

Economic Growth and Income Distribution: Linking Macroeconomic Models with Household Survey Data at the Global Level

*Maurizio Bussolo, Rafael E. De Hoyos and Denis Medvedev**

**Development Prospects Group
World Bank**

Do not quote

I. Introduction

This paper details the methodology for generating a counter-factual within-country distribution of income in the GIDD framework. GIDD stands for Global Income Distribution Dynamics and, as the words imply, it consists of a global microsimulation framework. The GIDD's framework is based on ex-ante microsimulation methodologies developed in the recent literature, including Bourguignon and Pereira da Silva (2003); Ferreira and Leite (2003, 2004); Chen and Ravallion (2003); and Bussolo, Lay, and van der Mensbrugghe (2005). The objective of the exercise is to create a hypothetical income distribution capturing three major changes in the structure of the population and the economy: (a) change in the age and skill composition of the population, (b) change in the allocation of workers across sectors in the economy, and (c) change in returns to labor by skill and occupation. Our analysis uses country-specific data at the *micro* level (household surveys) to *simulate* the impact of the three adjustments. Although in reality these changes take place simultaneously, in our simplified framework they are accommodated in a sequential fashion. This paper explains in some detail the steps of the microsimulation model.¹

II. Preliminary Discussion

Ex-post analysis of the distributional effects associated with changes in socio-demographic characteristics had traditionally made use of subgroup inequality decomposition methods (Shorrocks, 1982; Jenkins, 1995). According to this approach, the importance of population characteristic, *A*, is defined by the following counterfactual: *how would income inequality look like if characteristic A was the only difference in the population?* This question can be answer by dividing the population according to

* This paper is part of the Hewlett Foundation Trust Fund Program: Fertility, Reproductive Health, and Socioeconomic Outcomes. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors; they do not necessarily reflect the views of the World Bank, its Executive Directors, or the countries they represent. Corresponding author: Maurizio Bussolo, Development Prospects Group, The World Bank, 1818 H Street Washington, DC., email address: mbussolo@worldbank.org

¹ As a side product of this study, we developed a Stata module and a user guide containing all the necessary procedures to estimate the system described in this paper (see: web page here!)

characteristic A and computing the inequality between average incomes of the population subgroups formed by A .² The importance of A in overall inequality is determinant by the ratio of counterfactual to overall income inequality. Shorrocks (1982) shows that this ratio or proportion is entirely explained by differences in the size of the subgroups with respect the total population (population shares) and the relative incomes across subgroups. The methods presented in this paper provide an analytically consistent framework for ex-ante distributional analysis of *expected* changes in these two elements (population shares and relative incomes across subgroups).

Define I_t as a decomposable measure of total income inequality at time t . Let v_m denote the population share of subgroup m and let μ_m be the ratio between subgroup's m mean income, $\bar{y}_m$, and the average income of the population, $\bar{y}$. Shorrocks (1980) shows that given a partition rule, ξ , dividing total population into $m=(1, \dots, M)$ subgroups, I_t , can be expressed in terms of population shares $\mathbf{v}_t = (v_{1t}, \dots, v_{Mt})$, relative means $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{Mt})$ and inequality values within each subgroup $\mathbf{I}_t^p = (I_{1t}, \dots, I_{Mt})$:

$$I_t = (\mathbf{v}_t, \boldsymbol{\mu}_t, \mathbf{I}_t^p) \quad (2.1)$$

In particular, taking the Theil coefficient as the preferred measured of inequality, total income inequality (I_t) can be expressed as the sum of the un-equalizing effect attributable to income differences between groups and the effect due to income differences *within* the M subgroups:

$$\text{Theil}_t = \sum_{m=1}^M v_{m,t} \mu_{m,t} I_{m,t} + \sum_{m=1}^M v_{m,t} \mu_{m,t} \ln(\mu_{m,t}) \quad (2.2)$$

The first and second terms in the right hand side of equation (2.2) are the within and between subgroups components of inequality, respectively. The between component can be interpreted as the amount of inequality that can be *explained* by partition ξ (Cowell and Jenkins, 1995). Therefore, given partition ξ , ex-post analysis can explain the share of total inequality that is accounted for by differences in the subgroup population shares $\mathbf{v}_t = (v_{1t}, \dots, v_{Mt})$ and differences in relative means $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{Mt})$, i.e. the between component, leaving inequality within subgroups as an *unexplained* component.³

² Notice that for every population characteristic there is correspondent partition rule. For example, if we are interested in the distributional effects of education, the population will be partition in different educational groups.

