SOUTHMOD: Modelling Tax-benefit Systems in Developing Countries

André Decoster

KU Leuven, Leuven, Belgium.
andre.decoster@kuleuven.be

Jukka Pirttilä

University of Helsinki, VATT Institute for Economic Research, UNU-WIDER, Helsinki, Finland.
jukka.pirttila@helsinki.fi

Holly Sutherland

University of Essex, Colchester, UK.
hollys@essex.ac.uk

Gemma Wright

Southern African Social Policy Research Insights, Hove, UK.
Archie Mafeje Research Institute, University of South Africa, Pretoria, South Africa.
gemma.wright@saspri.org

INTRODUCTION

Social protection has attracted increasing interest in developing countries in recent decades and policies have been initiated in all developing regions. Cash transfer policies are already in place in many African countries, but their coverage is still limited, although it is expanding.[1] Many of the cash transfer programmes have been started as pilots with donor funding in certain regions or among certain target groups. Instead of isolated social protection projects, there is now an urgent need to move towards scaling up such policies to generate nation-wide social protection *systems*.

When countries build up their social protection systems, they need reliable information and tools on how the systems should be designed. Second, they need to explore new ways of financing social protection, since donors increasingly demand that the systems are financed using domestic means rather than donor support. This requires fostering tax capacity and developing the tax systems simultaneously with advances in social protection.

Tax-benefit microsimulation models are tailor-made tools to examine, from an ex ante perspective, the system-wide impacts of tax and benefit policies. They are especially well suited for the analysis of interactions between different policies, which is of key importance when determining the overall impact of taxes and transfers on poverty, inequality, and fiscal net revenues. Such models are routinely used by policy makers and researchers in developed countries. However, with few exceptions, developing countries have not had access to such models. Having an opportunity to test the likely impacts of policy reforms would be especially timely in the current circumstances where developing countries build up their policies from a very basic —or even non-existent— level of social protection.

This is the backdrop against which UNU-WIDER, a Helsinki-based research institute devoted to the study of economic development in poor countries, initiated the SOUTHMOD project in 2015.[2] The articles in this issue represent selected research outputs from the project.

The purpose of the project was to build new models for seven developing countries (Ecuador, Ethiopia, Ghana, Mozambique, Tanzania, Vietnam, and Zambia) and update and further develop two existing models (NAMOD for Namibia and SAMOD for South Africa). The models were prepared by local teams from the participating countries, with support from microsimulation experts from the EUROMOD team at the University of Essex, Southern African Social Policy Research Insights (SASPRI), and KU Leuven.

As the name may reveal, the models are built on the EUROMOD platform. This is because EUROMOD offers an easy-to-use, yet fully programmable, user interface for microsimulation. The second reason is that EUROMOD has embraced the goal of producing comparable simulation models across European countries. For an international research organization such as UNU-WIDER, comparability is most useful and can stimulate South-to-South learning among different participating countries.

The first versions of the models are now ready, and we have organized model launches and training courses in the countries in question to inform policymakers and researchers about the models and to teach potential users. It is planned to keep the models up-to-date by coding new policy rules as reforms take place and producing additional underpinning input datasets when new waves of data become available. And of course, the models are being developed further: one way to develop them is via research that is based on the models. The current issue reports on some of the research findings that are based on the first batch of SOUTHMOD models.

The rest of this introductory article is organized as follows. The next Section discusses matters pertinent to using EUROMOD as a base when building new microsimulation models. Next, special circumstances that need to be taken into account when modelling tax and benefit policies in developing countries are discussed. Finally, a summary of the papers in this issue is provided.

USING EUROMOD

EUROMOD is a multi-country tax-benefit microsimulation model for the 28 member states of the European Union (EU) based on representative household micro-data (Sutherland & Figari, 2013). It is designed for comparable simulations and analyses across countries as well as for policy swaps and EU level analysis (Sutherland, 2014). It has evolved and expanded since it was first initiated in 1996 and much of this development has been related to the need to accommodate the policies and data and to address the relevant research and policy questions in a consistent way; not only over time but also across a growing number of EU countries. The modelling platform as a whole (described further below) is designed to be both flexible and transparent as well as generic (not country-specific) but at the same time to provide a clear framework for the incorporation of new countries. It is also made freely available. Since it is unique in many of these respects EUROMOD was the obvious choice of modelling platform for the SOUTHMOD project. In addition, the recent successful incorporation of Croatia (which joined the EU in 2013) as well as the development of non-EU models using the EUROMOD platform (e.g. for Russia and Serbia)

demonstrated that EUROMOD was fit for purpose and, most relevantly, models for South Africa and Namibia had already been developed by adapting the EU model.[3]

