1 (which shows an estimated 90% take-up of OAG).

Comparing SAMOD V1.1 and SAMOD V1.2

Finally for this sub-section, Table 4 shows the simulated number of people eligible for the three largest grants–CSG, OAG and DG–using the original micro-dataset used by SAMOD V1.1 (Wilkinson, 2009a) and the new NIDS-based micro-dataset. The former was calculated using microdata relating to 2007, whereas the latter relates to 2008.[7] In both cases, however, the tax and benefit system applied in SAMOD related to 2008.

Using two totally different source micro-datasets, with a one year time difference between them, the estimated number of eligible people for these three grants is remarkably similar. Unfortunately, one cannot draw conclusions from this table unless one were to try to disentangle the impact of real change between the two time points (for example, changes in people's incomes, demographics, accuracy of administering these grants) from reliability of the two source micro-datasets. It is, however, possible to eliminate the impact of the change in policies between 2007 and 2008, as 2008 policies were applied to both datasets. Nevertheless, one can at least note that the NIDS estimates are within 2.1% of the original IES estimates for each of these grants.

Taxation

In 2007/08 Government raised 27.8% of its GDP through national taxes, totalling R572.9bn (National Treasury and SARS, 2008:1–2).[8] In 2008/09 the national tax revenue rose to R625.1bn. The main tax revenue sources are personal income tax, corporate income tax, and value-added tax (VAT); these made up approximately 80% of national tax revenue in 2007/08 (National Treasury and SARS, 2008:1). Table 5 shows the main sources of tax revenue and the tax take for 2007/08 and 2008/09. In descending order of size, the largest sources of tax in terms of percentage of GDP in 2007/08 were personal income tax (8.2%), VAT (7.3%) and Corporate Income Tax (6.8%); each of the other sources amounted to between 0.9 and 1.3% of GDP (National Treasury and SARS, 2008:17).

Table 3: Comparing simulated eligibility for Old Age Grant with reported receipt in NIDS

Eligible for OAG?	% receiving OAG	% not receiving OAG	Row total
Yes	75.9	24.1	100
No	0.8	99.2	100

Table 4: Simulated number of people eligible for social grants using IES and NIDS

Grant	Simulated number of eligibles (SAMOD V1.1, using the IES upweighted to 2007)	Simulated number of eligibles (SAMOD V1.2 using NIDS 2008)	% difference (NIDS/IES)
Child Support Grant	10 373 121	10 153 273	97.9
Old Age Grant	2 702 078	2 657 899	98.4
Disability Grant	1 477 351	1 497 840	101.4

7 It should be noted that we did not reweight the 2007 data used in SAMOD V1.1 to 2008 in any way for the purposes of this comparison, however the policies applied to both datasets are identical.

8 In addition to national taxes, there are provincial and local government taxes, but these are relatively small, amounting respectively to less than 1% and 3.6% of the total South African tax revenue in 2007/08 (National treasury and SARS, 2008: 1–2).

Apart from personal income tax, taxes are mainly not captured by SAMOD, as SAMOD is based on a household survey and so will only ever capture household (not business) expenditure. Attempts were made to simulate personal income tax, VAT, fuel level and excise duties, in the knowledge that the only policy for which close to the full tax take could be simulated was personal income tax. As will be elaborated below, only personal income tax worked satisfactorily in SAMOD V1.2.

The variables that were used to capture income are listed in Annex 1of the project report (Wright *et al*., 2011). Whenever possible, the individual (rather than the household) income variables were used, in order to assign incomes to individuals within the household most accurately. The non-derived variables were also given priority, as the derived variables were mainly constructed to be net of income tax (Argent, 2009), whereas the gross figures were required here in order to be able to simulate income tax liability. Overall, 6.7% of adults and proxies did not respond to the income questions and so UCT undertook extensive imputation work on the income data (Argent, 2009). Where income data was missing/ refused/unknown at the individual level, the derived income variables were used. This will result in a partial undercount of the gross income, but was the best option available within the scope of the project.

The NIDS-based version of SAMOD captures 5.2m taxpayers. At this level of aggregation this is the same as the National Treasury and SARS figure for registered taxpayers for 2007/08 (5.2m individuals) (National Treasury and SARS, 2008, p3)[9], and slightly less than the NT and SARS figures for 2008/09 of 5.54m individuals (NT and SARS, 2009 p6).