³ It is trivial to show that as the population partition rule ξ incorporates more characteristics, the *between* component increases. At the limit, when ξ is indeed incorporating all (observable and unobservable) characteristics the number of subgroups is actually equal to the number of individuals and the *between* component is equal to overall income inequality.

Introducing dynamics to equation (2.1) allows us to express changes in total inequality as a function of changes in the three elements defining I_t :

$$I_{t+1} - I_t = \Delta I = (\Delta \mathbf{v}, \Delta \boldsymbol{\mu}, \Delta \mathbf{I}^p) \quad (2.3)$$

While $\mathbf{v}_t$, $\boldsymbol{\mu}_t$ and $\mathbf{I}_t^p$ are known elements in an ex-ante analysis, in an ex-ante approach, equation (2.3) is define by the expected future values of its elements: $\hat{\mathbf{v}}_{t+1}$, $\hat{\boldsymbol{\mu}}_{t+1}$ and $\hat{\mathbf{I}}_{t+1}^p$. As it is the case in an ex-post analysis, the dynamic decomposition will provide us with information about the distributional effects of changes in the characteristics contained in partition rule ξ . However, we cannot say much about what is driving the changes in the within subgroups income inequality, $\mathbf{I}_t^p$, hence we will assume it remains constant over time. In other words, our approach creates a hypothetical income distribution capturing the *ceteris paribus* effect of changes in $\mathbf{v}_t$ and $\boldsymbol{\mu}_t$ given a partition rule ξ . The task in this ex-ante framework is to find the expected values: $\hat{\mathbf{v}}_{t+1}$, $\hat{\boldsymbol{\mu}}_{t+1}$ and incorporate them into the household survey. The resulting income inequality of this exercise will be a hypothetical index, I' capturing the distributional effects of the expected values of $\mathbf{v}_t$ and $\boldsymbol{\mu}_t$.⁴

$$I'_{t+1} = (\hat{\mathbf{v}}_{t+1}, \hat{\boldsymbol{\mu}}_{t+1}, \mathbf{I}_t^p) \quad (2.4)$$

As we mentioned before, economic theory points to three channels through which demographic changes can affect income distribution. The distributional effects of the three linking channels can be capture by the subgroup decomposition method. Firstly, Deaton and Paxson (1994) and Deaton and Paxson (1997) show that as long as the slope of the age-income profile is different from zero, i.e. mean income differ across age groups, aging will increase inequality. As population ages, the relative size of the different age groups become more homogenous, in other words population tends to be evenly distributed among age cohorts. Everything else constant, an equalization of the subgroup population shares, i.e. ν_m in equation (2.2), maximizes the value of the *between* component of inequality. Secondly, as long as different age groups are characterized by different within-group inequality, changes in population shares, ν_m , will affect the *within* component of inequality. The third channel considers the changes in inequality due to changes in the life-cycle income profile, this effect will manifest as changes in the relative means of the subgroup decomposition formula i.e. term μ_m in equation (2.2).

