PEP toolbox
Analytical tools

A top-down behaviour (TBD) microsimulation toolkit for distributive analysis: A manual

Luca Tiberti
Martin Cicowiez
John Cockburn

August 2017

partnership for economic policy
pep

A top-down with behaviour (TDB) microsimulation toolkit for distributive analysis: A manual

Macro/CGE models are often combined with microsimulation (MS) models to perform distributive impact analysis for fiscal or structural policies, or external shocks. This manual describes a user-friendly Stata-based toolkit to perform microsimulations combined with CGE models in a top-down fashion. It includes a module which estimates parametrically the income generation where the movements across sectors as well as the changes in wages and self-employment revenues are captured, by skills of workers. Then it estimates households' specific price deflators based on an individual utility function to capture the effects of changes of the purchasing of power by individuals over time. The changes estimated by a CGE model (or from other sources) in the employment (by skill and sector), in the wage payroll (by skill), in the revenues from self-employment activities (by skill) as well as in the commodities prices are fed into the MS model in a consistent way. Once this tool has estimated the new vector of real consumption or revenue, it performs a series of distributive analysis, such as standard poverty and inequality indices, their decomposition by income factor, robustness analysis and growth incidence curves, and compare the baseline with the simulation results. This will help us to run standard poverty and distributive analyses, and to see whether a given shock or policy have had some impact on household welfare and who are the most affected households. Based on such information, social protection policies can be accurately designed in order to minimise the, e.g., negative effects of a given shock in a cost-effective manner. An illustrative analysis is run on data from Uganda.

Authors

Luca Tiberti
Partnership for Economic Policy (PEP)
Université Laval
luca.tiberti@ecn.ulaval.ca

Martín Cicowiez
CEDLAS-UNLP
PEP
mcicowiez@gmail.com

John Cockburn
PEP
john.cockburn@pep-net.org

Acknowledgements

The authors would to thank Luc Savard and Jean-Yves Duclos for their insightful help, comments and discussions. Financial support from Oxford Policy Management (OPM) is greatly acknowledged (Contract Name: 5198 - Development of an Integrated Macro Economic Model for the Government of Uganda; PO Number: 5198 / PO10155). We also acknowledge the financial support from the Partnership for Economic Policy (PEP), with funding from the Department for International Development (DFID) of the United Kingdom (or UK Aid), and the Government of Canada through the International Development Research Center (IDRC).

Table of Contents

1. Overview

1.1 Introduction

This toolkit explains how to build and use a microsimulation (MS) model to analyse the distributive impacts of economic shocks and policies. Information on the latter - which are introduced into the MS model as exogenous changes in key variables - can come from various sources (e.g. macro projections, Computable General Equilibrium (CGE) model, hypothesised shocks, etc.). While this toolkit can be used in all these cases, it is particularly targeted to researchers who wish to construct a MS model where information on the "shock" variables come from a CGE model. The full toolkit (including this manual, the Stata codes and the example dataset) can be downloaded from https://www.pep-net.org/microsimulation-distributive-analysis.

Very generally, a MS model is a model of the behaviour of individual agents (individuals, households, firms). As such, it can be used to simulate the effects of economic policies or other shocks on those individual agents. In this toolkit, we focus on the impacts on individuals and households; not firms. Typically, an individual/household MS model:

- Requires micro-data from a household survey including:
  - socio-economic characteristics of individuals
  - labor-market status and labor incomes
  - household spending
- Simulates changes in household budget constraints.
- Operates in partial equilibrium.
- Does not simulate changes in macro and meso variables such as prices, wages, or employment rates.

MS models generally include two types of modeling:

- Parametric. This generally consists of a system of equations that determine:
  - occupational choice,
  - returns to labor and human capital,
  - consumer prices,
  - other household (individual) income components
- Non-parametric. Generally, this involves seeking individuals with similar characteristics to simulate certain changes (for example, a change in labor income for an individual that moves from unemployment to employment). For example, in order to assign a labor income for an individual that moves from unemployment to employment, labor income of individuals with similar observable characteristics but originally employed will be used. In turn, occupational shifts may be proxied by a random selection procedure within a

segmented labor market structure. For example, among the unskilled individuals that move from employment to employment a random selection is made

The final objective of the microsimulation module presented in this note is to simulate changes in per capita household income/welfare under various (counterfactual) scenarios. This information is then used to conduct standard poverty and distributive analyses of these changes and make policy recommendations. For example, in the case of negative shocks, social protection policies can be accurately designed to target the most affected households/individuals in order to minimise the negative effects in a cost-effective manner.

As mentioned above, we will focus on the particular case where the MS model draws information on the shock variables from a CGE model. This type of combined CGE-MS models has been used widely to evaluate the distributive impacts of macroeconomic shocks and policies such as:

- Public expenditures: Changes in size or composition;
- Tax/subsidy policies;
- Structural reforms: trade liberalization, privatization, labor market reforms, etc.;
- Global price shocks.

The idea to link macro (including CGE) and MS models for this purpose emerged in the mid 1980s (inter alia Meagher and Agrawal (1986) examinded the impacts of changes in the existing tax mix on the distribution of income in Australia). Combining a CGE model with a MS model makes it possible to capture and link the impacts of macro shocks on the economy as whole (CGE model) and on households/individuals (MS model). Together, they are able to capture both efficiency and distributive impacts.

CGE models allow the modeller to focus on:

- Winners and losers at the sectoral level
- Impact on macroeconomic variables
- General equilibrium price effects

However, they are not an adequate tool to perform distributional analysis given the lack of individual/household results and the representative agent assumption (i.e. the household in a CGE model is an aggregate household and not an average household).

On the other hand, MS models focus on household and/or individual behaviour. They are the key methodology to capture distributional effects of a policy change due to heterogeneity at the household or individual level. However, they are not able to capture the economy-wide effects of macro shocks (international trade, tax policy, public expenditures, etc.). Furthermore, they lack general equilibrium effects. If the macro shock or policy is substantial, there will likely be consequences for income distribution, sectoral prices, factor returns, employment, labor supply, GDP, etc., which are not captured by a MS model alone. For example, the reform of a tax-benefit system is likely to affect labour supply which, in turn, may affect wages and prices. Even in cases where the shock impacts at the micro level (e.g., a cash transfer programme), if it is on a large scale it is likely to have significant macro repercussions (e.g., changes in labour force participation, changes in household consumption) that may, in turn, feed back to the micro impacts.

CGE and microsimulations models are highly compatible. Both:

- Are used for comparative static analysis
- Are used to analyse policy reforms
- Allow similar hypotheses to be formulated in the counterfactual framework

The main elements for developing a MS model are:

- To have access to a micro-dataset that provides information on the socio-demographic and economic characteristics of individuals or households;
- To establish the characteristics of the policies to be simulated;
- To specify a behavioral response model for households and individuals (i.e., labor supply, savings, household composition, household preferences) (if a behavioural microeconomic approach is chosen).

When we want to use a combined CGE-microsimulation approach to evaluate the distributive effects of macroeconomic policies, particular care must be paid to the following issues, which will be referred to throughout this toolkit:

- Need to have a counterfactual scenario (i.e. the case of the policy not being implemented),
- Distinction between ex-post or ex-ante evaluation (whether the policy is already in place or not, and whether the outcomes are observed or estimated). The methodological tools are not the same for the two types of evaluation,
- Partial or General equilibrium effects (whether effects on prices and on economic equilibria are taken into account). A given macro shock normally produces both partial and general equilibrium effects.
- Behavioral or accounting approaches (whether agents are allowed or not to react – change their behavior – to a policy change).

1.2 CGE-microsimulation approaches

There are several major categories of CGE-micro simulation models found in the literature:

1. The representative household group approach: RHG (just briefly discussed here)
2. The fully integrated approach: FI (not discussed here)
3. The top down-micro-accounting approach: TD-MA (just briefly discussed here)
4. The Top down with behaviour approach: TD-WB (more details below and in next chapter, and application on real data)
5. The bottom up approach: BU (not discussed here)
6. And the Top down/bottom up or iterative approach: TD/BU or IA (not discussed here)

We focus primarily on the TD-WB approach, as it is the most widespread. However, before turning to this approach, let us just mention a few points concerning the RHG and TD-MA approaches, which are also quite common and allow us to better situate the advantages of the TD-WB approach. For the other approaches, refer to Cockburn, Savard and Tiberti 2014 for a recent review.

1. Standard Representative Household Groups (RHG) approach:

- Consists in decomposing several groups of households in a CGE model (e.g. from 2-3 to a hundred households)
- This is not a pure CGE-MS framework, as it still uses representative – not actual – households, but allows some distributive analysis as compared to CGE models with only one representative household
- Represents the oldest approach to using CGE models for distributional analysis
- Decomposes household in RH groups based on socio-economic characteristics

The standard RHG approach is not able to take fully into account the distributive effects of a reform for a number of reasons:

- The distribution of welfare within each RHG is assumed to be policy-neutral. Thus, only the distribution of welfare between RHG can be analysed.
- However, empirical evidence shows that changes in "within RHG" inequality can be equally or more important than "between RHG" inequality in explaining variations in "overall" inequality. More generally, within-group heterogeneity in socio-economic characteristics can imply that within-group impacts of economic shocks can be larger compared than between-group impacts (Huppi and Ravallion 1991, Mookherjee and Shorrocks 1982).
- Any distributional analysis is limited to the pre-defined RHGs
- The demographic weight of households in each RHG is assumed to be constant

This said, the RHG approach is the easiest to apply for efficiency and distributive impact analysis of macro shocks.

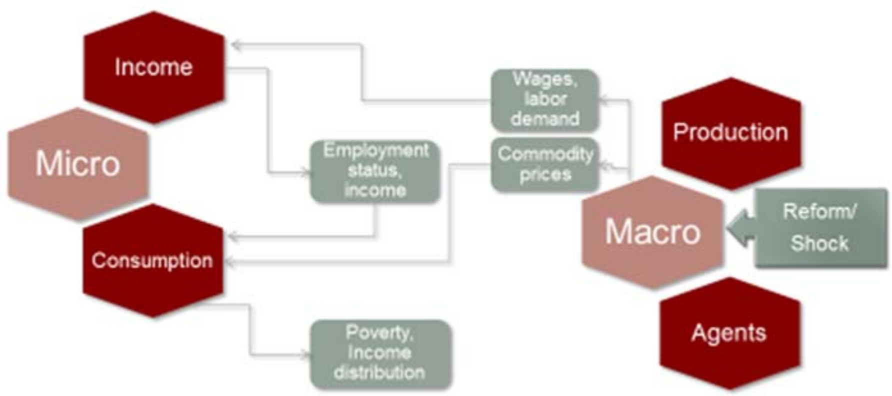
2. Top-down approaches (macro-micro linkages) - or layered approach, see Davis (2009)

Such approaches basically consist in taking results from a CGE model – such as changes in prices, factor returns, employments levels, etc – and feeding them into a MS model to analyse the distributional impacts.

We distinguish two main top-down approaches:

- Micro-accounting (TD-MA): no behaviour
- With Behaviour (TD-WB) (Labour supply – see Robilliard, Bourguignon and Robinson, 2008; Consumption and savings), which generally includes a micro-econometric estimation.

A top-down "micro-simulation" approach is illustrated here below:



For these kinds of approaches, some preliminary work is required to ensure compatibility between the CGE and MS models in terms of data and modelling hypotheses. Ideally, we should:

- calibrate the national account data (used in the CGE model) and household data (use in the MS model) in order to reach ``empirical consistency". For example, Robilliard and Robinson (2003) propose a method to reconcile these data by modifying the sampling weights;
- harmonize the categories (labor, commodities, products, sectors, etc.) and functional forms (e.g., consumer demand) for the two layers (CGE and MS).

However:

- when a sequential approach is followed, "full consistency is not required between the macro and the micro sides of the model. Indeed, all of the analysis using this model may be performed in terms of deviations from benchmarks that may not fit perfectly together" (Robilliard, Bourguignon and Robinson, 2008, pg. 106);

It should also be noted that the household or individual behaviour in the MS model often lacks the theoretical consistency of the CGE model. For example, in MS models, for occupational choices relative prices rarely enter the utility function.

We will focus on the TD-WB approach, but should point out a few salient characteristics of the TD-MA approach.

3. The top down-micro-accounting approach: TD-MA

Under this approach there is no explicit change in behaviour. Specifically, it follows the standard envelope theorem where the change in household consumption is simply the sum of the net consumption at the base year (as observed in the survey) multiplied by the relative change in consumer prices. In the same way, variations in labor income is equal to base-year labor income multiplied by the relative change in wage rates. Variations in other sources of income are similarily calculated from the changes in the relevant factor returns. If changes are small enough and there are no market failures, this approach is fully consistent with a full behavioral response model. According to the standard envelope theorem, formally we have that the change in household welfare is given by (see Chen and Ravallion, 2004):

$$\Delta y_i = \sum_j (q_{i,j}^0 - c_{i,j}^0) p_j^0 \left(\frac{\Delta p_j}{p_j}\right) + L_i^0 w_i^0 \left(\frac{\Delta w}{w}\right) + \Delta R_i$$

$$y_i^1 = y_i^0 + \Delta y_i$$

Where:

- Δy_i and y_i^1 stand, respectively, for the change in the equivalent welfare and the equivalent welfare after simulation 1 for household *i*. In welfare analysis, especially in developing countries, the welfare indicator is usually identified by the consumption variable. By "equivalent welfare" we mean the welfare indicator after any time and price differences are taken into account. More details are provided below.

- $y_{i,j}^0, q_{i,j}^0, c_{i,j}^0, p_j^0, L_i^0, w_i^0$ are observed in the household survey at the base year and identify, respectively, welfare, household production, consumption and prices of product/commodity *j*, worked hours and wage rates/income of household *i*. In a typical household survey, we observe the values of household production ($q_{i,j}^0 p_j^0$), consumption ($c_{i,j}^0 p_j^0$) and labor income ($L_i^0 w_i^0$).