While there are advantages in using an existing platform as a starting point rather than starting from scratch, there may also be disadvantages. Here we summarise the trade-offs as we experienced them. First of all, as in many contexts, it is undoubtedly easier to make an effective start on a new project if there is an existing structure and guidelines, even if these are questioned and amended later. Furthermore, each country starting from the same structure and guidelines meant that the end products would be more similar, even if not fully comparable, than if they had started independently from blank sheets. Having the same starting point also, in principle at least, helped to encourage dialogue and collective learning.

More specifically, the "existing platform" provided by EUROMOD consisted of (i) the EUROMOD software including a special purpose tax-benefit modelling language and variable-naming convention as well as a well-tested design, (ii) examples of policy rule implementations for EU countries (and South Africa and Namibia) as well as a simplified "SimpleLand" system to adapt, (iii) detailed guidelines as set out in the EUROMOD Modelling Conventions (EMC) document[4], (iv) training and support in the use of the software as well as regular maintenance and development of it, (v) Stata "do-file" templates for the derivation of EUROMOD input data from available micro-data and other documentation templates such as for the Country Report and Data Requirements Document. Some of these components are intended and designed to be generic and applicable, in principle at least, in any country. Others are specific to a particular context (such as the "do-files" which are designed for use with the European Union Statistics on Living Conditions (EU-SILC) micro-data but which are nevertheless useful in identifying an organised structure for data derivation).

There were of course some challenges associated with adopting the EUROMOD software, structure and guidelines. They were associated with four factors. First, since the aim was to produce a useable model that would be useful in a specific country context rather than comparatively in a relatively short period of time, some of the guidelines intended to ensure comparability were too burdensome or not a priority.[5]

Secondly, the developing country context rendered irrelevant some of the guidelines and conventions developed for the EU context; and these issues could be different across countries and especially across world regions. One example is the non-relevance of contributory

unemployment benefit, and hence of guidance about how to simulate it, in some of the African countries.

Thirdly, and relatedly, the data availability and also data requirements of the systems to be simulated were not the same as in the EU and often very different; they were also different across countries, and especially across regions. There are many examples and there were particular issues with income variables and considerable differences in the income reference time periods used. The inclusion of detailed information on household expenditures in African micro-data permitted the simulation of indirect taxes in a straightforward way that is not possible in the EU. To facilitate this while avoiding the imposition of a harmonised approach, EUROMOD was adapted to allow the input of detailed country-specific expenditure variables.

Finally, in many cases the trade-offs between starting from the EUROMOD framework or some other alternative depended on the existing experience and capacity of the national team. For example, needing to learn how to use a new software and a new modelling language was more of a disadvantage for teams that were already familiar and skilled with alternatives (such as Stata). The extent to which less-relevant aspects of the guidelines were a time-consuming distraction was something that reduced in importance as the project progressed and the team as a whole learned what was and was not a priority in the context of the countries being studied. The EMC was adapted to improve its applicability in the developing country context and particularly in view of the different data challenges on the one hand, and on the other, the need for simplicity given that the primary aim was to create a collection of pioneering usable and useful national models as an important first step.

TAILOR-MADE MODELS FOR DEVELOPING COUNTRIES

Members of the SOUTHMOD collaboration encountered significant data-related challenges during the models' development phase. As a consequence, the research papers included in this edition (discussed further below) each have a back-drop of at least two years of intensive gathering of the relevant data and policy information from government departments and stakeholders, and extensive collaborative data preparation. While survey datasets, tax-benefit rules, and annual reports containing key administrative data are available at a click of a button in some countries, in most of the SOUTHMOD countries the process of acquiring these information sources took many months, often requiring letters to be written, in-person visits to be arranged in different cities with different stakeholders, and multiple follow-ups for additional information.

Survey data

The first data-related challenge arose when preparing the underpinning datasets for each of the country models. Each model requires a single rectangular individual-level input file which has to be prepared using the standardized variable naming conventions that are required by the EUROMOD software. For some countries, nationally representative survey data is not available online and has to be requested for a particular purpose: this was the most straightforward hurdle to overcome.