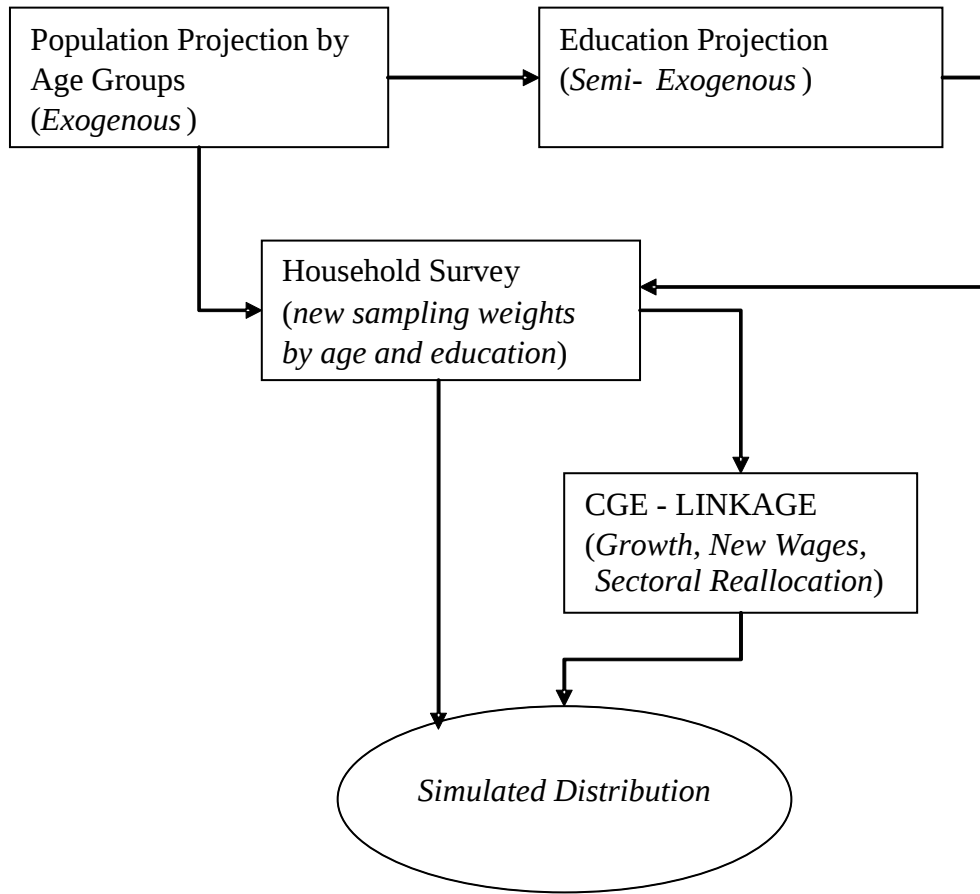
The second strand of the literature that is related both with the methodology developed in the present study and the subgroups decomposition framework is the recent

⁴ Notice how the value of income inequality *within* each subgroup ($\mathbf{I}_t^p$) is the same for the actual and simulated income distribution.

microsimulation decomposition analysis. The microsimulation analysis is basically an extension of the Oaxaca-Blinder decomposition where total income variation is decomposed into changes of observable characteristics (endowment effect), changes in the market price for those characteristics (price effect), and changes in unobservables (Bourguignon et al., 2005). The microsimulation decomposition finds a parallelism with the (non-parametric) framework so far discussed. The endowment effect is equivalent to the distributional impact caused by a change in population shares ($\Delta \mathbf{v}$); the price effect resembles the inequality impact of changes in mean incomes across subgroups ($\Delta \boldsymbol{\mu}$); and, as it was mentioned before, changes in *within* subgroups income inequality is taken as a residual.

Whilst constructing vectors $\hat{\mathbf{v}}_{t+1}$ and $\hat{\boldsymbol{\mu}}_{t+1}$ several methodological issues must be addressed. As it is explained in detail in the following sections, the population partition rule, ξ , includes age and education, two important characteristics affecting income distribution. On the other hand, changes in average incomes by population subgroups will derive from a Computable General Equilibrium (CGE) model. In fact the whole empirical framework is schematically represented in Figure 1. Micro-simulations include the expected changes in the shares of population by groups formed by age and education characteristics (top boxes of Figure 1). The future changes in population shares by age (upper left part of Figure 1) are taken as exogenous from the population projections provided by the World Bank's Development Data Group. Therefore, we assume that fertility decisions and mortality rates are determined outside the model. The change in shares of the population by education groups incorporates the expected demographic changes (linking arrow from top left box to top right box in Figure 1). Next, new sets of population shares by age and education subgroups are computed and household sampling weights are rescaled according to the demographic and educational changes above (larger box in the middle of Figure 1). In a second step, the demographic changes will impact overall labor supply by age and skill groups. These changes are incorporated into the CGE model to simulate overall economic growth, growth in relative incomes by education groups and sector reallocation of labor (link between the middle and bottom rectangles). Finally, the results of the CGE are passed-on to the re-weighted household survey (bottom link in Figure 1).

Figure 1 GIDD methodological framework



The following sections describe how each variable in the boxes of Figure 1 is estimated and how they are linked to each other.