- The CGE model, on the other hand, provides us with the following "shock" variables: $\frac{\Delta p_j}{p_j}, \frac{\Delta w}{w}, \Delta R$

Using the equation above, poverty and distributive analysis can then be conducted by calculating vectors Δy_i and y_i^1

TD-MA is simple to implement but has the following drawbacks:

- Loss of information due to assumed fixed behaviour
- Generally overestimates impacts as a consequence

- Limited to use for small price changes and short term impact
- Lack of micro-feedback effect to the CGE model
- Heterogeneity of households is limited to the structure of income and/or consumption
- Lack of constraint for behavioural consistency between models

4. The Top down with behaviour approach: TD-WB

This approach will be presented in detail in the following sections. Here we summarize its key characteristics. The key characteristic is that the MS model includes econometric modeling of household behaviour, allowing for full household/individual heterogeneity, e.g.:

- Wage rates
- Occupational choice (e.g. multinomial logit)
- Self-employment income
- Consumer behaviour

MS model is then used to simulate the effects on household income of introducing variations in a subset of variables (not only wages, prices, other income – as under the accounting approach – but also employment status) from the CGE model that influence the behaviors outlined above.

Consistency between the MS and CGE models requires that:

- Each have the same three labor markets (wage, informal urban self-employment and agricultural self-employment) and categories of consumption goods.
- Changes in average wage earnings with respect to the benchmark in the micro-simulations must be equal to changes in wage rates obtained in the CGE model for each segment of the wage labor market;
- Changes in self-employment income in the micro-simulations must be equal to changes in informal urban sector and agricultural sector income in the CGE model.
- Changes in the number of wage workers and self-employed by labor market segment in the micro-simulation models must be equal to those same changes in the CGE model;
- Changes in the consumption price vector must be consistent with the CGE model.
- Ideally, the household data underlying the MS model should be used to estimate several key parameters used in the CGE model (e.g. consumption demand).

Consistency can be ensured by:

- Introducing elasticities of labour supply that can also be used to model aggregate labor supply in the CGE model (see Bourguignon and Savard (2008))
- Recalibrating (e.g., changing the intercepts of the behavioural equations of MS model as in Robilliard, Bourguignon and Robinson (2008); ranking the individuals according to their estimated probabilities associated to each labour behaviour to change their status consistently with the macro employment results as in Cockburn, Robichaud and Tiberti (2017) – more description of the latter is provided below)
- Imposing consistency equations (as in Colombo (2010))

To verify, we check that changes in LAVs match with the changes in average values of the corresponding variables in the MS model.

To summarize, this approach:

- Introduces rich household/individual behaviour and heterogeneity
    - Captures extensive margin and intensive margin changes in labour supply
    - Captures discrete choices by individuals or households
- Is more complex, requiring econometric modelling and rich data

Its main limits are:

- The absence of a micro-feedback effect to the CGE model. The magnitude of this problem is linked to the size of the aggregation error from the micro households up to the aggregate households in the CGE model.

Introducing dynamics

These models normally keep key household characteristics (demographics, school participation, etc.) constant. Dynamic simulations, in the form of repeated simulations of the household income-generation model for period-specific changes in the shock variables, would require adding a population-ageing model to the micro-simulation model. For instance, this may require a micro-simulation model that ages the population over time using estimated or calibrated functions for fertility, mortality, migration, marriage, household formation, schooling, etc. (see e.g. Grimm (2005)).

It is useful to consider that, when combined with a dynamic CGE, the MS model should take into account the evolution of the population. This can be done by "sampling re-weighting" techniques or by changing the attributes of the individuals. We can in fact distinguish between:

- "**static ageing**" (sampling re-weighting): sampling weights are modified so that the simulated population matches the CGE aggregates. This approach is usually used in the short-to-medium run as no significant alteration in the population structure is expected. A well-known technique is the algorithm developed by Deville and Särndal (1992);
- "**dynamic ageing**": real life events (e.g. birth, death, etc) are modeled and, based on that, individual and household characteristics are updated at each period. These approaches are often combined with calibration mechanisms to align micro aggregates with CGE forecasts.

2. Top-down with behaviour CGE-microsimulation: Description of the methodology

In what follows, we present the econometric models that are used in the top-down with behaviour CGE-microsimulation approach. We illustrate with an application using data from

Uganda, which were collected in 2010-2011 (UNHS). Microeconomic behavioural parameters are estimated using cross-section data from 2010/2011 and kept constant over time.

The following behaviours are introduced in the MS model:

- employment
- consumption

2.1. Income Generation Model: Employment and income

Under a behavioural microsimulation approach, individuals and households adjust different types of behaviour. In the CGE-microsimulation literature, most of the contributions focus on the modelling of labor markets and incomes from working activities.

This module ultimately aims at estimating the variations in household welfare due to changes in employment status and income from all working activities. Consistently with the SAM and the CGE model hypotheses, we need to:

1. Identify:
    - the occupational choices available to household members
    - the categories of workers based on their skills
    - the economic sectors;
2. Establish the degree of rigidities in the labour market (e.g. degree of mobility of workers across the different occupational choices);
3. Identify the best suited function for the occupational choice and income, and estimate them;
4. For each simulation period, predict the individual employment status as well as the corresponding income, and estimate the total household income.

Occupational status[1]

The occupational choices available to household members are: (1) wage worker; (2) non-agricultural self-employed; (3) farmer; (4) not working (which also includes unemployed and apprentices). In order to keep the modelling manageable and due to the incompleteness of the information on second occupations, we only look at the first occupation. We thus did not create mixed categories (e.g. wage worker and self-employer), and the definition of each choice is exclusive. We distinguish between wage workers and the self-employed as we assume that the Ugandan labor market is imperfectly competitive. If this is not the case, we should find that the (actual or shadow) wage rates for these types of occupation are not different.

[1] This section takes some inspiration from Robilliard et al. (2008).

We have two categories of workers (based on their individual education skills): (1) skilled (if they completed at least primary education – P7); (2) unskilled (if not).

Concerning the labor mobility hypotheses, consistently with the CGE model, in each period each worker can potentially move between being a wage worker, farmer or non-agricultural self-employed, where the decision is determined by individual and household characteristics.

Estimation

Given that we generally do not have good data on the number of hours worked, the labor supply by a household member is defined as a discrete choice among the four alternatives presented above. As is commonly done with multiple discrete choices, the individual labor supply is estimated with a standard multinomial logit model,[2] where each choice is then modeled within a discrete utility-maximizing framework. The utility associated with each possible category is a function of a set of individual and household characteristics. The model we use to estimate the individual labor supply (E_i) is a reduced-form model in the sense that the remuneration rate (in each of the three working alternatives) does not enter the estimation of the labor supply. The general function to be estimated is:

$$\ln \frac{P(E_i = m)}{P(E_i = 4)} = \alpha_m + \sum_{j=1}^{J} \beta_{mj} X_{ij} + u_{ij} = Z_{mi}$$

where Z_{mi} represents the actual individual utility function associated to each occupational choice *m*[3]. Among the explanatory variables *X*, we include the individual's age and his/her highest educational degree. The individual residual terms u_{ij} are drawn following the methodology described in Bourguignon, Fournier and Gourgand (2001), where it is shown that the distribution of residuals of a multinomial logit model is independent, random and has a double exponential form. In particular:

1. The individual residual term in a MNL model is assumed to be an independently and identically distributed extreme value term. The distribution is also called Gumbel (or type I extreme value).
2. The cumulative distribution of ε_{im} is $F(\varepsilon_{im}) = e^{-e^{-\varepsilon_{im}}}$

[2] We are aware that with such approach we do not control for the potential selection in the estimation of income (from wage and profits) associated to each occupational alternative (this issue may be solved by using, for example, the method developed in Bourguignon, Fournier and Gourgand, 2007). However, for profits we only have information at the household level and we prefer not to decompose it into individual values (for more details, see below in the text).

[3] The fourth category – inactive – has a utility arbitrarily set to zero.

3. The difference between the two extreme value variables $\varepsilon^*_{ims} = \varepsilon_{im} - \varepsilon_{is}$ follows a logistic distribution: $F\left(\varepsilon^*_{ims}\right) = \dfrac{e^{\varepsilon^*_{ms}}}{1+e^{\varepsilon^*_{ms}}}$

Note that, empirically, the difference between extreme value and independent normal errors is indistinguishable (Train, 2003). Individual *i* chooses alternative *m* if

$$\alpha_m + \beta_m X_{im} + \varepsilon_{im} > Max\left(\alpha_s + \beta_{sj} X_{ij} + \varepsilon_{is}\right) \quad \text{with } m = 1,..,S, \forall m \neq s$$

Intuitively, we need then to find values for ε such that the observed employment statuses are validated. In other terms, the residuals are drawn randomly from the Gumbel distribution until the set of these values is in accordance with the observed occupational choices (and the inequality condition set above).

As is often done, we may suppose that there is some sequence in the choice of employment status across the members of the same household. That is, the household head is the first to decide and the other household members follow.

- Practically, we should[4]:
  - first estimate the multinomial logit model for the household head, whose choice is explained by both general household characteristics and those of all household members.
  - Then estimate the occupational choice of the other household members, conditional on the decision taken by the household head.

We can then estimate the individual probabilities of being in one of the four categories in each simulation period t as follow:

a) For other than reference categories: $P\left(E_i = m\right) = \dfrac{\exp\left(Z_{mi}\right)}{1+\sum_{m=1}^{m=3}\exp\left(Z_{mi}\right)}$

b) For the reference category: $P\left(E_i = 4\right) = \dfrac{1}{1+\sum_{m=1}^{m=3}\exp\left(Z_{mi}\right)}$

After the probabilities associated with each occupational status are estimated, we can proceed with assigning the new individual employment status. Differently from Robilliard et al. (2008), this is done by a "job queuing" approach. According to the CGE results concerning the employment status, the absolute numbers moving in or out the three working categories is

[4] Note that, in the application below, we do not follow any within-household sequence.

estimated. The individuals changing from one alternative to another are selected according to their probability of being in the concerned choice.

As an example, assume that the CGE predicts an increase from 20 to 23% of the ratio associated to "skilled wage workers". We rank all working-age skilled individuals according to their probabilities of being in that sector. We then select among individuals not previously working in the wage sector (i.e. farmers, non-agricultural self-employed or not employed) in descending order of probability of being in that sector until the predicted share of the workers in this category satisfies the CGE result.

As mentioned above, the CGE results need to be plugged into the microsimulation analysis in a consistent manner: i.e., the share of each working category in the CGE model and the MS model must be the same in each simulated period. Note that, if changes in individual or household observed characteristics are simulated, only the explained component of the model is recomputed. Ideally, the validity of this way of proceeding should be tested with panel data.

In the illustration below, we assume perfect mobility across sectors (wage workers can move to non-farm, farm self-employment or not employed, and vice versa), and perfect rigidity across type of workers (unskilled workers cannot move into the skilled pool, and vice versa).

Income from working activities

Once occupational choice is determined, we can proceed to calculating household labor income. As discussed earlier, we distinguish between wages (usually collected at the individual level) and income from self-employment activities (generally collected at the household level).

Estimation

Individual wage

The wage equation follows the Mincer model and is estimated by a two-step procedure (Heckman model). We first estimate the "selection equation" (i.e. being in the wage sector or not, as identified by the binary variable s_i) through a probit model as:

$$s_i^* = \gamma z_i + u_i$$

$$s_i = \begin{cases} 1 & if\ s_i^* > 0 \\ 0 & if\ s_i^* \leq 0 \end{cases}$$

with z_i representing suitable selection variables (e.g., marital status, number of children).

We then estimate the "wage" equation as: $w_i = \begin{cases} \beta x_i + \varepsilon_i & if\ s_i^* > 0 \\ . & if\ s_i^* \leq 0 \end{cases}$

As is well known, this procedure is suggested because, if u_i and ε_i are correlated, then β is biased and inconsistent ($E(\varepsilon_i) \neq 0$). In simpler terms, in such a case the unobservable variable in the selection equation (being a wage worker) also affects the wages.

- As u_i and ε_i can be correlated, the conditional expectation of wages is:

$$E\left(w \middle| x_i, s^* > 0\right) = \beta x_i + \sigma_{sw}\lambda\left(\gamma z_i\right)$$

where σ_{sw} is cov(u_i , ε_i) and λ is the inverse Mills' ratio (or non-selection hazard – how much the conditional wage is shifted up due to the selection).

In addition, for simulation purposes, ε_i, which can be identified as individual unobserved fixed effects (or, heterogeneity of individual earnings), is easily estimated only for individuals employed in the wage sector at the base (observed) year. For individuals who did not report information on wages (i.e. non-wage workers), this residual term is estimated by drawing randomly from a normal distribution with the relevant (e.g. skilled or unskilled) observed variance. Note that, after the simulations, only the deterministic component of the model is recomputed (by using the parameters' regression estimated in the base year).

Finally, we can integrate the variations in the wage rates as predicted by the CGE. Note that changes in average wages (i.e., total payroll) with respect to the baseline figures in the micro-simulation must be equal to changes in wage rates obtained in the CGE model for each type of worker category.

Household profit (farmers and non-agricultural self-employed)

Household profit (farmers and non-agricultural self-employed) is defined as the difference between the value of output and the value of inputs observed in a given period. In the profit function, we should also consider the imputed value of self-production activities. This component, usually associated with agricultural products, should be valued at producer market prices.

The profit equation is estimated through an instrumental variable approach as some explanatory variables – e.g. number of family workers – may be endogenous and thus correlated with the error term.

The basic model to be estimated for household profit is estimated through a Cobb-Douglas function (so, all observed variables enter in ln):

$$\ln \Pi_{h,t}^{j} = \alpha + \beta_1 \ln X_{h,t} + \beta_2 \ln N_{h,t}^{j,sk} + \beta_3 \ln N_{h,t}^{j,unsk} + u_h$$

with $N_{h,t}^{j,sk/unsk} = \sum_{i=1}^{N} I(E_{i,t}^{sk/unsk} = j)$; and $j = 2$ if non-agricultural sector and $j = 3$ if agricultural sector.

Henceforth, $\ln N_{h,t}^{sk}$ and $\ln N_{h,t}^{unsk}$ are, respectively, the total number of skilled and unskilled family workers in period t and $X_{h,t}$ identifies some usual household characteristics. In addition, we should consider that:

- The number of family members involved in self-employment is likely to be endogenous (due to, e.g., omitted variables) so that it is desirable to instrument these variables (a Hausman test to see whether these variables are endogenous or not should be run subsequently)
- As usually done in the literature, the exogenous variables of family composition can be suitable instruments. As well-known, the exogenous regressor(s) (instrument(s)) must not be correlated with the error term in the outcome regression (*cov*(Z,u)=0), but they should be correlated with the endogenous explanatory variable(s)

Similarly to the case of wages, we need to recover the error term u_h both for those households who reported a positive profit value during the base year and for those with missing values. Again, note that, after the simulations, only the deterministic component of the model is recomputed (by using the regression parameters estimated at the base year).