More challengingly, several of the SOUTHMOD countries' bureaux of statistics do not provide pre-prepared rectangular datasets and so these had to be hand-crafted by the research teams, often in the absence of any metadata except the survey questionnaires. Extensive data-cleaning was required for some of the countries' datasets, particularly to ensure that the household relationship data was cleaned, and that there were no excessive outliers in the income and expenditure data that would distort the simulations. The assumptions, imputations, and programmatic and manual cleaning steps that were undertaken are summarized in each of the "Country Reports", and in practice this phase was by far the most labour-intensive.

Preparation of the survey data on income was often a "voyage of discovery", as in many of the countries the income data had not been used previously for research purposes, especially as most of the countries use consumption-based measures of poverty and inequality which do not require usage of the income data even though it is collected. There was very little information available to enable external validation of the income data. The SOUTHMOD programme has an ongoing work stream to explore more sophisticated data cleaning processes for the income data, including a range of multiple imputation methods for missing and implausible data, and the use of administrative data on personal income taxes to improve the income data. There is great scope for cross-country sharing of experiences and techniques in this regard.

Information about tax-benefit policy rules and administrative data for external validation

The second data-related challenge that was faced by the SOUTHMOD team was the difficulty encountered when trying to obtain information about current tax and benefit systems, both in terms of collating the policy rules for incorporation into the country models, but also obtaining administrative data about the current policies for the purposes of external validation. The Tanzanian paper in this edition provides an example of the intricate (and fluid) rules that exist in Tanzania for social security benefits. Regarding the acquisition of external validation data, in some

instances it was simply not possible though this was never due to access being denied. It is hoped that over time such information will become more readily accessible and that research activities such as the SOUTHMOD programme will help to provide an impetus alongside larger initiatives such as the United Nations World Data Forum[6] and the African Union's African Charter on Statistics which also promotes transparency and access to data.[7]

The SOUTHMOD team comprises an inter-disciplinary and multi-lingual group of individuals with fluency in languages including: Bulgarian, English, Finnish, German, isiXhosa, Portuguese, Spanish, and Swahili. This was much more of a strength than an impediment, as it required there to be meticulous clarity of communication within each country model team, between the country model team and the country stakeholders, between the country model teams, and when converting policy rules into instructions for each country model using the EUROMOD software. As highlighted above, the EUROMOD project itself promotes harmonized terminology and naming conventions, and so EUROMOD was well suited to accommodate these realities. That having been said, all of the models were produced and annotated in English. Some of the training events were undertaken in other languages, for example partly in Swahili in Tanzania and almost exclusively in Portuguese in Mozambique. Training materials for civil servants are increasingly being produced in languages other than English in order to facilitate access to the models. Much more could be done to promote access to the models for non-English speakers within and outside the team, as well as to promote awareness of the models within the respective countries amongst stakeholders in the vernacular.

Importance of indirect taxes and informal sector

The Government Revenue dataset of UNU-WIDER reveals that low income countries rely for 67% of their tax revenue on indirect taxes, and for 19% on individual income taxes. For high income countries this is respectively 55% for indirect taxes and 35% for personal income taxes.[8] This illustrates two things. First, in developing countries, indirect taxes, in the form of import and export duties, excise duties and sales taxes (on intermediate and final transactions, or retail sales taxes and VAT), typically play a larger role as revenue generating instrument, than the personal income tax system. Second, during the process of development there often is a gradual move away from more distortive general sales taxes to less harmful indirect taxes such as VAT and to taxes which do take into account individual abilities to pay (such as personal income taxes).[9] Monitoring the budgetary and distributional effects of this important change in the revenue structure of developing countries, made it compelling to integrate from the very beginning of the project

indirect tax analysis into the models. The SOUTHMOD models for Ethiopia, Tanzania, Mozambique, Zambia, Ghana and Ecuador each have a VAT policy, plus an Excise Duty policy for tobacco, alcohol, and fuel. The South Africa and Namibia country models have VAT policies, but no excise.