III. Socio-Demographic and Educational Changes

The starting point of our microsimulation exercise is a set of changes in the demographic structure. The relative size of the different age groups is modified following the World Bank's Development Data Group population projections. Additionally, the changes in the demographic structure have an impact on the average educational attainment in the population, i.e. a "pipeline" effect; therefore, educational endowments are modified accordingly. The microsimulation model accounts for these changes by adjusting (or re-calibrating) the household survey data by means of two alternative re-weighting procedures.

Begin with a matrix of *individual* sampling weights $W=[w_{mn}]$, where N is the number of observations in the sample and m is a vector of individual-level characteristics targeted by the microsimulation model. Since in the majority of surveys the household, rather than the individual, is the sampling unit, the individual weight is often, but not always, the

household weight divided by the number of household members.⁵ The sum of all weights in W gives us total population P :⁶

$$P = \sum_{m=1}^M \sum_{n=1}^N w_{m,n} = W \mathbf{i}_n' \quad (3.1)$$

where $\mathbf{i}_n$ and $\mathbf{i}_m$ are identity column vectors. The row sums define the totals of the relevant population sub-groups P_m :

$$P_m = \sum_{n=1}^N w_{m,n} = W \mathbf{i}_n \quad \forall m = 1, \dots, M \quad (3.2)$$

In the current application of the GIDD, these population sub-groups are calculated as intersections of age and education projections, although the methodology can incorporate any number of additional partition rules: by gender, geographic area, ethnicity, etc. The demographic projections between 2000 and a future year are obtained from the World Bank's Development Data Group in 5-year cohorts, ranging from 0 to 100 years of age.⁷ Educational projections are based on the forecasted demographic structure by exploiting the heterogeneity of educational attainments across age groups.

Assume that at time t young individuals are more educated than older ones. As the population ages, the old and unskilled of today will be replaced by the young and more skilled individuals. Therefore at time $t+1$, the overall skill endowments increase as a consequence of the change in the structure of the population—even in the absence of policies intended to increase educational attainments. In other words, this “pipeline” effect maintains a constant distribution of skills within age groups, but leads to gradually rising average educational attainments at the national level. For example, if at time t half of the population in the cohort formed by individuals whose age is between 25 and 30 have post-secondary education, then, after 10 year (at $t+1$), half of the population between 35 and 40 will have post-secondary education.

Combined with the exogenous population forecasts, these *semi-exogenous* projections of skill levels (Figure 1) yield the target (or expected) population in each sub-group m such that:

$$\hat{P}_m = \sum_{n=1}^N a_{m,n} w_{m,n} = (A.W) \mathbf{i}_n \quad \forall m = 1, \dots, M \quad (3.3)$$

⁵ Certain surveys (e.g., Brazil and Venezuela) target certain individual-level characteristics (such as the gender composition of the sample) and therefore adjust the sampling weights at the individual level to be consistent with the census data.

⁶ In most cases, aggregate statistics like census data will differ from the sum of micro sources such as household surveys; a cross-entropy method to reconcile household survey and national accounts data is developed in Robilliard and Robinson (2003).

⁷ The assumptions behind these projections can be found in: <http://esa.un.org/unpp/>

where $A=[a_{mn}]$ is a matrix of multipliers which ensure that the m constraints on the future structure of the population $\hat{P}$ are satisfied and $(A.W)$ is the hadamard product.⁸ This system has $(mxn)-1$ variables but only m constraints and is therefore underdetermined. The two possible solutions are to add equations to make the system exactly identified, or to solve an optimization problem that minimizes the distance between the original matrix W and the final matrix $(A.W)$. Both solutions are available in the GIDD.

The first approach imposes the restriction that the multipliers must be equal for each sub-group m :

$$a_{m,n} = \bar{a}_m \quad \forall m = 1, \dots, M \quad (3.4)$$

This approach reduces the problem to a system of m equations and m unknowns and thus yields an easy solution:

$$\bar{a}_m = \hat{P}_m \left(\sum_{n=1}^N w_{m,n} \right)^{-1} \quad \forall m = 1, \dots, M \quad (3.5)$$

Beyond its simplicity, there is one additional advantage of this method: it maintains the original distribution of personal characteristics within each of the m population sub-groups. In other words, the distribution of personal characteristics in $\hat{P}$ differs from the distribution in P only due to changes in the between-group variance. Therefore, within the m groups, the original survey design remains unaltered.