The estimated household profit (including the residual terms) is finally divided by the total number of household members (skilled and unskilled) working in the farm or in the non-agricultural family enterprise. The per-worker profit is thus obtained (and defined as π_h).

The changes in profits from farming and non-agricultural self-employment activities as simulated by the CGE are then fed into the MS model. The changes in net income from self-employment activities in the micro-simulation must be equal to changes in income per worker in non-agricultural and farming sectors resulting from the CGE model.

We can finally estimate the total household income at time *t* ($Y_{h,t}$) as:

$$Y_{h,t} = \sum_{i=1}^{N} w_t^{sk} I(E_{i,t}^{sk} = 1) + \sum_{i=1}^{N} w_t^{unsk} I(E_{i,t}^{unsk} = 1) + \sum_{j=f,na} \pi_{h,t}^{j} N_{h,t}^{j,sk/unsk} + y_{h,t}^{ex}$$

where the first two components on the RHS give the total income from wages summed at the household level for skilled and unskilled wage workers and the third component represents total household profit. Note that the last term ($y_{h,t}^{ex}$) identifies eventual exogenous private (e.g. remittances) or public (e.g. government program of cash transfer) transfers to the household. The changes predicted by the CGE for $w_t^{sk/unsk}$, $E_t^{sk/unsk}$ (for $E_t^{sk/unsk} = 1,2,3$) and , π_t^j are integrated into the microsimulation as described above.

We can then easily calculate the relative change of income between the base year and t as:

$$r_{h,t} = \left(Y_{h,t} / Y_{h,0}\right) - 1$$

which is then applied to total household consumption to estimate the change in total income. The value we obtain is then added to our welfare variable (in most contexts, this is the initial expenditure variable) (see below). In such a way, we implicitly assume that there is no change in the marginal savings rate. Another reason for following this strategy is that the resulting change in income is compatible with the expenditure variable. This would not be necessarily the case if we add the absolute change in income estimated.

Some final remarks:

- Ideally, all components influencing household income generation (employment choice, wage, income from self-employment) should be estimated simultaneously.
- Indeed, the unobservable terms in the different equations are likely to be correlated. This can be particularly relevant in a framework of utility maximisation, for which the random terms in the occupational choices is likely to be dependent on the error terms in the wage and income functions (Bourguignon and Ferreira (2005)).
- However, the estimation of these simultaneous equations might be particularly complex and the estimations not robust (Bourguignon and Ferreira, 2005).
- As discussed, the estimations of employment and incomes include a random term. This implies that the household income for the counterfactual scenarios is random. As Bourguignon and Ferreira (2005) say, "[T]his is not too great a problem if the microsimulation relies on a sufficiently large number of observations." If this is not the case, we should replicate the exercise a sufficiently large number of times in order to get a distribution of counterfactual distributions.

2.2. Household consumption and prices

This module serves to estimate the predicted change in real household consumption. We are then able to estimate the variation in poverty and inequality for each simulation period. The change in real consumption comes from the variation in household incomes (as derived in the previous modules) and consumer prices.

Per capita consumption in constant prices is our variable of interest to estimate changes in poverty and inequality across the different simulation scenarios.

The first step is to define the categories of commodities available from the household survey. These categories are determined by mapping the categories in the underlying micro and CGE data and then aggregating by nature of commodity.

Definition of household consumption

Consumption values in the household survey have to be expressed in the same time period as the poverty line (usually monthly or annual). Total consumption is obtained by aggregating purchases, self-consumption and gift values over all household consumption categories to calculate total household consumption. In this definition, we exclude the purchase of durables and the value of their services, except for the real or imputed value of rent for housing, which is included. Finally, user fees and charges, as well as local service taxes, are included in the definition of household consumption.

Equivalent income

To take into account the heterogeneity of the effect of price changes across households, it is important to calculate a household-specific price index. To do so, we relied on King's (1983) approach to define the concept of *equivalent income*. For a given budget constraint the equivalent income of household *h* living in cluster *c* in year *t* is defined as the income level that, at the reference price system $\boldsymbol{p^r}$, yields the same utility level as attained with consumption $\boldsymbol{x^t_{h,c}}$ and prices $\boldsymbol{p^t_c}$

$$v\left(p_c^t, x_{h,c}^t\right) = v\left(p^r, e_{h,c}^t\right)$$

where *v(.)* is the indirect utility function, and $e^t_{h,c}$ is the *equivalent income* function specific to the household *h*. By inverting the indirect utility function, we can obtain the equivalent income as an expenditure function:

$$e_{h,c}^t = e\left(p^r, p_c^t, x_{h,c}^t\right)$$

Let assume that $\boldsymbol{p^r} = \boldsymbol{p^0}$, where $\boldsymbol{p^0}$ represents the vector of prices faced by households living in the reference cluster or region (in our case, Kampala) at the base year (i.e. before policy reforms). By following King (1983), the change in household welfare (equivalent gain) after the reform (t=1) is given by:

$$EG_{\mathrm{h},c} = e(\mathbf{p}^0, \mathbf{p}_c^1, x_{\mathrm{h},c}^1) - e(\mathbf{p}^0, \mathbf{p}_c^0, x_{\mathrm{h},c}^0)$$

If $EG_{\mathrm{h},c}$ <0, then a transfer to household *h* equal to $EG_{\mathrm{h},c}$ would compensate the impacts of the reform.

Equivalent income based on a Stone-Geary utility function

Ideally, using richer cross section survey data, a complete demand system should be estimated to derive the equivalent income functions. After all, the demand system used in the microeconomic module should be the same as that used in the CGE model. Here we assume K-commodity Cobb-Douglas preferences. Its derived equivalent income is:

$$e_{h,c}^{t}\left(\mathbf{p}^{0},\mathbf{p}_{c}^{t},x_{h,c}^{t}\right)=\frac{x_{t,h}}{\prod_{k=1}^{K}\left(\frac{p_{k}^{t}}{p_{k}^{0}}\right)^{w_{k,h}}}$$

where the denominator identifies the household specific price deflator, with $w_{k,h}$ representing the household budget share associated to category *k*.

Poverty and distributive analysis[5]

For the purpose of poverty and distributive analysis, we used the standard poverty gap indices and the Gini inequality index. We also run some stochastic dominance analysis to assess the robustness of our results throughout the simulation period.

The FGT poverty indices are defined as:

$$P_{\alpha}^{t}(z)=\frac{1}{\mathbf{N}}\sum_{h=1}^{H}\rho_{h,c}n_{h,c}\left(\frac{z-e_{h,c}^{t}(\mathbf{p}^{0},\mathbf{p}_{c}^{t},y_{h,c}^{t})}{z}\right)_{+}^{\alpha}$$

where z is the monthly poverty line in the reference cluster (Kampala) in the base year (32106.24 UGS); $f_+ = \max(0, f)$; **N** is the number of households in the survey (and corresponds to the sum of the sampling weights); $N_{h,c}$ is the size of household *h*; $\rho_{h,c,t}$ is the sampling weight of *h* in time *t* (remember that the sampling weight evolves according to the population growth, as explained above); and α is a parameter that captures "aversion to poverty".

The variation in the FGT index between the base year (t=0) and the simulated period t can be written as:

$$\Delta P_{\alpha}^{t}(z)=\frac{1}{\mathbf{N}}\left(\sum_{h=1}^{H}\rho_{h,c}n_{h,c}\left(\frac{z-e_{h,c}^{t}(\mathbf{p}^{0},\mathbf{p}_{c}^{t},y_{h,c}^{t})}{z}\right)_{+}^{\alpha}-\sum_{h=1}^{H}\rho_{h,c}n_{h,c}\left(\frac{z-e_{h,c}^{0}(\mathbf{p}^{0},\mathbf{p}_{c}^{0},y_{h,c}^{0})}{z}\right)_{+}^{\alpha}\right)$$

According to the stochastic dominance literature, if

$$\Delta P^{s}(z)=P^{t=1,\ldots,T}(z;s)-P^{t=0}(z;s)<0$$

for any choice of poverty index within a class of ethical order $s=\alpha+1$ and for any poverty line $[0,z^{+}]$, then we can unambiguously affirm that poverty in scenario t is decreasing (the distribution under this scenario is said to stochastically dominate the distribution in the base year). If ΔP^{s} changes sign, then the poverty results should be interpreted with some caution.

[5] Poverty and distributive estimates, as well as significance tests and dominance curves, are carried out using the DASP program (Araar and Duclos, 2007).

To assess the distributive effects of the different policy simulations, we use the well-known Gini Index. Starting from the class of single-parameter Gini indices:

$$I\left(\rho, p^0, p^t, y^t\right) = \int_0^1 \left(s - L\left(s\right)\right) \kappa\left(s;\rho\right) dp$$

for *ρ=2*, we obtain the standard Gini index, with ρ being an ethical parameter; *L(s)* being the cumulative percentage of total income held by the cumulative proportion *s* of the population (ranked according to increasing consumption values *y*); *κ(s, ρ)* being the percentile-dependent weights to aggregate the distances s-L(s).

The variation in the Gini index between the base year *(t*=0) and the simulation *t* can be written as:

$$\Delta I_{Gini}^t = I_{Gini}^t\left(p^0, p^t, y^t\right) - I_{Gini}^0\left(p^0, p_c^{\ 0}, y^0\right)$$

Finally, we analyse the poverty decomposition by income sources (notably, wages, profits and consumer prices) and the growth incidence curves. Decomposition by income source is performed using the Shapley/Shorrocks approach (for more details, see Shorrocks (2013) and Azevedo et al. (2013)). The growth incidence curve is a useful tool to have an overall picture of impacts over the whole distribution. It is estimated as the difference of the logarithm of the welfare variable under the simulation and the baseline scenarios at each percentile of the distribution (see Ravallion and Chen, 2003).

2.3. Final methodological remarks

As discussed, the estimations of employment and incomes include some random terms. This implies that the household income for the counterfactual scenarios is random. We proceeded by replicating each simulation a sufficiently large number of times (100 times). Then, we took the median values of these 100 replicates for each relevant estimator (namely, per adult equivalent income, and poverty and inequality indices) in each simulated year, and estimated their corresponding confidence intervals.

3. Feeding the microeconomic model and how to use the Stata codes

In this section, we describe how the MS model described in section 2 can be implemented with a given set of shocks (e.g., those obtained from a CGE model) in order to obtain poverty and inequality results. In particular, we explain how this approach is implemented in Stata.

In our MS model the labor market and the consumer prices are considered to be the main transmission channels of the impact of the simulated policy or exogenous shocks scenarios on poverty and income inequality. For example, individuals may change position in the labor market (and hence also affect household income) due to external shocks, trade reforms, or other policy changes. In turn, workers may shift from one sector to another, change occupation or lose their jobs. From a practical point of view, the methodological issue is to find a procedure that can account for such labor-market shifts and identify which individuals are most likely to shift position in order to be able to simulate a new, counterfactual income distribution. In addition, our MS model also considers changes in commodity prices through a household-specific consumer price index.

In the literature, various microsimulation methods have been proposed to deal with changes in the labor market. In our case, we implement a method that involves the estimation of a microeconomic, partial-equilibrium household income generation model through a system of equations that determine occupational choice, returns to labor and human capital, consumer prices and other household (individual) income components (see, for instance, Bourguignon, Fournier and Gurgand, 2001).

In a nutshell, for a number of variables of the labor and commodity markets (i.e., changes in skilled and unskilled employment and wages, and changes commodity prices), changes for each year of the scenarios are calculated relative to a particular year, or base year of the microsimulations. As seen above, some of the results generated by the CGE model are fed into our MS model. More precisely, each scenario definition should provide changes with respect to the base year (2010) for the following variables:

1. employment:
    - rural skilled employed
    - urban skilled employed
    - rural unskilled employed
    - urban unskilled employed
2. wage payroll:
    - skilled workers
    - unskilled workers
3. income from self-employment activities
    - skilled workers
    - unskilled workers

4. (categories of) commodity prices

Unfortunately, the available information in the CGE and micro data do not always perfectly match. So, some adjustment that makes the two sources as comparable as possible is sometimes necessary.

In the specific case of this toolkit, for (1) and (2), the CGE data distinguishes 16 employment categories: 4 skill levels (unskilled, semi-skilled, skilled, high-skilled) X 2 locations (rural, urban) X 2 genders (male, female). To be consistent with the micro data, CGE results have been averaged into four categories: 2 skill levels (skilled=semi-skilled, skilled or high-skilled; unskilled=unskilled) X 2 locations (rural, urban).

In terms of sectors of activity, since the CGE data do not distinguish between "wage", "non-agricultural self-employed" and "agricultural" workers (as in the micro data), we made the hypothesis that "wage" and "non-agricultural self-employed" workers are approximated by "urban", while "agricultural" workers are associated to "rural" workers.

Concerning (3), we used the change in capital returns for agricultural and non-agricultural activities to simulate changes in self-employment income for agricultural and non-agricultural sectors in the MS model.

For (4), the CGE categories of commodities have been regrouped and mapped with micro categories: (1) coffee; (2) tea; (3) flowers; (4) maize; (5) rice; (6) other cereals; (7) cassava; (8) roots; (9) pulses; (10) matoke; (11) other agricultural items; (12) fish; (13) mining; (14) other food; (15) beverages & tobacco; (16) textile; (17) petrochemical; (18) other manufacturing; (19) electricity; (20) water & sanitation; (21) construction; (22) public administration; (23) private primary education; (24) private secondary education; (25) public primary education; (26) public secondary education; (27) private health; (28) public health; (29) other infrastructure; (30) other services.

The changes simulated by the CGE model and used to feed the MS model are in the Excel files:

- "***uga2010-base-defn***" (results for the baseline)
- "***uga2010- example-defn***" (results for the simulation scenario).

3.1. Implementation in Stata

The steps that need to be taken to implement the above-described approach are all explained in this section. A microsimulation toolkit has been prepared and it has been packaged in the file called "TDBmicrosim.zip". This file needs to be decompressed in one folder in order to use the content of its three folders: Do-Files, Output, and Raw-Data.