As already noted above, in the brief description of the EUROMOD backbone of the project, this integration of indirect taxes in the policy scope of the models, was in fact less cumbersome a process than it has been for many tax benefit developers in rich countries. Contrary to the available income surveys or administrative datasets underlying the EU-countries in EUROMOD, which in most cases do not contain detailed expenditures, most of the countries of the SOUTHMOD project did use surveys that provide detailed information on expenditures, and even quantities, for a detailed break-down of total expenditures. This allowed the incorporation of retail sales tax or VAT and the most important excise duties in a rather straightforward way. Since no behavioural model was linked with the personal income tax module, we have chosen to postpone also the integration of behavioural reactions such as income or relative price effects into the consumer allocation model. Changes in consumer prices, triggered by a change in either VAT or an excise duty, are affecting after-tax disposable income by calculating the change in indirect taxes to be paid at unchanged quantities. But it is well-known that even this impact effect is an acceptable first order approximation of the true welfare effect of price changes (see Capéau, Decoster, & Phillips, 2014).

Also, other assumptions follow the mainstream of most arithmetic microsimulation models but might be somewhat more restrictive in the context of developing countries. The underreporting of some expenditure categories when compared to national account aggregates is in itself a well-known issue in richer countries but is mostly fixed by ad hoc adjustments for revenue predictions. Of course, this underreporting changes nature and becomes more relevant from a welfare point of view when it reflects purchases in the informal sector. The quick fix, consisting of uprating expenditures and/or revenues to calibrate the simulated revenue in the baseline with some external indirect tax total, obviously becomes less attractive for two reasons. First, by definition no taxes are paid in the informal sector. And second, the margin of behaviour of shifting from formal to informal sector will matter.[10] This caveat holds a fortiori for the distributional effects when the underreporting and/or reliance on purchases in the informal sector is more prevalent in expenditure categories which are relatively important for the poor.

The same consideration holds, mutatis mutandis, for the pass-through of general sales taxes (i.e. not only falling on the final purchase, but also paid on transactions of intermediate products) and

the treatment of non-reclaimable VAT. The latter might be particularly relevant in developing countries, since it is probable that many traders have a turnover below the applicable registration threshold for VAT. In that case, the increased production costs due to un-reclaimable VAT charged on inputs to production might be reflected in the final price paid by the consumer.[11] There exist methods to estimate these effects of the VAT on the prices of exempt goods, and incorporate these implicit tax rates into microsimulation models (see Ahmad & Stern, 1984, for a seminal; and Du, 2015 for a recent illustration using input-output tables). However, in line with standard practice of most microsimulation models, at this stage we have only included indirect taxes paid directly by consumers and do not take account of the indirect effects of non-reclaimable VAT or of excises paid at earlier stages of production. We leave these —to wit, important— extensions for a further phase of the project.

SUMMARY OF THE ARTICLES IN THIS ISSUE

The first article in the issue, following this introductory article, by Jouste and Rattenhuber presents a comparison of simulating old-age pensions in a number of countries. While some countries for which SOUTHMOD models exist have a functioning old-age pension programme, others have pension coverage only for workers in the public sector, in addition to some voluntary contributed schemes. The paper simulates different common pension reforms for Ecuador, Ghana, South Africa, and Tanzania. The policies modelled differ in terms of their generosity and the threshold age (above 60 or 70 years). For countries that already run an existing pension system, the new systems are assumed to either replace or complement the systems already in place. The results suggest that the impact of the pension systems on reducing poverty among the target group is potentially considerable, for instance reducing poverty headcount by 33% in the case of Tanzania. But the costs can also be large, requiring a major increase in the tax-to-GDP ratio or the GDP. The paper demonstrates the similarities and differences that arise in different countries when modelling these kinds of policies.

The second paper by Jara and Varela examines the degree of redistribution in Ecuador using two different approaches: direct use of reported taxes and benefit in survey data and use of simulated taxes and benefits. This is possible since taxes paid and transfers received are actually reported in the data underpinning the ECUAMOD microsimulation model. The results suggest that the simulated values better predict the actual total sums of taxes and transfers, based on external information. Similarly, distributive impacts are larger when estimated on the basis of simulated

values. This suggests that in further, counterfactual, analysis, it makes sense to use the simulated values as a benchmark.

In the third paper, Wright, Leyaro, Kisanga, and Byaruhanga critically evaluate the policy transparency in providing social benefits (via the Productive Social Safety Net (PSSN)) in Tanzania. It is shown how complex the eligibility criteria is. Issues can arise from community-based targeting, and it is difficult to find clear documentation of the system. Needless to say, these issues also complicate reliable simulation of the system. The authors suggest different ways of making the system more transparent and effective and also to provide simulation results for a number of benefits that are arguably easier to implement.