As can be seen in Table 6 below, NIDS V1.2 simulated a personal income tax take of R197Bn. This amounts to 101% of the reported tax take for 2008/9 (NT and SARS, 2009).This compares very favourably to the SAMOD V1.1 figures where only 69% of the 2008/09 income tax take was simulated. The NIDS-based simulated income tax figure (R197Bn) is much larger than the total reported income tax paid by NIDS respondents (R19.3Bn), suggesting that it is a very difficult survey question for respondents to answer accurately, or that respondents have difficulty in distinguishing net and gross incomes.

Table 5: Sources of tax revenue in 2007/08 and 2008/09

Sources of tax revenue	2007/08 R million	2008/09 R million
Personal income tax	168 774.4	195 115.0
Corporate income tax	140 119.8	165 378.3
Secondary tax on companies	20 585.4	20 017.6
Value-added tax	150 442.8	154 343.1
Fuel levy	23 740.5	24 883.8
Customs	27 081.8	22 852.4
Specific excise duties	18 218.4	20 184.5
Other	23 907.4	22 325.4
Total	**572 870.6**	625 100.2

Source: NT and SARS, 2008 p.16 for 2007/08; NT and SARS, 2009 p34 for 2008/09.

9 The document cautions that "it must be noted that the figures for registered taxpayers for income tax purposes exclude an estimated 4 million formal employees (standard income tax on employees (SITE)-only taxpayers) earning an annual taxable income below R60 000 from employment only, who are not required to register with SARS." (National Treasury and SARS, 2008 :3). These are people who are earning so little that they are not paying income tax.

The other types of tax shown in Table 6 are much less satisfactory, as SAMOD V1.2 is simulating only around a fifth of the reported figures for these types of tax income. Only 19% of VAT is captured, 21% of excise duty and 20% of fuel levy. Although not all expenditure in South Africa is made by households, and therefore much of the shortfall will legitimately be made up by non-household expenditure, the size of the shortfall also suggests that there is under-reporting of expenditure in NIDS. Indeed, Finn, Franklin, Keswell, Leibbrandt & Levinsohn (2009) report that although there were quite high response rates, more than 40% of all households had at least one case of missing expenditure data. Although imputation work has been undertaken by UCT on the expenditure data, this could not drawn upon here as the imputed variables are too composite. So, for example, though UCT has created a variable for food expenditure with full imputations and another for non-food expenditure with full imputations, this does not enable one to distinguish between VAT-able, zero-rated and VAT-exempt items[10]. In order to improve this result, a fresh set of imputations would need to be undertaken on each of the expenditure variables, or some approximation used derived from the IES data, but these options were beyond the scope of the project.For the purposes of the project, therefore, it was necessary to prioritise focus on personal income tax and the various forms of social assistance listed in Figure 1 above.

Table 6: Direct and indirect taxation

Tax A	Actuals 2008/09 B	SAMOD V1.2 (NIDS) C	Extent of capture (C/B) % D
Income tax	195 115 000 000	197 100 000 000	101.0
VAT	154 343 100 000	29 810 000 000	19.3
Excise	20 184 500 000	4 241 000 000	21.0
Fuel levy	24 883 800 000	5 028 000 000	20.2

10 See National Treasury and SARS, 2008: 116-118 for a discussion about the efficacy of VAT zero-rating as a pro-poor intervention.

4 A SMALL CASE STUDY

SAMOD V1.1 has been used to assess the impact of tax and benefit policies on child poverty (Wilkinson, 2009b) and the impact of possible new grants on poverty and inequality with a particular focus on lone parents (Ntshongwana, Wright & Noble 2010) and children (Dinbabo, 2011).

In this section the version of SAMOD is used that is underpinned by the NIDS micro-dataset to measure the impact on poverty of introducing a hypothetical "income replacement grant" (IRG) as an additional policy.

The IRG was designed so that it was targeted at people who are healthy (that, is not in receipt of the disability grant), aged16–59 inclusive for women and 16–62 inclusive for men, and have a low income (applying the same means-test as for the OAG). The grant was set at R200 per person per month.

In Table 7 below, Hoogeveen and Özler's (2006) lower- and upper-bound poverty lines are used in order to present the impact on poverty of the IRG. The figures are presented for a situation with no grants, the 2008 tax and benefit system simulated by SAMOD (assuming 100% take-up), and the 2008 tax and benefit system simulated by SAMOD plus the IRG (again assuming 100% take-up).