3.2. Installation of the Microsimulation Toolkit

In order to install the microsimulation toolkit, the user has to follow the four steps below:

1. **Create a folder within which the contents of the file *TDBmicrosim.zip* will be unzipped**. For example, *C:\TDBmicrosim-app*, where app would be replaced by the name of your application; for example, a three-letter acronym that stands for the country for which the method will be applied. The base year of the application could also be added as part of the application name; so, for an application for a country like Uganda whose base year is 2010, app could be represented by, for example, uga2010.
2. **Unzip the file *TDBmicrosim.zip* in the above newly-created folder; keep the folder structure unchanged**! The file contains three major folders: **Do-Files, Output, and Raw-Data**. The first folder, in turn, contains all the "do files" by means of which the microsimulations are implemented using a Stata code whose structure is outlined in Figure 1. In addition, the subfolder Scen-Files is used to store the definition of the base and non-base scenarios to be simulated (see below). The Output folder will be used to store the microsimulation results together with intermediate files produced when running the MS model. Finally, the Raw-Data folder contains the original micro dataset.
3. **Subsequently, the folder Do-Files should be configured**. This is done by adjusting the file master.do, in order to edit the "base" path (following the instructions in point 2, you should have "cd "C:\TDBmicrosim\Do-Files"". This is important for Stata to be able to call up the various do, input, and output files. Save the file. Besides, specify (a) the number of years to be considered in each scenario (local macro yrnum), and (b) the number of times the microsimulations will be run repeatedly for each year (local macro called iternum). Last but not least, save the file master.do.
4. **Likewise, the contents of the folder Raw-Data should be provided**. Specifically, save in this folder the processed household survey (Uganda2009.dta).

Figure 1: organization of Stata code MS model

3.3. Preparation of the Micro Dataset

As part of step 4, the user has to prepare a micro dataset with the information at both individual and household levels that is required to conduct microsimulations. Generally speaking, this micro dataset should contain the following information: population weights, household identifier, individual identifier, individual labor income, family consumption spending by commodity, labor status (employed, unemployed and inactive), employment sector, occupational category, educational level (in order to establish skill levels), and other individual characteristics such as age, sex, and so on. This information is usually available in a household or labor force surveys. More specifically, the required variables needed to prepare the dataset are grouped for households, all individuals, and employed individuals, as outlined in panels (a) to (c) of Table 1.

Table 1a: individual level variables in dataset for the microsimulations

Name	Description
worker_sim0	employment status
	1 = wage
	2 = self-empl NA
	3 = farmer
	4 = not-working
wgt09_sim0	sampling weight
hhid	Unique HH Identifier Across Panel Waves
pid	Unique Person Identifier Across Panel Waves
age_sim0	age
skilled_sim0	1=skilled;0=otherwise
worker_sim0	employment status (=1 wage;2=self-empl NA;3=farmer;4=not-working)
gender	gender
	1 = male
	2 = female
hh_head	0 = not hh head
	1 = hh head
region	Region
education	education macro degrees
married	0 = not married
	1 = married
n_children	number of children
wgt09_sim0	sampling weight
	note: wgt09_sim1, wgt09_sim2, ... are used to consider pop growth
urban	Rural/urban
education	education macro degrees
salaried_sim0	0=not wage-worker
	1=wage worker
married	0=not married
	1=married
pov_line	Poverty line in 2005/06 constant prices
	(can vary across regions)
cpexp30	Monthly household expenditures in constant prices after adjusting for regional p
nrrexp30	Monthly nominal household expenditures in market prices & after regional price a
exp_1 - exp_30	expenditure on various commodities
equiv	Adult equiv
	(if equal to 1, welfare is computed on a per capita basis)
pid	Unique Person Identifier Across Panel Waves
ea	Enumeration Area
stratum	UNPS Strata of Representativeness

Table 1b: Household level non-agricultural self-employment variables in dataset for the microsimulations

Name	Description
gender_hh	gender of the hh head
hheduc	education degrees for hh head (primary,secondary,tertiary)
ln_hhwork_2_sim0	number of family workers in non-agr SE (ln)
ln_hhwork_2_1_sim0	number of skilled family workers in non-agr SE (ln)
ln_hhwork_2_0_sim0	number of unskilled family workers in non-agr SE (ln)
hsize	household size (#)
n_worker2_sim0	number of family workers in non-agr SE
rev_2_tot_0	household profit from non-agr SE

Table 1c: Household level agricultural self-employment variables in dataset for the microsimulations

Name	Description
gender_hh	gender of hh
hheduc	education level of hh
ln_hhwork_3_sim0	number of family workers in farming (ln)
ln_hhwork_3_1_sim0	number of skilled family workers in farming (ln)
ln_hhwork_3_0_sim0	number of unskilled family workers in farming (ln)
hsize	household size
n_worker3_sim0	number of family workers in farming
rev_3_tot_0	household profit from farming

For income variables, data should be recorded on an annual basis. Otherwise, the user will have to adjust the contents of the *sim-hhd-inc.doi* file. Specifically, simulated income variables are divided by 12 at the same time that a monthly poverty line is used to compute poverty indicators. In the Uganda2010 example, the poverty line is expressed at 2005/06 prices, while income variables are expressed at 2009/10 prices. Thus, we need to deflate income variables using the ratio between two price indexes.

3.4. Running microsimulations and generating results

Once the above steps have been duly implemented, the following should allow the user to run microsimulations. The first step is to edit the five tables stored in the Excel files in the ***Do-Files\Scen-Files*** folder, which contain the shocks associated with the base and non-base scenarios to be simulated – for example, see the files *uga2010-base-defn* and *uga2010-example-defn*, respectively. This edited (or new) files are used to impose a counterfactual scenario in order to simulate impacts on poverty and inequality of given simulated changes. The scenarios are defined by changing tables representing deviations from the labor-market structure as observed from a given household survey, for each of the parameters of the labor-market structure. Besides, changes in commodity prices can also be considered. The dimension of these tables is predefined: for example, they are designed to work with two skill levels. This means, for example, that these matrices should be modified in the case that the user wants to expand the number of skill levels.

The second step to implement our MS model is to run the file (double click on) *master.do*, which is located in the ***Do-Files*** folder. Notice that this will take some time – depending on the speed of the processor. As a result of this routine, the microsimulation results are stored in the file *welfare_sim*, which will be located in the **Output** folder. In addition, some descriptive statistics from the household survey can be found in the file *PEP-Microsimulations.log*, which is also located in the **Output** folder. Finally, a summary report is saved to the *report.xlsx Excel* file -- also located in the **Output** folder – which contains the following indicators for each simulation year:

Sheet distindic, reporting the following indicators

- time = simulation period
- FGT0_base = poverty rate for baseline
- FGT0_example = porverty rate for non-base simulation
- FGT1_base = poverty gap rate for baseline
- FGT1_example = porverty gap rate for non-base simulation
- Gini_base = Gini coefficient for baseline
- Gini_example = Gini coefficient for non-base simulation

Sheet fgtdecomp, reporting the following indicators

- time = simulation period
- welfare = zero by definition (as it serves as benchmark value in the baseline)
- wage = contribution of wage changes to overall change in welfare
- nonWageNonAg = contribution of non-wage non-agricultural income to overall change in welfare
- nonWageAg = contribution of non-wage agricultural income to overall change in welfare
- CPI = contribution to (household-specific) price changes to overall change in welfare
- Total

Sheet difgtindic, reporting the following indicators

- time = simulation period
- FGT0_base = poverty rate for baseline
- FGT0_example = porverty rate for non-base simulation
- diff = absolute difference between FGT0_base and FGT0_example
- standardDev = standard deviation of diff
- lowBound = lower bound for diff
- upperBound = upper bound for diff
- FGT1_base = poverty gap rate for baseline
- FGT1_example = porverty gap rate for non-base simulation
- diff = absolute difference between FGT1_base and FGT1_example
- standardDev = standard deviation of diff
- lowBound = lower bound for diff
- upperBound = upper bound for diff

Sheet diginiindic, reporting the following indicators

- time = simulation period
- gini_base = Gini for baseline
- standardDev_base = standard deviation of Gini for baseline
- gini_example = Gini for non-base simulation
- standardDev_example = standard deviation of gini for non-base simulation
- diff = absolute difference between FGT0_base and FGT0_example
- standardDev_diff = standard deviation of the difference
- lowBound_diff = lower bound for diff
- upperBound_diff = upper bound for diff

Also, the following graphs, for each simulation year, are stored in the **Output\Graphs** folder :

- file cfgt-t.gph = FGT curves (along an axis of poverty lines) for simulation period t
- file cfgts2d-t.gph = differences between FGT poverty curves with confidence interval for simulation period t
- file cnpe-t.gph = growth-incidence curve for simulation period t, estimated using parametric regression curves

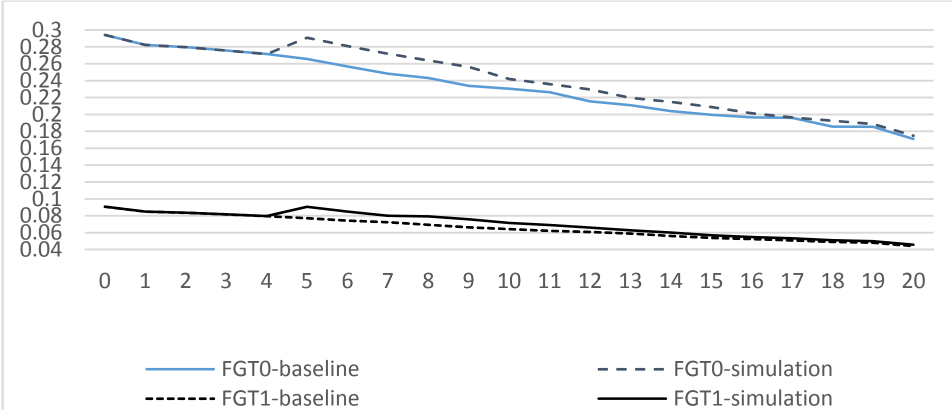
4. Interpreting the microeconomic results of a given shock

In the Ugandan illustration, the CGE simulation period is from 2011 to 2030. Also, for the baseline simulation, forecast data (regarding GDP growth and population growth) from the IMF are used. Regarding the policy scenario, the first phase of the simulation (2015-2017) concerns large investments in the oil sector when there is no production from this sector. This is followed up with a construction phase from 2018 – 2030. In this second phase, the oil sector generates production, and profit and labour income generated in the oil sector goes to Ugandan households. It is also in this phase that profits and labour payments are generated. In this simulation we assume that that foreigners fund the investment in the oil sector (as a "gift" from the rest of the world to Uganda).

As said earlier, the main goal of the microsimulation component is to estimate the impacts on households and individuals. In particular, we want to identify the most affected and advantaged population groups in the distribution following the proposed investment in the oil sector.

As shown in Figure 1, the poverty rate and poverty gap decrease along the whole simulation period. However, their trajectory differ between the baseline and the simulation scenarios. As for the simulation setup, the two scenarios do not differ until 2014. As said earlier, the simulated policy intervention starts in 2015. The short to medium term impact on household welfare and poverty is negative, as detected by the upward jump in 2015 (or year 5 of the simulation) for both the incidence of poverty and the poverty gap. Specifically, the proposed investment would cause a rise in the incidence of poverty by 2.5 percentage points and by around 1.3 percentage points in the poverty gap. In other words, such a policy would push back the poverty indicators to their 2010 value. In the long term, the baseline and reference scenarios tend to converge.

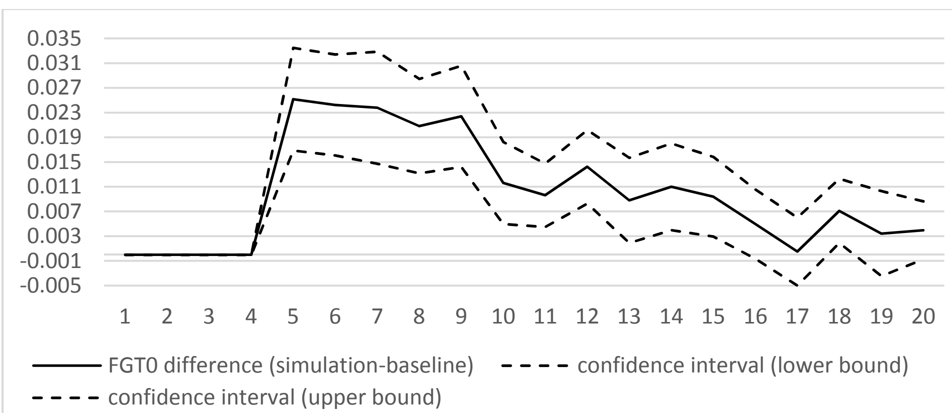
Figure 1: Evolution of the FGT0 and FGT1 between 2010 and 2030



Source: authors' estimations
Note: Graph based on the Excel file "report"

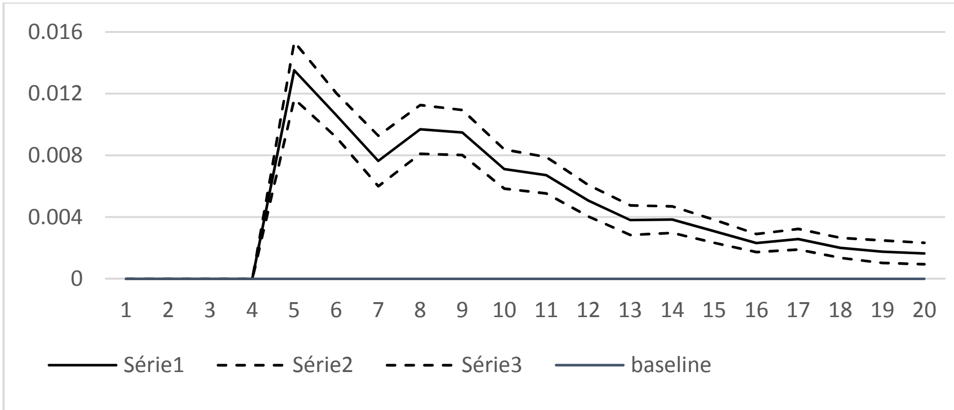
According to Figure 2a, the poverty headcounts in the two scenarios are not statistically different starting from 2027 (or year 17), with the only exception of 2028 (where the poverty rate is significantly higher in the simulation scenario). In contrast, as shown in Figure 2b, the poverty gap under the simulation scenario is always statistically higher than in the baseline along the whole period, although the difference approaches zero near the end. The policy intervention also has a negative impact on inequality (Figure 2c). The effects are higher when the oil sector starts to produce and to generate income (i.e. from 2018) and slowly decrease over time, but being always statistically significantly higher than under the baseline.