Despite these advantages, the above method can produce significantly flawed results if the sampling units are sufficiently dispersed across the m sub-groups. For example, if the variable of interest is household per capita consumption and the m sub-groups span across age and skill endowments, relatively few households would fall entirely into one sub-group. For households spanning more than one sub-group, the re-weighting procedure will then assign higher sampling weights to some household members and lower weights to others. This is unsatisfactory for two reasons. First, the intention of any re-weighting procedure is to produce “clones” of observations in the initial dataset. However, the structure of an average household in $\hat{P}$ will differ from the structure of the average household in P . Second, the procedure can also have unintended consequences for the distribution of per capita consumption.

As an example, consider two households: one is composed of two “old” individuals, while the other contains one “old” and one “young” member. With an upward-sloping

⁸ Note that we are not imposing the total population constraint $\hat{P} = \sum_{m=1}^M \sum_{n=1}^N a_{m,n} w_{m,n} = (A.W)\mathbf{i}_n \mathbf{i}_m'$,

which would make the system over-determined in m variables. The underlying assumption is that the sub-group targets $\hat{P}_m$ add up to the total population $\hat{P}$ (either originally or following normalization by the user), which makes one of the equations linearly dependent of the others and allows us to drop it.

age-consumption profile, the per capita consumption of the first household would generally be above those of the second. As the population ages, the first household will become more representative of the overall demographic structure and the average consumption in the population will increase. However, in the procedure described by equation (3.5), the increase in consumption due to higher weight of the first household will be somewhat offset by the rising contribution of the second household which has lower per capita consumption (because both the sampling weights are increased for both households). Therefore, the upward-sloping age-consumption profile observed in the cross-section may not be accurately reflected in the outcome of the re-weighting procedure. In order to address these shortcomings, the GIDD allows for a second alternative for estimating the A matrix.

The procedure works by minimizing a distance function $D(w_{mn}, a_{mn}w_{mn}) = D(a_{mn})$ subject to a set of constraints in equation (3.3). It is therefore similar to the methodology of Robilliard and Robinson (2003) and Cai, Creedy and Kelb (2006). However, it differs from the previous efforts in one crucial aspect by explicitly recognizing the importance of maintaining the household structure of the original survey while respecting the individual-level constraints of equation (3.3). Consider minimizing a simple distance function of the following form:

$$\sum_n 0.5 \left(\frac{\sum_{m=1}^M a_{m,n} w_{m,n}}{\sum_{m=1}^M w_{m,n}} - 1 \right)^2 \quad (3.6)$$

subject to the constraints in equation (3.3) and an additional set of constraints below:

$$r_{m,n} = \frac{w_{m,n}}{\sum_{m=1}^M w_{m,n}} = \frac{a_{m,n} w_{m,n}}{\sum_{m=1}^M a_{m,n} w_{m,n}} \quad (3.7)$$

The solution to this minimization problem is a matrix A that penalizes the squared percentage deviations of $(A.W)$ from W while meeting the set of sub-group constraints $\hat{P}_m$ and keeping the original ratio of individual to household weights unchanged for each household in the sample (equation 3.7). Equation (3.7) implies that:

$$a_{m,n} = \bar{a}_n \quad \forall n = 1, \dots, N \quad (3.8)$$

which allows for a convenient re-statement of the minimization problem by simplifying equation (3.6) and combining equations (3.3) and (3.8):

$$\min \sum_n 0.5(\bar{a}_n - 1)^2 \quad \text{s.t.} \quad \hat{P}_m = \sum_{n=1}^N a_{m,n} w_{m,n} = \sum_{n=1}^N \bar{a}_n w_{m,n} \quad (3.9)$$

The first order conditions are:

$$\bar{a}_n = 1 + \sum_{m=1}^M \lambda_m w_{m,n} \quad (3.10)$$

$$\hat{P}_m = \sum_{n=1}^N \bar{a}_n w_{m,n} \quad (3.11)$$

These can be written in matrix form as follows:

$$\begin{bmatrix} I & -W' \\ W & 0 \end{bmatrix} \begin{bmatrix} A \\ \Lambda \end{bmatrix} = \begin{bmatrix} \mathbf{i}_n \\ \hat{P} \end{bmatrix} \quad (3.12)$$

The solution is:

$$\begin{bmatrix} A \\ \Lambda \end{bmatrix} = \begin{bmatrix} 0 & W'(WW')^{-1} \\ -(WW')^{-1}W & (WW')^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{i}_n \\ \hat{P} \end{bmatrix} \quad (3.13)$$

which gives a simple expression for Λ :

$$\Lambda = (WW')^{-1}(\hat{P} - W\mathbf{i}_n) \quad (3.14)$$

The matrix to invert is mxm , which considerably reduces the dimensionality of the problem. Once the values for Λ are known, the first order condition (3.10) can be used to obtain a solution for the A matrix.