Applying the lower bound poverty line of R515 per person per month, 59% of people would fall below the poverty line if there were no grants. Using SAMOD's simulated grants and assuming 100% take-up, the poverty rate would fall to 52%. [11]The introduction of an IRG would further reduce the poverty rate by five percentage points to 47%.

Using the upper poverty line of R949 per person per month, the poverty rate would fall less dramatically from 68% (with no grants) to 64% (simulated 2008 grants plus IRG).

It has been demonstrated elsewhere that existing social grants for adults – in particular the Disability Grant and Old Age Grant - are more effective at reducing child poverty than the Child Support Grant, as the CSG is paid at a much lower rate than the adult grants (e.g. Hall and Wright, 2010). To what extent does this hypothetical IRG help with further reducing levels of poverty in households containing children? Table 8 presents analysis for people living in households containing one or more children. Four scenarios are compared: a situation with no grants; a situation with 2008 grants excluding CSG; a situation with 2008 grants (i.e. including CSG); and a situation with 2008 grants plus the IRG.

Table 7: Impact on poverty of an income replacement grant

Total population	Lower poverty line (R515 pcm)			Upper poverty line (R949 pcm)		
	P^0	P^1	P^2	P^0	P^1	P^2
No grants	0.59	0.50	0.46	0.68	0.56	0.51
2008 grants (simulated)	0.52	0.30	0.22	0.66	0.44	0.33
2008 grants plus IRG (simulated)	0.47	0.21	0.11	0.64	0.37	0.25

11 Leibbrandt *et al*. (2010: 35) state that 54% of people were in poverty using the same poverty line and taking into account income from reported receipt of grants in NIDS. This falls between the figures shown in Table 7 of 59% (no grants) and 52% (simulated full take-up of 2008 grants).

Table 8 shows that poverty rates are higher among people living in households that contain one or more children, than for the population as a whole (which was shown in Table 7 above). Using the lower poverty line of R515 per capita per month, we estimate that 65% of people living in households containing one or more child (including the children themselves) would be in poverty if there were no grants, compared to 59% for the population as a whole.

By introducing all of the 2008 grants apart from the CSG, this falls to 62%; if CSG is additionally included, the poverty rate falls to 58%. This demonstrates the impact of the CSG, based on a scenario of full take-up. If CSG had not existed in 2008 it is estimated that the poverty rate (still using the R515 line) would have been four percentage points higher (at 62%) for people in households containing children. The poverty rate falls a further five percentage points with the introduction of an IRG, down to 53%. Even though the hypothetical IRG is paid at the relatively low amount of R200 per month, the children would nevertheless benefit in terms of the resources entering the household.

Table 8: Impact on child poverty of an income replacement grant

People living in households containing one or more children	Lower Poverty Line (R515 pcm)			Upper Poverty Line (R949 pcm)		
	P^0	P^1	P^2	P^0	P^1	P^2
No grants	0.65	0.54	0.50	0.74	0.61	0.56
2008 grants (simulated, without CSG)	0.62	0.41	0.33	0.73	0.54	0.44
2008 grants (simulated)	0.58	0.33	0.22	0.72	0.48	0.36
2008 grants plus IRG (simulated)	0.53	0.23	0.12	0.71	0.41	0.28

5 CONCLUSION

It has been possible to convert the NIDS data into a source micro-dataset for SAMOD. In general, the NIDS data contained the requisite variables in order to simulate tax and benefit policies. The benefits were largely quite successful, although the disability-related and foster child grants required a certain amount of pre-model preparation of flags for disability status and potential foster child status using proxy questions available in the survey data. The amount of simulated personal income tax came very close to the 2008/09 figures released by the National Treasury and SARS. However, only about a fifth of the main indirect taxes were simulated.

There are a number of ways in which this work could be developed further, both methodologically and for policy purposes. First, just as new systems for 2008, 2009 and 2010 have been created for the IES-based dataset, it would be possible to produce NIDS-compatible systems for 2009 and 2010, thereby enabling more up-to-date tax and benefit policies to be applied to the data.

In principle, the NIDS micro-dataset could also be reweighted (for example, using ASSA data and more up to date income/expenditure data) to relate to 2009 and 2010. However, it is likely that the forthcoming 2008/09 Living Conditions Survey will be more amenable to being reweighted in this way as it is a much larger survey.