Figure 2a: Difference between the FGT0 under the simulation and the baseline



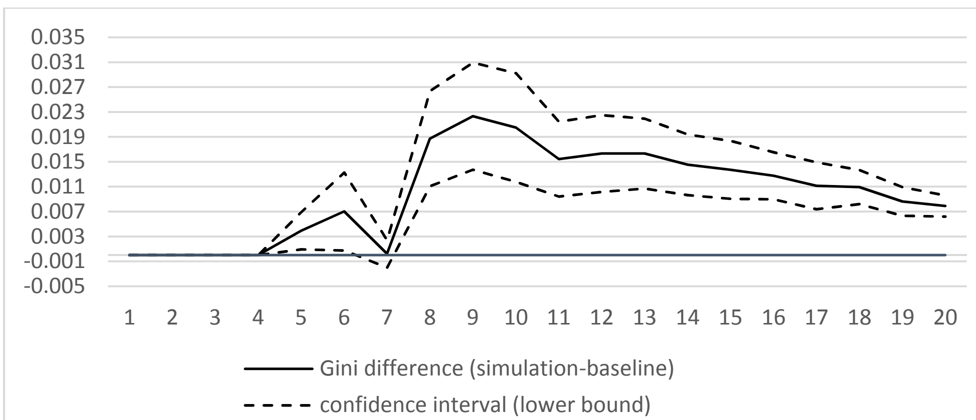
Source: authors' estimations
Note: Graph based on the Excel file "report" ("distindic" sheet)

Figure 2b: Difference between the FGT1 under the simulation and the baseline



Source: authors' estimations
Note: Graph based on the Excel file "report" ("difgtindic" sheet)

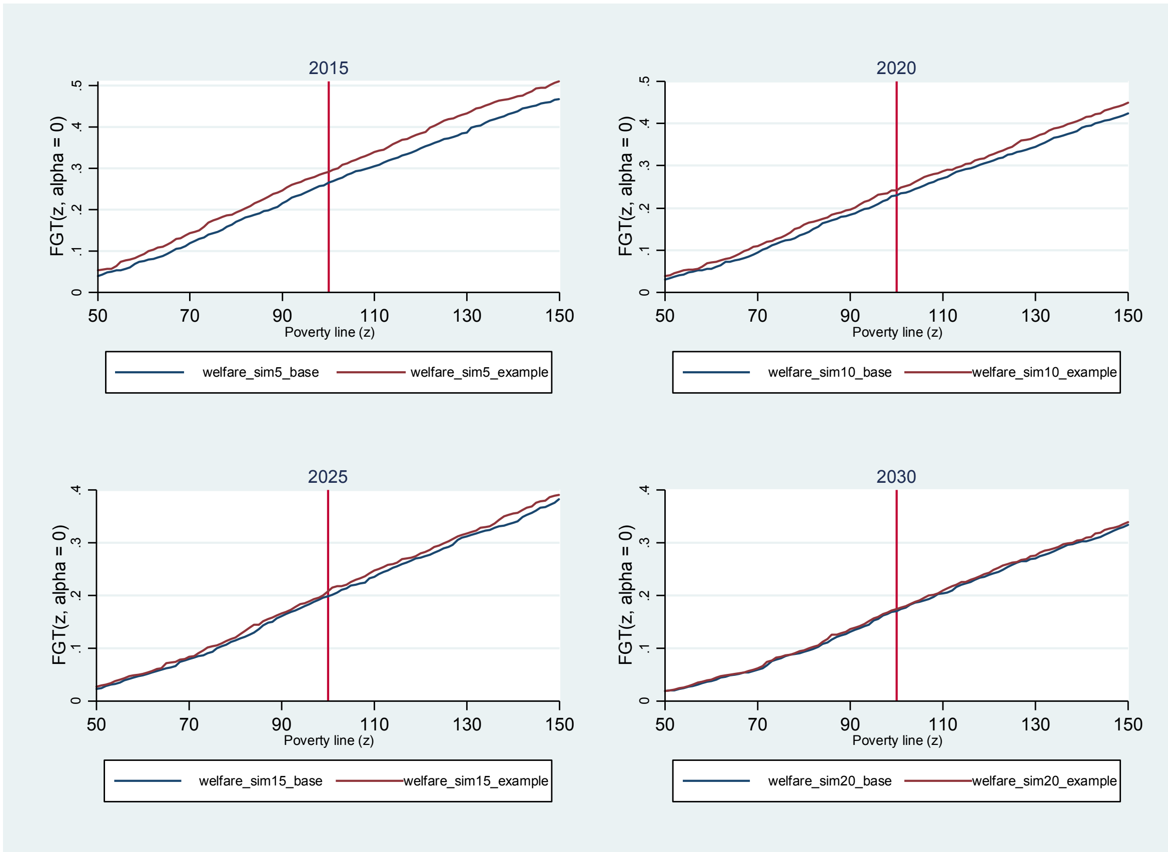
Figure 2c: Difference between the Gini under the simulation and the baseline



Source: authors' estimations
Note: Graph based on the Excel file "report" ("diginiindic" sheet)

Our results are robust over a large range of possible poverty lines. As shown in Figure 3, the FGT0 curve associated to the baseline always dominates (i.e. lies below or poverty is lower than) the simulation FGT0 curve. In a separate analysis, we found (see "cfgts2d" graphs) that for the majority of poverty lines in the adopted range the difference between the simulation and baseline curves are statistically different from zero. As said earlier, starting from 2017, the baseline curve does not clearly dominates the simulation curve.

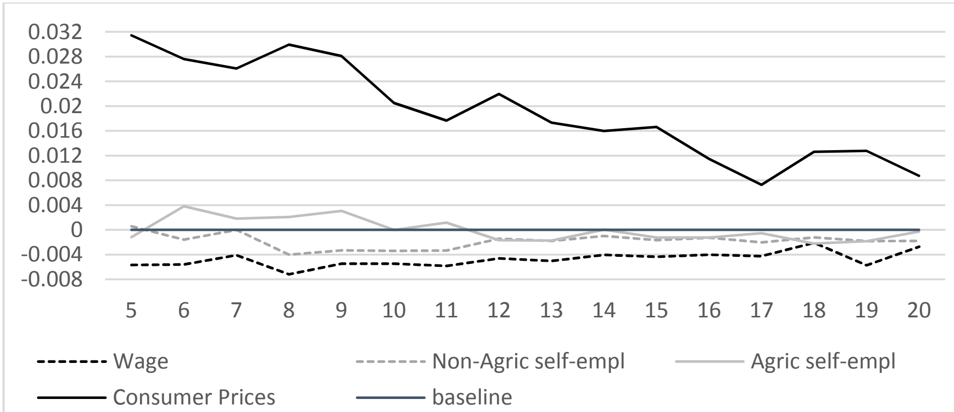
Figure 3: FGT0 curves for selected years (5, 10, 15 and 20)



Source: authors' estimations
Note: Graph based on "cfgt" (5, 10, 15 and 20)

As discussed above, the proposed policy intervention would particularly affect households during the first years of implementation. Out of a total increase by 2.5 percentage points in 2015 between the simulation and baseline scenarios, Figure 4 showns that 3.2 points are attributable to the large rise in consumer prices due to the investment. While the simulated changes in the income from the agricultural and non-agricultural sectors do not significantly affect the poverty difference, variations in wage would reduce poverty under the simulation scenario (with respect to the baseline) by around 0.5 points. Between 2016 and 2021, the change in agricultural income would negatively affect poverty (with respect to its baseline counterfactual); after that, it would slighly decrease or have no effects on poverty. Income from non-agricultural self-employment would help – though marginally (0.2-0.4 points a year) – decrease poverty from 2018. Wages would help to reduce poverty along the entire period by 0.2 to 0.8 points a year. Consumer prices are the major factor of the increase in headcount poverty in the whole period, even if their (negative) contribution decreases over time, falling from a contribution of 3.2 points in 2015 to 0.8 points in 2030.

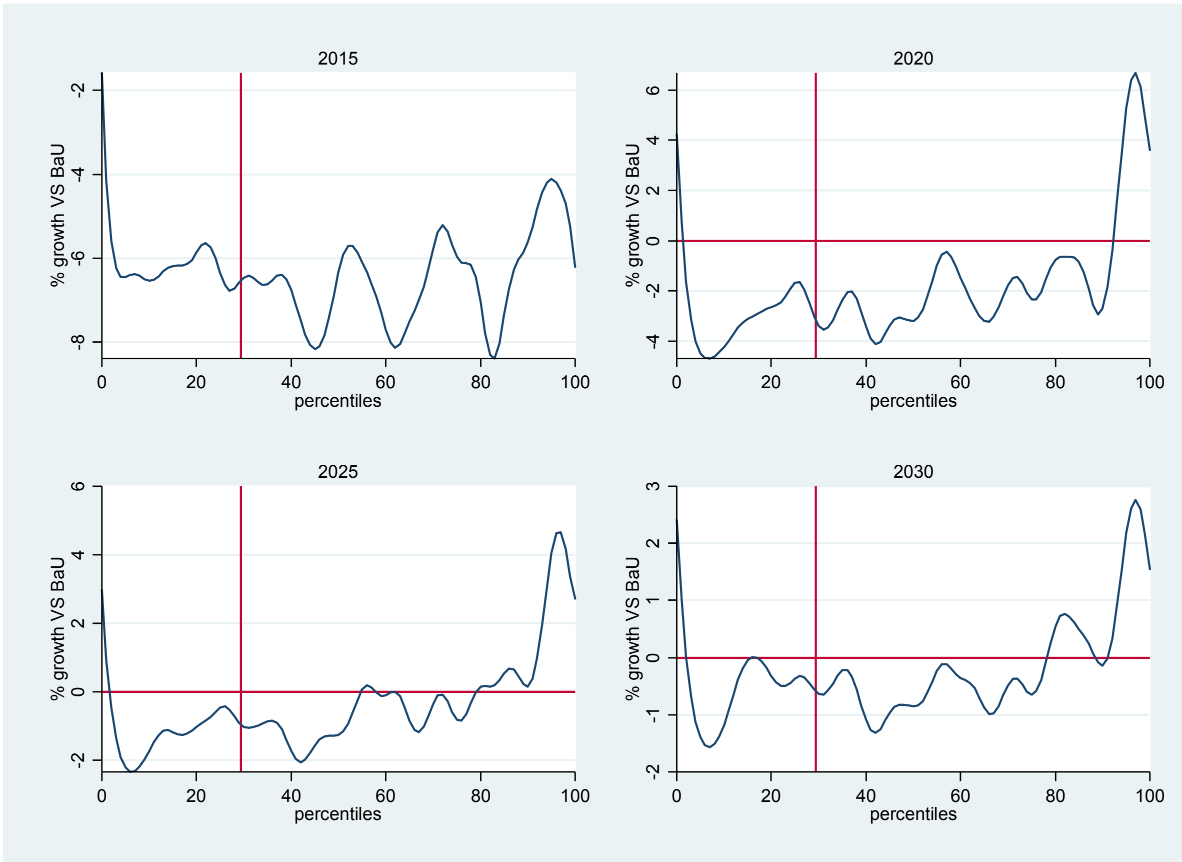
Figure 4: Decomposition of FGT0 by income factors



Source: authors' estimations
Note: Graph based on the Excel file "report" ("fgtdecomp" sheet)

Finally, we propose to look at the distributive effects along the whole distribution for some selected years. This is done by using Growth Incidence Curves, which draw the percentage change of consumption under the simulation scenario (with respect to the baseline – normalised to zero in the graphs below) for all the percentiles in our population. As shown by Figure 5, soon after the policy implementation (in 2015) all percentiles would be negatively affected. The change in consumption is indeed always negative, with those in percentiles 40th to 80th being particularly affected (up to -8%), and with those below the poverty line experiencing a 6% reduction. In later periods, richer percentiles (around the top 10%) would have higher income than under the baseline; in contrast, lower percentiles would continue to experience some reduction, though at a smaller rate over time (about -3% in 2020, -2% in 2025 and -1% in 2030) and without significant differences between those lying below the poverty line and those above. Overall, the results shown below confirm an increase in inequality over time, with a rise by 0.8 points by 2030.

Figure 5: Growth Incidence Curves for selected years



Source: Authors' estimations
Note: Graphs based on "cnpe" (5, 10, 15 and 20)

References

Azevedo, J.P., G. Inchauste, S. Olivieri, J. Saavedra and Winkler, H. (2013) "Is Labor Income Responsible for Poverty Reduction? A Decomposition Approach", World Bank Policy Research Working Paper 6414, Washington DC.

Bourguignon, F., M. Fournier and M. Gurgand (2007). "Selection Bias Corrections Based On The Multinomial Logit Model: Monte Carlo Comparisons," Journal of Economic Surveys, Wiley Blackwell, vol. 21(1), pages 174-205, 02.

Bourguignon, F., M. Fournier and M. Gurgand (2001). "Fast Development with a Stable Income Distribution: Taiwan, 1979-94," Review of Income and Wealth, International Association for Research in Income and Wealth, vol. 47(2), pages 139-63, June.

Bourguignon, F., & Ferreira, F. (2005). "Decomposing changes in the distribution of household incomes: methodological aspects" in Bourguignon, Ferreira Lustig (eds.) The microeconomics of income distribution dynamics in East Asia and Latin America, 17-46.

Bourguignon, F. and L. Savard (2008). "A CGE Integrated Multi-Household Model with Segmented Labour Markets and unemployment" in Bourguignon, Pereira Da Silva and Bussolo (eds.) (2008) The Impact of Macroeconomic Policies on Poverty and Income Distribution, The World Bank, Oxford University Press.

Chen, S., & Ravallion, M. (2004). Welfare impacts of China's accession to the World Trade Organization. The World Bank Economic Review, 18(1), 29-57.

Cockburn, J., V. Robichaud and L. Tiberti (2017) "Energy Subsidy Reform and Poverty in Arab Countries: A Comparative CGE-Microsimulation Analysis of Egypt and Jordan", forthcoming in Review of Income and Wealth

Cockburn, J., L. Savard and L. Tiberti (2014) "Macro-Micro Models" by in C. O'Donoghue (ed.) Handbook of Microsimulation Modelling (Contributions to Economic Analysis, Volume 293) Emerald Group Publishing Limited, pp.275 – 304

Colombo, G (2010). "Linking CGE and Microsimulation Models: A Comparison of Different Approaches", International Journal of Microsimulation 3(1): 72-91.

Deville, J.C., and Särndal, C.E. (1992).Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376-382.

Grimm, M. (2005). Educational policies and poverty reduction in Côte d'Ivoire. Journal of Policy Modeling, 27(2), 231-247.

Huppi, M., & Ravallion, M. (1991). The sectoral structure of poverty during an adjustment period: evidence for Indonesia in the mid-1980s. World development, 19(12), 1653-1678.