The final paper by Pirttilä, Osei, and Rattenhuber examines expansion of social protection in Ghana. They consider both unfunded and fully funded scenarios, where the current LEAP (Livelihood Empowerment Against Poverty) transfer system is made more generous, especially to older people. In the budget-neutral reform, the benefit expansion is financed by an increase in social security contributions. In addition to static simulation results, the authors also take into account potential behavioural changes due to the increase in the tax burden. For this end, they incorporate estimation results regarding the change in the share of formal sector work when the tax burden on formal-sector work is increased. The results suggest that the desirable effects of the reform remain valid, even when taking into account the behavioural reactions.

REFERENCES

Ahmad, A., & Stern, N. (1984). Theory of reform and Indian indirect taxes. *Journal of Public Economics*, *25*(3), 259-298.

Barrientos, A. (2013). *Social Assistance in Developing Countries*. Cambridge: Cambridge University Press.

Capéau, B., Decoster, A., & Phillips, D. (2014). Consumption and Indirect Tax Models. In C. O'Donoghue (Ed.), *Handbook of Microsimulation Modelling* (Contributions to Economic Analysis, Volume 293, pp. 223-273). Emerald Group Publishing Limited.

Du, L. (2015). The effects of China' s VAT enlargement reform on the income redistribution of urban households. *China Finance and Economic Review*, *3*(1), 1-15.

International Labour Organization (2017). *World Social Protection Report 2017-2019: Universal social protection to achieve the Sustainable Development Goals*. Geneva: International Labour Office.

Kleven, H., Kreiner, C., & Saez, E. (2016). Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. *Economica 83*(330), 219–246.

Sutherland, H. (2014). Multi-Country Microsimulation. In C. O'Donoghue (Ed.) *Handbook of Microsimulation Modelling* (Contributions to Economic Analysis, Volume 293, pp. 77-106). Emerald Group Publishing Limited.

Sutherland, H., & Figari, F. (2013). EUROMOD: the European Union tax-benefit microsimulation model. *International Journal of Microsimulation*, *6*(1), 4-26.

Waseem, M. (2018). Informality and income shifting. Evidence from recent Pakistani tax reform. *Journal of Public Economics* 157, 41-77.

Wright, G., Noble, M., Barnes, H., McLennan, D. & Mpike, M. (2016a). *SAMOD, a South African tax-benefit microsimulation model: recent developments*. UNU-WIDER Working Paper 2016/115. Helsinki: UNU-WIDER.

Wright, G., Noble, M., Barnes, H., McLennan, D. & Mpike, M. (2016b). *Updating NAMOD, a Namibian tax-benefit microsimulation model*. UNU-WIDER Working Paper 2016/143. Helsinki: UNU-WIDER.

NOTES

[1] See International Labour Organization (2017) and Barrientos (2013) for overviews.

[2] For more information about the project, please see https://www.wider.unu.edu/project/southmod-simulating-tax-and-benefit-policies-development.

[3] See Wright, Noble, Barnes, McLennan, and Mpike (2016a; 2016b) for details about SAMOD and NAMOD respectively.

[4] see https://www.euromod.ac.uk/publications/euromod-modelling-conventions

[5] Indeed, it is quite likely that guidelines to ensure comparability designed for an African or Latin American multi-country model would not all be the same as those designed for comparability across European countries.

[6] see for example https://undataforum.org/

[7] See https://au.int/sites/default/files/documents/32838-doc-charte_en.pdf

[8] We used data from the Government revenue dataset available at https://www.wider.unu.edu/project/government-revenue-dataset, release of July 2017 and consulted on Nov 29th 2017. For tax revenue, we used total tax revenue excluding social security contributions and natural resource revenue.

[9] Kleven, Kreiner, and Saez (2016) describe this process by making a distinction between "traditional taxes" and "modern taxes", the latter consisting of individual and corporate income taxes, payroll taxes and social security contributions, and value-added taxes.

[10] Using quasi-experimental variation created by a tax reform in Pakistan in 2009, Waseem (2018) shows that this kind of behavioural responses of firms in an environment of low enforcement and large informality are indeed large.

[11] Also, the VAT paid on goods purchased as inputs to the production of goods that are exempt from the VAT cannot be reclaimed, and obviously also excise might be passed through into the consumer prices.