Third, the model can be used to simulate modified or new policies, and the impact on poverty and inequality of such changes can then be measured by analysing SAMOD's output data.

Fourth, it will be possible to update the microdata when the next wave of NIDS becomes available. This data will relate to 2010 and is due to be released in 2012. The longitudinal nature of NIDS does have some potential advantages. For example, for unemployed people it would be possible in Wave 2 to find out whether they were employed previously and in what occupation, and therefore enhance the model in relation to eligibility for Unemployment Insurance Fund.

Lastly, as part of the project, a course on micro-simulation was run for students at the University of the Western Cape as part of the methods course for the Masters in Development Studies. Eight students took part in this course and several of them have expressed interest in focusing on micro-simulation in their subsequent studies. There is great scope for using SAMOD for further Masters and Doctoral level study.

References

Adelzadeh, A. (2005). *The South African Tax and Transfer Simulation Model (SATTSIM)*. World Institute for Development Economics Research. See http://african-models.wider.unu.edu .

Adelzadeh, A. (2007). *Halving poverty and unemployment in South Africa: Choices for the next ten years,* Oxfam Research Report.

Argent, J. (2009). *Household income: report on NIDS Wave 1*, NIDS Technical Paper No.3, Cape Town: Saldru, University of Cape Town.

Black Sash (2010). *Social Assistance: A reference guide for paralegals,* Mowbray: Black Sash.

Chitiga, M., Cockburn, J., Decaluwe, B., Fofana, I. and Mabugu, R. (2010). 'Case Study: A Gender-focused Marco-Micro Analysis of the Poverty Impacts of Trade Liberalization in South Africa', *International Journal of Microsimulation,* 3(1): 104–108.

Department of Work and Pensions (DWP); (2010). *Income related benefits: estimates of take up in 2008-09,* London: DWP.

Finn, A., Franklin, S., Keswell, M., Leibbrandt, M. and Levinsohn, J. (2009). *Expenditure: Report on NIDS Wave 1,* NIDS Technical Paper No.4, Cape Town: Saldru, University of Cape Town.

Haarman, C. (2000). *Social assistance in South Africa: Its potential impact on poverty*, PhD thesis. University of the Western Cape.

Hall, K. (2010). *Income and Social Grants – Take-up of the Foster Child Grant*, Children Count Fact Sheet,Cape Town: Children's Institute.

Hall, K. and Wright, G. (2010). 'A profile of children living in South Africa in 2008', *Studies in Economics and Econometrics,* 34(3): 45–68.

Herault, N. (2005). *A Micro-Macro Model for South Africa: Building and Linking a Microsimulation Model* to a CGE Model, Working Paper No. 16/05, Melbourne Institute, University of Melbourne.

Leibbrandt, M., Woolard, I., Finn, A. and Argent, J. (2010). *Trends in South African Income Distribution and Poverty Since the Fall of Apartheid,* OECD Social, Employment and Migration Working Papers No. 101, Paris: The Organisation of Economic Co-operation and Development.

May, J. (2004). 'Poverty, Social Policy and the Social Wage', paper presented at a conference on 'The Politics of Socio-Economic Rights in South Africa: 10 Years after Apartheid', 8–9 June 2004, Oslo, Norway.

McEwen, H., Kannemeyer, C. and Woolard, I. (2009). *Social Assistance Grants: Analysis of the NIDS Wave 1 Dataset,* NIDS Discussion Paper No. 10, Cape Town: Saldru, University of Cape Town.

Meintjes H., Budlender, D., Giese, S. and Johnson, L. (2003). *Children 'in need of care' or in need of cash? Questioning social security provisions for orphans in the context of the South African AIDS pandemic.* Joint working paper of the Children's Institute and the Centre for Actuarial Research, University of Cape Town. Available at: www.ci.org.za/depts/ci/pubs/pdf/hiv/workp/fostercare/foster_care.pdf

Meth, C. (2008). *Social income in South Africa: an economy marred by high unemployment, poverty and extreme inequality,* Pretoria: Human Sciences Research Council.

Mitton, L., Sutherland, H. and Weeks, M. (Eds.) (2000). *Microsimulation modelling for policy analysis,* Cambridge: University Press.