Meagher, G. A., & Agrawal, N. (1986). Taxation reform and income distribution in Australia. Australian Economic Review, 19(3), 33-56.

Mookherjee, D., & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. The Economic Journal, 92(368), 886-902.

Ravallion, M., & Chen, S. (2003). Measuring pro-poor growth. Economics letters, 78(1), 93-99.

Robilliard, A. S., Bourguignon, F., & Robinson, S. (2008). Examining the social impact of the Indonesian financial crisis using a macro-micro model. Bourguignon, F., M. Bussolo, and LA Pereira da Silva.

Robilliard, A. S., & Robinson, S. (2003). Reconciling household surveys and national accounts data using a cross entropy estimation method. Review of Income and Wealth, 49(3), 395-406.

Shorrocks, A. (2013) "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value", Journal of Economic Inequality, 11: 99-126

Appendix A

In this appendix we present all Stata do files that comprise the microsimulation toolkit.

master.do

```
clear all

set more off

* set path
cd "C:\TDBmicrosim\Do-Files"

* open log file
capture log close
log using "..\Output\PEP-Microsimulations.log", replace

/*
notation:
  i = current year of simulation
  k = previous year of simulation (i.e., k=i-1)
  s = type of worker according to his/her skills (s=0 for unskilled; s=1 for
skilled)
  e = employment category (e=1 for wage workers; e=2 for non-agricultural self-
employed; e=3 for farmers; e=4 for not working)
  j = activity sector (1=non-agriculture; 2=agriculture)
  t = loop within the same simulation
  c = commodities (1...30)
*/

* note: in the current version, pov+dist analysis only works with two simul
* named base and example

local sim ///
  base    ///
  example

foreach scen in `sim' {

  * seed for random numbers generator; relevant for replicability
  set seed 121212

  * number of iterations, used in master.do and compute-welfare.doi
  local iternum = 2
  * number of years, used in master.do and pov+dist-analysis.do
  local yrnum = 20


  * start: Definition of Scenario -----------------------------------

//  include "Scen-Files\scen-defn-`scen'.doi"
  include "Scen-Files\scen-defn.doi"

  * end: Definition of Scenario -------------------------------------


  * define the program that generate thes random errors for the mlogit
  * reference is Bourguignon, Fournier and Gurgand (2001)
  do "calibrate-ocup.do"

  * iterate over years
```

```
  forvalues i = 1(1)`yrnum' {

    * if changes are estimated from the preceding year, define the year
    * preceding the current year; otherwise, use the base year - 0 - if changes
    * are with respect to the base year
//  local k = `i' - 1
    local k=0

    * define the loops for sim`i' (i.e., how many times run sim`i' - it is
    * suggested to run it a sufficiently large number of times; e.g., 100)
    forvalues t = 1(1)`iternum' {

      * load data
      use "..\Raw-Data\Uganda2009.dta", clear

      * static ageing = population growth
       * in this example, already included in data -- no need to run pop-
       *growth.doi
      * include pop-growth.doi

      * sim changes in occupational categories
      include "sim-ocup.doi"

      * sim changes in labor incomes
      include "sim-labor-inc.doi"

      * sim changes in total household income
      include "sim-hhd-inc.doi"

      }

    * poverty+dist analysis
    include compute-welfare.doi

  }

}

* poverty and distributive analysis
do pov+dist-analysis.do `yrnum'

log close
```

calibrate-ocup.do

```
capture program drop draw

program define draw
  capture drop u1-u4
  capture drop v1-v4
  capture drop pr

  tempvar i

  local y `1'
  local i 1
  local k 1

  gen v1 = uniform()
  gen v2 = uniform()
```

```
  gen v3 = uniform()
  gen v4 = uniform()

* note: given what is done below, no need to generate u1-u4 as random numbers
  gen u1 = uniform()
  gen u2 = uniform()
  gen u3 = uniform()
  gen u4 = uniform()

  gen ur = 0
  gen pr = 0

  while `i' < 5 {
    replace u`i' = -ln(-pr_`i'*ln(v`i')) if `i' == `y'
    replace ur = u`i' if `i' == `y'
    replace pr = pr_`i' if `i' == `y'
    local i = `i' + 1
  }

  while `k' < 5  {
    replace u`k' = -ln(exp(-ur)*(pr_`k'/pr) - ln(v`k')) if `k' ~= `y'
    local k = `k' + 1
  }

end
```

* pop-growth.doi

```
* this do file shows how to change the population through static ageing
* note: no need to run it; already included in Uganda data
* in this case, it is assumed that population growth is 3%
* generate categorical variables for regions
* in Uganda there are 4 categories -> generate 4 dummy variables
* tabulate region, gen(iregion)
* make a row matrix of population totals (by regions)

* read from Excel
preserve
import excel using "Scen-Files\uga2010-base-defn.xlsx", sheet("pop")
cellrange(A2) firstrow clear
* each yr = one matrix
forvalues i = 1(1)`yrnum' {
  mkmat reg* if _n==`i', matrix(M`i')
}
restore

forvalues r = 1(1)20 {
  calibrate , marginals(iregion*) poptot(M`r') entrywt(wgt09_sim0)
exitwt(wgt09_sim`r')
}
```

sim-ocup.doi

```
* initialize ocuppational categories and skills for t=i using t=i-1=k
* in this case, k=0 is hard-coded
gen worker_sim`i'  = worker_sim0
gen skilled_sim`i' = skilled_sim0
```

```
* iterate over skills (0=unskilled; 1=skilled)
forvalues s=0(1)1 {

  * iterate over occupational categories
  * (e=1 for wage workers; e=2 for non-agricultural self-employed; e=3 for
  * farmers; e=4 for not working)
  forvalues e = 1(1)3 {


  * start: transform matrix with shocks as shares in percentage changes
  * note: not needed if shocks in scen-defn-sim.doi are expressed as percentage
  * changes


/*
    tabulate worker_sim0 [iw=wgt09_sim0] if skilled_sim0==`s' & worker_sim0!=.
    local tot_force_sim0 = r(N)

    tabulate worker_bau`i' [iw=wgt09_sim`i'] if skilled_bau`i'==`s' &
worker_bau`i'!=.
    local tot_force_bau`i' = r(N)

    scalar   chng_`s'`e' =
((`empl_sh_`s'_bau`i'_`e''*`tot_force_bau`i'')/(`empl_sh_`s'_sim0_`e''*`tot_for
ce_sim0'))-1
*/

  * end: transform matrix with shocks as shares in percentage changes


    * start: compute population growth

    tabulate worker_sim0 [iw=wgt09_sim0] if worker_sim0==`e' &
skilled_sim0==`s'
    local abs_0 = r(N)

    tabulate worker_sim`i' [iw=wgt09_sim`i'] if worker_sim`i'==`e' &
skilled_sim`i'==`s'
    local abs_`i' = r(N)

    local diff_`i' = (`abs_`i''/`abs_0')-1

    * end: compute population growth


    * substract population growth
    scalar chng_`s'`e' =  empl_sh_`s'_sim`i'[1, `e'] - `diff_`i''

  }

}


* iterate over skills (0=unskilled; 1=skilled)
forvalues s = 0(1)1 {

  * estimate multinomial logit for occupational category
  include estimate-ocup.doi


  * job queuing
  * for each occupational category, sort individuals according to their
```

```
  * estimated probability

  * iterate over occupational categories
  forvalues e = 1(1)3 {

    * the procedure to single out individuals that change their ocuppational
    * category is different depending on whether the change is positive or
    * negative; thus, two cases are considered


    *--- start: case 1 - if decrease in occupational category `e'

    * if employment in "e" decreases by "chng_s_e" ("n" people), replace the
    * employment status with 4 for the "n" people employed in this sector
    * AND with the lowest probability of being employed in this sector
    if chng_`s'`e'  <= 0 {
      display as text "change for s=`s' and e=`e' is <=0"

      * estimate change% (jobs lost) in status `e' wrt the worker status in
      * sim`i' (with the population weight in sim`i')
      tabulate worker_sim0 [iw=wgt09_sim0] if worker_sim0==`e' &
skilled_sim0==`s'
      local perc = abs(chng_`s'`e')/100

      * identify the percentile of the probability (as predicted above)
      * corresponding to the change% just estimated, amongst (un)skilled
      * workers in sector `e'
      centile pr`e' if worker_sim0==`e' & skilled_sim`i'==`s', centile(`perc')

      * replace with 4 (as there is a job lose) the worker status to workers
      * showing a probability lower than the one identified by the centile
      replace worker_sim`i'=4 if ((pr`e'<=r(c_1) & pr`e'!=.) &
(worker_sim0==`e')) & skilled_sim`i'==`s'

      * ONLY for negative changes: generate a variable identifying individuals
      * moving from employment status 1 to employment 4 - to be used later
      gen migr_`e'4=(worker_sim0==`e' & worker_sim`i'==4) & skilled_sim`i'==`s'

    }


    *--- end: case 1 - if decrease in occupational category `e'


    *--- start: case 2 - if increase in occupational category `e'

    * if employment in `e' increases by "chng_s_e" ("n" people), replace the
    * employment status with 1 for the "n" people not employed in this sector
    * AND with the highest probability of being employed in this sector
    if chng_`s'`e'  > 0 {
      display as text "change for s=`s' and e=`e' is >0"

      * estimate the change% (new jobs) in status `e' wrt worker status in
      * sim0 with the population weight in sim0
      tabulate worker_sim0 [iw=wgt09_sim0] if worker_sim0==`e' &
skilled_sim0==`s'
      local absolute_change = r(N)*abs(chng_`s'`e')
      tabulate worker_sim0 [iw=wgt09_sim0] if worker_sim`i'!=`e' &
worker_sim`i'!=. & skilled_sim`i'==`s'
      local perc = (r(N) - `absolute_change') / r(N) * 100

      * identify the percentile of the probability (as predicted above)
      * corresponding to the % change as just estimated, amongst (un)skilled
```

```
      * people who did not work in `e'
      centile pr`e' if worker_sim0!=`e' & worker_sim0!=. & skilled_sim`i'==`s',
centile(`perc')

      * replace with `e' the status of (un)skilled people who did not work in
      * `e', showing a probability higher than the one identified by the
centile
      replace worker_sim`i'=`e' if pr`e'>=r(c_1) & worker_sim0!=`e' &
worker_sim0!=. & pr`e'!=. & skilled_sim`i'==`s'
    }

    *--- end: case 2 - if increase in occupational category `e'

  }


  * final adjustments necessary to satisfy macro changes
  * we have first left people adjusting across working categories (1,2,3)
  * according to their utility (as done above) and then we search among
  * non-working people to fill the remaining jobs which are needed to satisfy
  * the macro changes

  * iterate over occupational categories
  forvalues e = 1(1)3 {

    * estimate the absolute number of changes based on the new occupational
    * status and then the difference needed to satisfy the macro results
    tabulate worker_sim0 [iw=wgt09_sim0] if skilled_sim0==`s' &
worker_sim0==`e'
    local tot_force_sim0_`e'=r(N)

    tabulate worker_sim`i' [iw=wgt09_sim0] if skilled_sim`i'==`s' &
worker_sim`i'==`e'
    local tot_force_sim`i'_`e'=r(N)

    scalar   chng_`s'`e'_abs = `tot_force_sim0_`e''*(1+chng_`s'`e') -
`tot_force_sim`i'_`e''
  }


  * for each working occupational choices (e=1,2,3) allow exceeding (missing)
  * workers to move towards (from) non-working (working) status

  * iterate over occupational categories
  forvalues e = 1(1)3 {


   * case 1 - exceeding workers in sector `e'
    if (chng_`s'`e'_abs < 0) {

      * define the "exceeding workers" rate and use (100-) this rate to
identify
      * the percentile of probabilities. Again, such an approach is followed to
      * select workers moving out of sector `e'
      tab worker_sim`i' [iw=wgt09_sim0] if worker_sim`i'==`e' &
skilled_sim`i'==`s'
      local perc = 100 - ((r(N)+chng_`s'`e'_abs)*100/r(N))
      centile pr`e' if worker_sim`i'==`e' & skilled_sim`i'==`s',
centile(`perc')
      replace worker_sim`i'=4 if pr`e'<r(c_1) & pr`e'!=. & skilled_sim`i'==`s'
& worker_sim`i'==`e'
    }
```

```
    * case 2 - missing workers in sector `e'
    if chng_`s'`e'_abs > 0 {

      * define the "missing workers" rate and use this rate to identify the
      * percentile of probabilities. Again, such an approach is followed to
      * select workers moving to sector `e'
      tabulate worker_sim`i' [iw=wgt09_sim0] if worker_sim`i'==4 & pmax`e'==1&
skilled_sim`i'==`s'
      local perc = (r(N)-chng_`s'`e'_abs)*100/r(N)
      centile pr`e' if worker_sim`i'==4 & pmax`e'==1 & skilled_sim`i'==`s',
centile(`perc')
      replace worker_sim`i'=`e' if pr`e'>=r(c_1) & worker_sim`i'==4 & pr`e'!=.
& pmax`e'==1 & skilled_sim`i'==`s'
    }


  }

  * for the changes in self-employment revenues (which is estimated at the
  * household level), we need to aggregate new employment status in sectors 2
  * and 3 at the household level



  preserve

  keep if skilled_sim`i'==`s'
  keep hhid pid worker_sim`i' skilled_sim`i'

  gen worker2_`s'_sim`i' = worker_sim`i'==2
  gen worker3_`s'_sim`i' = worker_sim`i'==3

  sort hhid
  by hhid: egen n_worker2_`s'_sim`i' = total(worker2_`s'_sim`i')
  by hhid: egen n_worker3_`s'_sim`i' = total(worker3_`s'_sim`i')
  drop worker2_`s'_sim`i' worker3_`s'_sim`i'
  save "..\Output\skilled`s'_sim`i'.dta", replace

  restore


}

* we have now estimated the new employment statuses in sim`i' - we create a
* single dataset including skilled and unskilled workers and then merge to
* the original dataset
use "..\Output\skilled0_sim`i'.dta", clear
append using "..\Output\skilled1_sim`i'.dta"
merge 1:1 hhid pid using "..\Raw-Data\Uganda2009.dta"
drop _merge

* show employment transition matrix and check the simulation results
tabulate worker_sim0 worker_sim`i'
tabulate worker_sim0 worker_sim`i' if skilled_sim0==1
tabulate worker_sim0 worker_sim`i' if skilled_sim0==0
tabulate worker_sim`i' [iw=wgt09_sim`i'] if skilled_sim0==1
tabulate worker_sim`i' [iw=wgt09_sim`i'] if skilled_sim0==0
```