IV. Macroeconomic Changes

The socio-demographic changes captured by the above procedure are likely to have important consequences for economic growth and the distribution of income within a given country. For example, population aging is generally correlated with declining saving rates and changing demand patterns, while rising average skill endowments could reduce the observed skill wage premiums. In an increasingly globalizing world, the direction and magnitude of these changes will also be affected by the changing patterns of international flows of goods, services, and capital. In order to capture all of these effects in a consistent fashion, the GIDD is linked to a global computable general equilibrium (CGE) model to obtain a set of price (factor returns) and quantity (factor volumes) changes for a future time period. Currently, the CGE model used with the GIDD is the World Bank's global LINKAGE model, although the microsimulation methodology is compatible with any CGE model that has sufficient factor market detail.

LINKAGE is a relatively standard CGE model with many neoclassical features (for the full model description, see [van der Mensbrugghe 2006](#)). It is currently based on the Global Trade Analysis Project (GTAP) Release 6.3 dataset with a 2001 base year.⁹ The model is solved in a recursive-dynamic mode in which a series of end-of-period equilibriums are linked with a set of equations that update the main macro variables. The three particularly relevant aspects of LINKAGE (for the purposes of the GIDD) are its multi-sectoral nature and its detailed treatment of factor markets and international trade and capital flows.

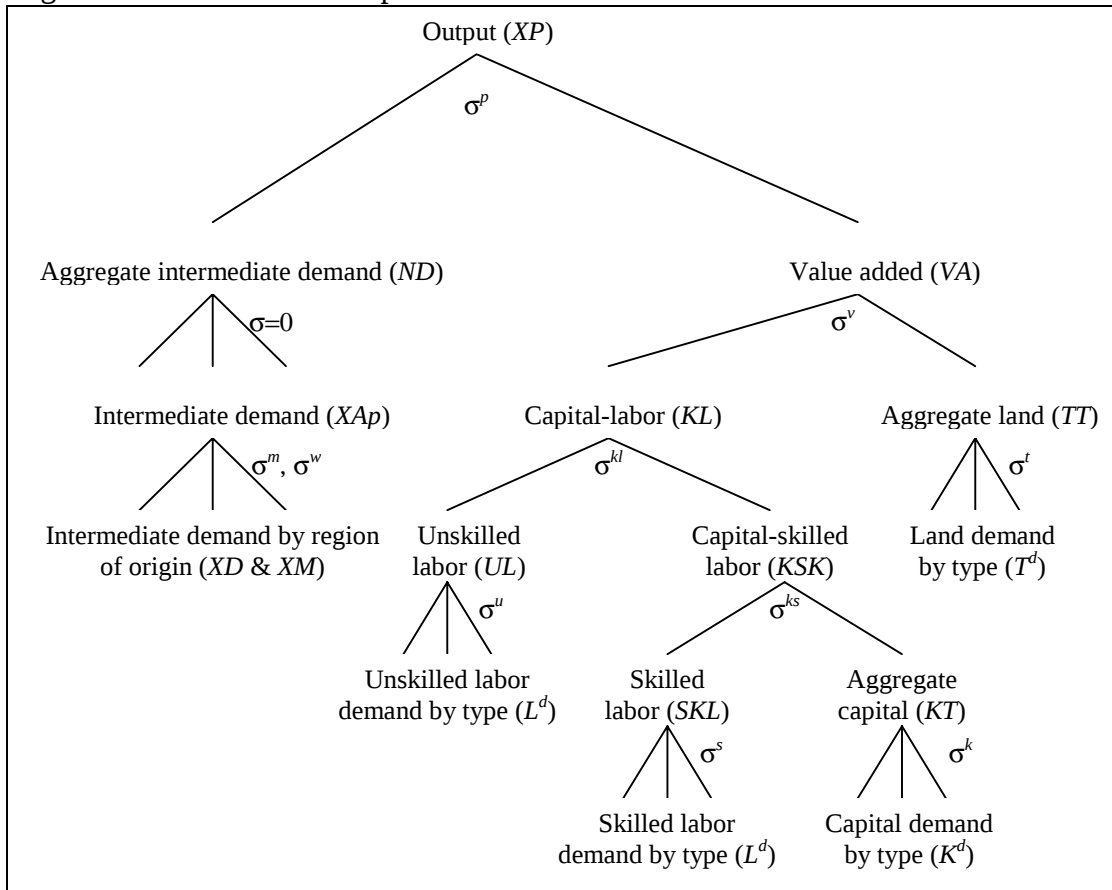
The inclusion of multiple productive activities and multiple commodities allow for a rich production and demand structure. Productivity trends are sector- and factor-specific, and