National Treasury and South African Revenue Service (2008). 2008 *Tax Statistics*, Pretoria: National Treasury and SARS.

National Treasury and South African Revenue Service (2009). 2009 *Tax Statistics*, Pretoria: National Treasury and SARS.

Ntshongwana, P., Wright, G. and Noble, M. (2010). *Supporting Lone Mothers in South Africa: Towards Comprehensive Social Security*, Pretoria: Department of Social Development.

Republic of South Africa (2004). *Social Assistance Act of 2004 (Act No. 13 of 2004).*

Republic of South Africa (2008). *Social Assistance Amendment Act of 2008 (Act 6 of 2008).*

Samson, M., Babson, O., Haarman, C., Haarman, D., Khathi, G., Mac Quene, K. and van Niekerk, I. (2002) *Research Review on Social Security Reform and the Basic Income Grant for South Africa,* EPRI Research Report No. 31, International Labur Organisation and the Economic Policy Research Unit.

Samson, M., Lee, U., Ndlebe, A., Mac Quene, K., van Niekerk, I., Gandhi, V.,Harigaya, T. and Abrahams, C. (2004). *The Social and Economic Impact of South Africa's Social Security System,* Economic Policy Research Unit.

Smit, E.v.d.M. (2010). (ed.) Special Edition on the South African National Income Dynamics Study, *Studies in Economics and Econometrics,* 34(3).

South African Social Security Agency (SASSA). (2009). *Statistical report on social grants, Report No. 14 (28 February 2009),* Pretoria: Department of Monitoring and Evaluation, SASSA.

Southern Africa Labour and Development Research Unit (Saldru). (2009). *National Income Dynamics Study Wave 1 July 2009:* User Document, Cape Town: Saldru, University of Cape Town.

Sutherland, H. (2001). *EUROMOD: an integrated European Benefit-Tax model - Final Report,* EUROMOD Working Paper EM9/01.

Sutherland, H., Figari, F., Lelkes, O., Levy, H., Lietz, C., Mantovani, D. and Paulus, A. (2008). *Improving the capacity and usability of EUROMOD - Final report,* EUROMOD Working Paper EM4/08.

Wilkinson, K. (2009a). *Adapting EUROMOD for use in a developing country – the case of South Africa and SAMOD,* EUROMOD Working Paper No. EM5/09, University of Essex.

Wilkinson, K. (2009b). 'The impact of taxes and transfers on child poverty in South Africa', Paper prepared for the International Microsimulation Association Conference, Ottawa, June 8-10, 2009.

Wilkinson, K., Wright, G. and Noble, M. (2009). *SAMOD User's Handbook: Installing and running the South African Microsimulation Model, Version 1.1,* Pretoria: Department of Social Development, Republic of South Africa.

Woolard, I., Leibbrandt, M. and de Villiers, L. (2010). 'The South African National Income Dynamics Study: Design and Methodological Issues', *Studies in Economics and Econometrics,* 34(3): 7–24.

Woolard, I., Simkins, C., Oosthuizen, M. and Woolard, C. (2005). *Final report Tax incidence analysis for the fiscal incidence study being conducted for the national Treasury.*

Wright, G., Noble, M., Dinbabo, M., Ntshongwana, P., Wilkinson, K. and Le Roux, P. (2011). *Using the national income dynamics study as the base micro-dataset for a tax and transfer South African microsimulation model*, Report produced for the Office of the Presidency, South Africa.

Zaidi, A. Harding, A. and Williamson, P. (Eds.) (2009). *New frontiers in microsimulation modelling*, Vienna: Ashgate.

Front page footnotes:

1 Senior Research Fellow, Oxford Institute of Social Policy, University of Oxford and Deputy Director of the Centre for the Analysis of South African Social Policy. Senior Research Associate, Rhodes University.
2 Professor of Social Policy, Oxford Institute of Social Policy, University of Oxford and Director of the Centre for the Analysis of South African Social Policy. Honorary Fellow of the Human Sciences Research Council. Honorary Professor, Rhodes University. Visiting Professor, University of KwaZulu-Natal.
3 Research Officer, Institute for Social Development, University of the Western Cape, South Africa.

University of Stellenbosch
Institute for Social Development
School of Government Building
Bellville

Soreaso (Social Research Solutions)
Somerset West
Cape Town
svdb@sun.ac.za