sim-labor-inc.doi

```
* start: wages for unskilled and skilled workers ------------------------------

* iterate over skills (0=unskilled; 1=skilled)
forvalues s = 0(1)1 {

  preserve

  keep if skilled_sim`i'==`s'
  gen salaried_sim`i'=(worker_sim`i'==1)

  * Heckman selection model
  xi: heckman ln_wage_`s' gender hh_head i.region age_sim0 age2_sim0 education,
/// select(salaried_sim0=married age_sim0 gender i.region n_children)

  * save standard errors of residuals
  global sigw_`s' = e(sigma)

  * predict log wage
  * option ycond = E(y | y observed)
  predict lnwage_fitted_`s'_sim`i', ycond

  * generate residuals for individuals with observed wage
  gen res_`s'=ln_wage_`s' - lnwage_fitted_`s'_sim`i' if e(sample)

  * generate residuals for individuals without (unobserved) wage -- draw
  * randomly from a normal distribution with the relevant (skilled or
  * unskilled) observed variance
  replace res_`s' = invnorm(runiform())*$sigw_`s' if ln_wage_`s'==.

  * convert ln wages to level values
  gen wage_`s'_fitted_sim`i' = exp(lnwage_fitted_`s'_sim`i' + res_`s')
  gen wage_`s' = exp(ln_wage_`s')

  * wage changes
  * chng_wage_1 = change in wage for skilled workers
  * chng_wage_0 = change in wage for unskilled workers
  if `s'==0 {
    scalar   chng_wage_`s' = chng_wage_sim`i'[1, 2]
  }

  if `s'==1 {
    scalar   chng_wage_`s' = chng_wage_sim`i'[1, 1]
  }

  * assign fitted wages for individuals with missing wage
  replace wage_`s'=wage_`s'_fitted_sim`i' if worker_sim0==1 & wage_`s'==.
  gen wage_`s'_hat=0 if worker_sim0==1 & worker_sim`i'!=1

  * initialize the simulated wage
  replace wage_`s'_hat = wage_`s'_fitted_sim`i' if worker_sim0==1 &
worker_sim`i'==1
  replace wage_`s'_hat = wage_`s'_fitted_sim`i' if worker_sim0!=1 &
worker_sim`i'==1
  replace wage_`s' = 0 if wage_`s'==. | wage_`s'==1
  replace wage_`s'_hat = 0 if wage_`s'_hat==. | wage_`s'_hat==1

  gen wage_`s'_sim`i' = wage_`s' if worker_sim0==1 & worker_sim`i'==1
  replace wage_`s'_sim`i'=wage_`s'_hat if worker_sim0!=1 & worker_sim`i'==1

  * compute the average wages in sim`k' and in sim`i' and save them
```

```
  sum wage_`s' [aw=wgt09_sim`k'] if worker_sim0==1 & wage_`s'>1
  scalar imean = r(mean)

  sum wage_`s'_sim`i' [aw=wgt09_sim`i'] if worker_sim`i'==1 & wage_`s'_sim`i'>1
  scalar ffmean = r(mean)


  * calibrate the change of wages of all workers in skill `s'
  * Note: if less productive people move out between 0 and `i', even if macro
  * changes in wages are positive we might be forced to downward adjust wages
  * of workers in time `i', so that some individuals wage might be lower than
  * that observed in time 0 --> generating delta<0
  global agwage = (1+chng_wage_`s')*(imean/ffmean)
  replace wage_`s'_sim`i'=wage_`s'_sim`i'*$agwage
  sum wage_`s'_sim`i' [aw=wgt09_sim`i'] if worker_sim`i'==1 & wage_`s'_sim`i'>1

  * compute the absolute change in wage
  gen delta_`s' = wage_`s'_sim`i' - wage_`s'
  replace delta_`s' = 0 if delta_`s'==.
  replace delta_`s' = -wage_`s' if worker_sim0==1 & worker_sim`i'!=1

  * sum wage changes of all household members and collapse to one observation
  * per household
  collapse(sum) delta_`s', by(hhid)
  save "..\Output\wage_after`s'.dta", replace

  restore

}

* merge delta in wages for skilled and unskilled together
use "..\Output\wage_after0.dta", clear
merge 1:1 hhid using "..\Output\wage_after1.dta"
drop _merge

* changes in wages at the household level
egen delta_hh_wage_sim`i'_`t'=rsum(delta_1 delta_0)
save "..\Output\wage_after_sim`i'.dta", replace


* end: wages for unskilled and skilled workers --------------------------------

* start: profits for non-agricultural self-employed and farmers ---------------

* iterate over occupational categories
* non-agricultural self-employed (e=2) and farmers (e=3)
forvalues e = 2(1)3 {

  * load the (individual level) datasets with new occupational status
  * (skilled and unskilled)
  use "..\Output\skilled0_sim`i'.dta", clear
  append using "..\Output\skilled1_sim`i'.dta"

  * "collapse" to a household level dataset
  collapse n_worker`e'_1_sim`i' n_worker`e'_0_sim`i', by(hhid)
  replace n_worker`e'_1_sim`i'=0 if n_worker`e'_1_sim`i'==.
  replace n_worker`e'_0_sim`i'=0 if n_worker`e'_0_sim`i'==.

  merge 1:m hhid using "..\Raw-Data\Uganda2009_all.dta"
  drop _merge
  *for simplicity, just keep 1 observation per household
```

```
  keep hhid hheduc gender_hh region urban n_worker`e'_sim* n_worker`e'_1_sim*
n_worker`e'_0_sim* ln_hhwork_`e'_sim* ln_hhwork_`e'_1_sim* ln_hhwork_`e'_0_sim*
rev_`e'_tot_0 ln_rev_`e'_tot_0 ln_inp_2_tot_0 hsize fem_ratio mage wgt09_sim0
wgt09*
  ds hhid, not
  collapse `r(varlist)', by(hhid)

  * estimate farmer profit function (Cobb-Douglas production function)
  regress ln_rev_`e'_tot_0 urban i.region gender_hh hheduc ln_hhwork_`e'_1_sim0
ln_hhwork_`e'_0_sim0 [aw=wgt09_sim0*hsize] if rev_`e'_tot>0 & rev_`e'_tot!=.

  /* with instrumental variables
  ivregress 2sls ln_rev_`e'_tot urban i.region gender_hh hheduc
ln_hhwork_`e'_1_sim0 (ln_hhwork_`e'_0_sim0=hsize mage) [aw=wgt09_sim0*hsize] if
rev_`e'_tot>0 & rev_`e'_tot!=., first
  estat endogenous, forceweights
  estat firststage
  */

  * predict the ln revenue
  predict ln_rev_`e'_hat_0

  * generate residuals for those where wage is observed
  predict res`e' if e(sample), r
  * save standard error of residuals
  gen sigw = sqrt(r(Var))


  * generate residuals for those where self-employment revenues are not
  * observed - draw randomly from a normal distribution with the relevant
  * (skilled or unskilled) observed variance
  replace res`e' = invnormal(runiform())*sigw if rev_`e'_tot_0==. |
rev_`e'_tot_0==0

  * replace the number of working hh members with new numbers
  replace ln_hhwork_`e'_1_sim0 = ln(n_worker`e'_1_sim`i') if
n_worker`e'_1_sim`i'!=n_worker`e'_1_sim0
  replace ln_hhwork_`e'_1_sim0 = 0 if ln_hhwork_`e'_1_sim0==.
  replace ln_hhwork_`e'_0_sim0 = ln(n_worker`e'_0_sim`i') if
n_worker`e'_0_sim`i'!=n_worker`e'_0_sim0
  replace ln_hhwork_`e'_0_sim0 = 0 if ln_hhwork_`e'_0_sim0==.

  * predict estimated revenues and change with zero if missing (revenues or
  * workers)
  predict ln_rev_`e'_hat
  gen rev_`e'_hat = exp(ln_rev_`e'_hat+res`e')
  replace rev_`e'_tot = 0 if rev_`e'_tot_`k'==.
  replace rev_`e'_hat = 0 if rev_`e'_hat==.
  replace n_worker`e'_sim0 = 0 if n_worker`e'_sim0==.

  * generate the output per worker - as the CGE provides the remuneration rate
  * per worker
  egen n_worker`e'_sim`i' = rsum(n_worker`e'_1_sim`i' n_worker`e'_0_sim`i')
  replace rev_`e'_tot_`k' = 0 if rev_`e'_tot_`k'==. | rev_`e'_tot_`k'==1
  replace rev_`e'_hat = 0 if rev_`e'_hat==. | rev_`e'_hat==1
  gen revpw_`e'_0 = rev_`e'_tot_`k'/n_worker`e'_sim0
  gen revpw_`e'_hat_`i' = rev_`e'_hat/n_worker`e'_sim`i'
  replace revpw_`e'_hat_`i' = 0 if revpw_`e'_hat_`i'==. | revpw_`e'_hat_`i'==1
  replace revpw_`e'_0 = 0 if revpw_`e'_0==. | revpw_`e'_0==1

  * define the change of revenue for sectors 2 and 3 as scalars
  if `e'==2 {
    scalar   chng_revse_`e' = chng_revse_sim`i'[1, 1]
```

```
  }

  if `e'==3 {
    scalar   chng_revse_`e' = chng_revse_sim`i'[1, 2]
  }

  * estimate the average revenues at `k' and `i'
  * note: individuals with zero income are excluded
  sum revpw_`e'_`k' [aw=wgt09_sim`k'*hsize] if n_worker`e'_sim0>0 &
revpw_`e'_0>1
  scalar imean=r(mean)
  gen revpw_`e'_`i'=revpw_`e'_0 if (n_worker`e'_sim0==n_worker`e'_sim`i') &
n_worker`e'_sim0>0
  replace revpw_`e'_`i'=revpw_`e'_hat_`i' if
(n_worker`e'_sim0!=n_worker`e'_sim`i')

  * Note: if less productive people move out between 0 and `i', even if macro
  * changes in reveues are positive
  * we might be forced to downward adjust revenues in time `i', so that some
  * revenue might be lower than that observed in time 0-->generating delta<0)
  sum revpw_`e'_`i' [aw=wgt09_sim`i'*hsize] if n_worker`e'_sim`i'>0 &
revpw_`e'_`i'>1
  scalar fmean=r(mean)
  global agwage=(1+chng_revse_`e')*(imean/fmean)
  replace revpw_`e'_`i' = revpw_`e'_`i'*$agwage
  sum revpw_`e'_`i' [aw=wgt09_sim`i'*hsize] if n_worker`e'_sim`i'>0 &
revpw_`e'_`i'>1

  * generate the household revenue
  gen rev`e'_`i' = revpw_`e'_`i'*n_worker`e'_sim`i'

  * estimate the change between `i' and `k'
  gen delta_`e'_sim`i'_`t' = rev`e'_`i' - rev_`e'_tot_`k'
  replace delta_`e'_sim`i'_`t'=0 if delta_`e'_sim`i'_`t'==.

  keep hhid delta_`e'_sim`i'_`t'
  save "..\Output\rev`e'_sim`i'.dta", replace


}


* end: profits for non-agricultural self-employed and farmers -----------------
```

sim-hhd-inc.doi

```
* compute total household income
use "..\Output\wage_after_sim`i'.dta", clear
merge 1:1 hhid using "..\Output\rev2_sim`i'.dta"
drop _merge
merge 1:1 hhid using "..\Output\rev3_sim`i'.dta"
drop _merge

* express changes in income in monthly terms
* note: monthly poverty line is used
replace delta_hh_wage_sim`i' = delta_hh_wage_sim`i'/12
replace delta_2_sim`i' = delta_2_sim`i'/12
replace delta_3_sim`i' = delta_3_sim`i'/12

* compute total change in household revenues from working activities
```

```
egen delta_rev_sim`i'_`t' = rsum(delta_hh_wage_sim`i' delta_2_sim`i'
delta_3_sim`i')

keep hhid delta_rev_sim`i'_`t' delta_hh_wage_sim`i'_`t' delta_2_sim`i'_`t'
delta_3_sim`i'_`t'
save "..\Output\delta_rev_sim`i'_`t'.dta", replace


* build database for poverty and distributive analysis

* load raw database and merge with results database
use "..\Raw-Data\Uganda2009.dta", clear
merge m:1 hhid using "..\Output\delta_rev_sim`i'_`t'.dta"

replace delta_rev_sim`i'_`t'=0 if delta_rev_sim`i'_`t'==.
replace delta_hh_wage_sim`i'=0 if delta_hh_wage_sim`i'==.
replace delta_2_sim`i'=0 if delta_2_sim`i'==.
replace delta_3_sim`i'=0 if delta_3_sim`i'==.

* geographical price deflator
* note: Kampala is the region with highest PL
* defl is regional price deflator; from regions other than Kampala prices to
* Kampala prices
sum pov_line
gen defl = r(max)/pov_line
replace pov_line = r(max)

* PL is defined at 2005/06 prices
* cpexp30/nrrexp30 is price deflator; from 2009/10 prices to 2005/06 prices


* estimate the welfare variable for sim`i' in iteration `t'
* -> add delta revenues (after spatial price differences are taken into
account)
* to the base welfare variable
* adjustment of expenditures and computation of the incompressible demand
* note: (cpexp30/nrrexp30) should be the 2005-2009 price deflator

forvalues c = 1(1)30 {

  replace exp_`c' = exp_`c'*(cpexp30/nrrexp30)*defl

}


forvalues c = 1(1)30 {

  replace exp_`c' = exp_`c'/equiv

}

* welfare = total consumption spending
* baseyr welfare
egen welfare = rowtotal(exp_1-exp_30)

* generate welfare variable for sim
gen welfare_sim`i'_`t' = welfare +
(delta_rev_sim`i'_`t'*(cpexp30/nrrexp30)/equiv)*defl

* generate budget shares
forvalues c = 1(1)30 {
  gen w_`c'=exp_`c'/welfare
}
```

```
* (cpexp30/nrrexp30) is the 2005-2009 price deflator
* poverty line is in 2005/05 prices

replace delta_rev_sim`i'_`t' =
delta_rev_sim`i'_`t'*(cpexp30/nrrexp30)*defl/equiv

replace delta_hh_wage_sim`i'_`t' =
delta_hh_wage_sim`i'*(cpexp30/nrrexp30)*defl/equiv
replace delta_2_sim`i'_`t' = delta_2_sim`i'_`t'*(cpexp30/nrrexp30)*defl/equiv
replace delta_3_sim`i'_`t' = delta_3_sim`i'_`t'*(cpexp30/nrrexp30)*defl/equiv

* why make initial prices equal to 1?
forvalues c = 1(1)30 {

  gen double p_ref_`c' = 1

}


forvalues c = 1/30 {
  scalar chng_cons_p_sim`i'_`c' = chng_cons_p_sim`i'[1,`c']
  qui gen p_sim`i'_`c' = (1 + chng_cons_p_sim`i'_`c')*p_ref_`c'
}

* generate household specific price index assuming a Cobb-Douglas utility
function
gen double cpi_sim`i' = 1
forvalues c = 1(1)30 {
  replace cpi_sim`i' = cpi_sim`i'*((p_sim`i'_`c'/p_ref_`c')^w_`c')
}

* compute equivalent income after simulation
replace welfare_sim`i'_`t'= welfare_sim`i'_`t'/cpi_sim`i'
qui replace         welfare_sim`i'_`t'        = 0 if welfare_sim`i'_`t'<0

*
*bacon welfare_sim`i'_`t', generate(out1) percentile(1)
*gen zero=(welfare_sim`i'_`t'==0)

*drop if out1==1
*drop if zero==1

keep hhid pid welfare welfare_sim`i'_`t' delta_hh_wage_sim`i'_*
delta_2_sim`i'_* delta_3_sim`i'_* cpi_sim`i'
save "..\Output\welfare_`t'.dta", replace
```

compute-welfare.doi

```
use "..\Raw-Data\Uganda2009.dta", clear

sum pov_line
replace pov_line = r(max)

forvalues t = 1(1)`iternum' {
  merge 1:1 hhid pid using "..\Output\welfare_`t'.dta"
  keep if _merge==3
  drop _merge
}

* note: median and not mean is computed
```

```
* estimate the welfare variable in sim`i' by taking the mean value of all `t'
* iterations
egen welfare_sim`i' = rowmedian(welfare_sim`i'_*)
egen delta_hh_wage_sim`i' = rowmedian(delta_hh_wage_sim`i'_*)
egen delta_2_sim`i' = rowmedian(delta_2_sim`i'_*)
egen delta_3_sim`i' = rowmedian(delta_3_sim`i'_*)
sort welfare, stable
gen perc = sum(wgt09_sim0)
replace perc = perc/perc[_N]
gen perc_sim0 = ceil(perc*100)

* welfare is multiplied by 100 and divided by PL; then, use 100 as PL
replace welfare = welfare*100/pov_line
replace welfare_sim`i' = welfare_sim`i'*100/pov_line
replace delta_hh_wage_sim`i' = delta_hh_wage_sim`i'*100/pov_line
replace delta_2_sim`i' = delta_2_sim`i'*100/pov_line
replace delta_3_sim`i' = delta_3_sim`i'*100/pov_line

* clone simulated welfare and ??delta?? variables so that they are linked to a
* specific scenario
clonevar welfare_sim`i'_`scen' = welfare_sim`i'


* variable sim is used to idenfity scenarios in pov+dist-analysis.do
gen sim = "`scen'"

save "..\Output\welfare_sim`i'-`scen'.dta", replace

* save data for decomposition analysis
keep hhid pid sim wgt09_sim`i' ea stratum welfare welfare_sim`i' ///
  delta_hh_wage_sim`i' delta_2_sim`i' delta_3_sim`i' cpi_sim`i'
gen z=100
save "..\Output\sim`i'-`scen'.dta", replace
```

pov+dist-analysis.do

```
args yrnum

* to test:
if "`yrnum'" == "" {
  local yrnum = 5
  capture log close
  log using "..\Output\tmp-rep.log", replace
}


drop _all


* define matrix to store simulation results; one row per year
local dum = `yrnum'
matrix fgtdecomp = J(`dum',7,.)
matrix colnames fgtdecomp = time welfare wage nonWageNonAg nonWageAg CPI Total
matrix list fgtdecomp

* initialize counter; second row of matrix distindic
local cnt = 1


* decomposition analysis by income factors (Shapley rule - adecomp routine)
forvalues i=1(1)`yrnum' {
```

```
  use "..\Output\sim`i'-base.dta", clear

  summ welfare*

  append using "..\Output\sim`i'-example.dta"
  * some minor adjustments are necessary as the sum of factors sometimes do not
  * give exactly the new real income
  * (i.e., the sum of median values!=median value)
  gen check_`i' = (welfare + delta_hh_wage_sim`i' + delta_2_sim`i' +
delta_3_sim`i') / cpi_sim`i'
  gen diff_`i'_ = (welfare_sim`i' - check_`i') * cpi_sim`i'
  gen diff_`i'  = (welfare_sim`i' - check_`i') * cpi_sim`i'
  egen sum_`i'  = rsum(delta_hh_wage_sim`i' delta_2_sim`i' delta_3_sim`i')

  foreach var of varlist delta_hh_wage_sim`i' delta_2_sim`i' delta_3_sim`i' {
    gen w_`var'   = `var'/sum_`i'
    replace `var' = `var' + w_`var'*diff_`i' if diff_`i'_!=0
    replace `var' = 0 if `var'==.
  }

  * adecomp does not allow the use of string variable in by option
  egen sim_alt = group(sim)
  adecomp welfare_sim`i' welfare delta_hh_wage_sim`i' delta_2_sim`i'
delta_3_sim`i' cpi_sim`i' ///
    [w=wgt09_sim`i'], id(pid) by(sim_alt) eq((c1+c2+c3+c4)/c5) varpl(z)
in(fgt0) std
  * start: save results
  matrix tmp = r(b)
  matrix fgtdecomp[`cnt',1] = `i'
  matrix fgtdecomp[`cnt',2] = tmp[1,3]
  matrix fgtdecomp[`cnt',3] = tmp[2,3]
  matrix fgtdecomp[`cnt',4] = tmp[3,3]
  matrix fgtdecomp[`cnt',5] = tmp[4,3]
  matrix fgtdecomp[`cnt',6] = tmp[5,3]
  matrix fgtdecomp[`cnt',7] = tmp[6,3]

  * end: save results

  * increase counter
  local cnt = `cnt' + 1
}


* the following commands use DASP
* to install DASP, follow the instructions below)
/*
net from http://dasp.ecn.ulaval.ca/modules/DASP_V2.3/dasp
net install dasp_p1, force
net install dasp_p2, force
net install dasp_p3, force
net install dasp_p4, force
*/

* declare survey design
svyset ea [pweight=wgt09_sim`i'], strata(stratum)

* define matrix to store dist results; one row per year
local dum = `yrnum'+1
matrix distindic = J(`dum',7,.)
matrix list distindic
matrix colnames distindic = time FGT0_base FGT0_example FGT1_base FGT1_example
Gini_base Gini_example
```

```
* observed (initial) poverty rate
ifgt welfare, alpha(0) pline(100)
matrix tmp = e(est)
matrix distindic[1,1] = 0
matrix distindic[1,2] = tmp[1,1]
matrix distindic[1,3] = tmp[1,1]
* observed (initial) poverty gap rate
ifgt welfare, alpha(1) pline(100)
matrix tmp = e(est)
matrix distindic[1,4] = tmp[1,1]
matrix distindic[1,5] = tmp[1,1]
* observed (initial) gini coeff
igini welfare
matrix tmp = e(est)
matrix distindic[1,6] = tmp[1,1]
matrix distindic[1,7] = tmp[1,1]

* define matrix to store diff gini results; one row per year
matrix diginiindic = J(`dum',9,.)
matrix list diginiindic
matrix colnames diginiindic = time gini_base standardDe_base gini_example
standardDe_example diff ///
standardDe_diff lowBound_diff upperBound_diff ///


* define matrix to store diff fgt results; one row per year
matrix difgtindic = J(`dum',13,.)
matrix list difgtindic
matrix colnames difgtindic = time FGT0_base FGT0_example diff0 standardDev0
lowBound0 upperBound0 ///
FGT1_base FGT1_example diff1 standardDev1 lowBound1 upperBound1


* initialize counter; second row of matrix distindic
local cnt = 2

forvalues i=1(1)`yrnum' {

  use "C:\uga2010\microsim-2017-01-19\Output\welfare_sim`i'-base.dta", clear
  merge 1:1 hhid pid using "..\Output\welfare_sim`i'-example.dta"

  * declare survey design
  svyset ea [pweight=wgt09_sim`i'], strata(stratum)

  ifgt welfare welfare_sim`i'_base welfare_sim`i'_example, alpha(0) pline(100)
  * start: save results
  matrix tmp = e(est)
  matrix distindic[`cnt',1] = `i'       //time
  matrix distindic[`cnt',2] = tmp[2,1]  //fgt0 sim base
  matrix distindic[`cnt',3] = tmp[3,1]  //fgt0 sim example

  ifgt welfare welfare_sim`i'_base welfare_sim`i'_example, alpha(1) pline(100)
  matrix tmp = e(est)
  matrix distindic[`cnt',4] = tmp[2,1]  //fgt1 sim base
  matrix distindic[`cnt',5] = tmp[3,1]  //fgt1 sim example

  igini welfare welfare_sim`i'_base welfare_sim`i'_example
  matrix tmp = e(est)
  matrix distindic[`cnt',6] = tmp[2,1]  //gini sim base
  matrix distindic[`cnt',7] = tmp[3,1]  //gini sim example

  matrix list distindic
  * end: save results
```

```
  display "===######=== Running digini ===######==="

  digini welfare_sim`i'_base welfare_sim`i'_example
  matrix tmp1 = e(d1)
  matrix tmp2 = e(d2)
  matrix tmp3 = e(di)
  matrix diginiindic[`cnt',1] = `i'
  matrix diginiindic[`cnt',2]  = tmp1[1,1] //gini base
  matrix diginiindic[`cnt',3]  = tmp1[1,2] //gini base standard dev
  matrix diginiindic[`cnt',4]  = tmp2[1,1] //gini non-base base
  matrix diginiindic[`cnt',5]  = tmp2[1,2] //gini non-base standard dev
  matrix diginiindic[`cnt',6]  = tmp3[1,1] //gini difference
  matrix diginiindic[`cnt',7]  = tmp3[1,2] //gini difference standard dev
  matrix diginiindic[`cnt',8]  = tmp3[1,5] //gini difference low bound
  matrix diginiindic[`cnt',9]  = tmp3[1,6] //gini difference high bound

  matrix list diginiindic


  display "===######=== Running difgt ===######==="

  difgt welfare_sim`i'_base welfare_sim`i'_example, alpha(0) pline1(100)
pline2(100)
  matrix tmp1 = e(b)
  matrix tmp2 = e(di)
  matrix difgtindic[`cnt',1] = `i'
  matrix difgtindic[`cnt',2] = tmp1[1,1] //FGT0 base
  matrix difgtindic[`cnt',3] = tmp1[1,3] //FGT0 non-base
  matrix difgtindic[`cnt',4] = tmp2[1,1] //Difference
  matrix difgtindic[`cnt',5] = tmp2[1,2] //Standard Dev
  matrix difgtindic[`cnt',6] = tmp2[1,3] //Low Bound
  matrix difgtindic[`cnt',7] = tmp2[1,4] //Upper Bound
  matrix list difgtindic

  difgt welfare_sim`i'_base welfare_sim`i'_example, alpha(1) pline1(100)
pline2(100)
  matrix tmp1 = e(b)
  matrix tmp2 = e(di)
  matrix difgtindic[`cnt',8]  = tmp1[1,1] //FGT1 base
  matrix difgtindic[`cnt',9]  = tmp1[1,3] //FGT1 non-base
  matrix difgtindic[`cnt',10] = tmp2[1,1] //Difference
  matrix difgtindic[`cnt',11] = tmp2[1,2] //Standard Dev
  matrix difgtindic[`cnt',12] = tmp2[1,3] //Low Bound
  matrix difgtindic[`cnt',13] = tmp2[1,4] //Upper Bound
  matrix list difgtindic


  display "===######=== Running cfgt ===######==="

  cfgt welfare_sim`i'_base welfare_sim`i'_example, alpha(0) type(nor) min(50)
max(150) xline(100)
  graph save "..\Output\Graphs\cfgt-`i'.gph", replace

  display "===######=== Running cfgts2d ===######==="

  cfgts2d welfare_sim`i'_base welfare_sim`i'_example, alpha(0) min(50) max(150)
  graph save "..\Output\Graphs\cfgts2d-`i'.gph", replace


  * prepare the data to draw growth incidence curves
```

```
  collapse(mean) welfare_sim`i'_base welfare_sim`i'_example wgt09_sim`i' ea
stratum [aw=wgt09_sim`i'], by(perc_sim0)

  gen ln_perc_base`i' = ln(1 + welfare_sim`i'_base)
  gen ln_perc_example`i' = ln(1+welfare_sim`i'_example)

  gen growth_sim`i' = (ln_perc_example`i' - ln_perc_base`i')*100

  * draw non-parametric curve of growth rates along the whole distribution
  cnpe growth_sim`i', xvar(perc_sim0) min(0) max(100) band(3.0) ytitle(% growth
VS BaU) ///
    xtitle(percentiles) yline(0) title("") ytitle(, size(small)) xtitle(,
size(small)) legend(size(small)) xline(29.4)
  graph save "..\Output\Graphs\cnpe-`i'.gph", replace


  * increase counter
  local cnt = `cnt' + 1

}


*### show and save fgt decomp results
matrix list fgtdecomp
* save results in distindic.xlsx
drop _all
* create variables from matrix fgtdecomp
svmat fgtdecomp, names(col)
export excel using "..\Output\report.xlsx", firstrow(variables)
sheet("fgtdecomp") sheetmodify cell(A2)


*### show and save diff gini results
matrix list diginiindic
* save results in distindic.xlsx
drop _all
* create variables from matrix difgtindic
svmat diginiindic, names(col)
export excel using "..\Output\report.xlsx", firstrow(variables)
sheet("diginiindic") sheetmodify cell(A2)

*### show and save diff fgt results
matrix list difgtindic
* save results in distindic.xlsx
drop _all
* create variables from matrix difgtindic
svmat difgtindic, names(col)
export excel using "..\Output\report.xlsx", firstrow(variables)
sheet("difgtindic") sheetmodify cell(A2)


*### show and save dist results
matrix list distindic

* save results in distindic.xlsx
drop _all
* create variables from matrix distindic
svmat distindic, names(col)
export excel using "..\Output\report.xlsx", firstrow(variables)
sheet("distindic") sheetmodify cell(A2)
```