⁹ The Global Trade Analysis Project (GTAP) database and model are disseminated by Center for Global Trade Analysis of Purdue University. See <http://www.gtap.org> and [Hertel \(1999\)](#).

are calibrated to be consistent with historical evidence as well as World Bank's near- and medium-term GDP growth forecasts. The allocation of household budget (for a single representative household in each country) across saving and a vector of consumption commodities is determined simultaneously through maximization of an extended linear expenditure system (ELES). The system captures various substitution possibilities across commodities as well as a gradual shift in demand towards commodities with higher income elasticities (e.g., manufacturing and services) over time.

Production is modeled in a nested constant elasticity of substitution (CES) fashion to reflect various substitution possibilities across inputs (see Figure 2). This allows for a rich treatment of factor markets, where returns to factors of production—unskilled and skilled labor, capital, land, and natural resources—can be type- and sector-specific. In standard GIDD applications, capital and well as skilled labor are perfectly mobile across sectors within a country, while the market for unskilled labor is segmented into farm- and non-farm categories. Within each segment, labor is perfectly mobile across activities, but mobility across segments is limited by a migration function which responds to changes in the farm/non-farm wage premiums. The LINKAGE model also allows for international mobility of labor and capital as well as changes in the unemployment rate, but none of these possibilities are currently modeled within the GIDD.

Figure 2 Nested structure of production in LINKAGE



Source: van der Mensbrugghe, 2006

International trade is modeled using the nested Armington specification, in which consumer products are differentiated by region of origin and combined using CES functions.¹⁰ On the supply side, producers allocate output to domestic and export markets according to a constant elasticity of transformation (CET) specification. The global nature of the model means that all countries have some degree of market power, goods and services markets clear at the international level, and global capital flows are balanced. The degree of international openness—both trade and capital—affects domestic factor prices directly but also has important consequences for the growth of factor productivity.

V. Labor Reallocation

Changes in the rate of exit of workers from the traditional agricultural sector into manufacturing and services may occur as an outcome of the baseline growth process or as a result of specific policy interventions that affect the wage gap between the two types of activities. Workers will choose to abandon the agricultural sector if this choice represents an increase in their expected earnings. Therefore, any change in the rate of re-allocation of labor across sectors will have an impact on income distribution. At the macro level, the CGE model will predict the number of workers moving out of the traditional agricultural sector into the relatively modern industrial and service sectors. At the micro level, the macro constraint of moving N workers out of agriculture and into manufacturing and service activities can be satisfied by a large number of potential combinations of workers. Some studies (e.g., Bussolo et al 2007) resolve this ambiguity by randomly picking migrants from the agricultural labor supply until the aggregate constraint is satisfied. The GIDD employs a more sophisticated methodology by estimating the conditional probability function of being a worker in the non-agricultural sector, ranking the workers in the agricultural sector according to their probability score, and assigning migrant status to workers with the highest score until N workers have been selected. Currently, this procedure is implemented at the household level—where the head of household makes the migration decision and takes the rest of the household members with her—although the methodology can also be applied at the individual level.¹¹

The probability of observing that individual j works in the non-agricultural sector is modeled with a probit equation:

$$\Pr(NA_j = 1) = P(X_j, Z_j) \quad (5.1)$$

¹⁰ See Armington (1969).

¹¹ The choice for implementing the migration routine at the household level is driven by data constraints. In a large number of GIDD surveys (particularly consumption-based surveys, which make up 54 of the 73 surveys in the GIDD) contributions of individual incomes to total household income cannot be identified, forcing us to operate at the household level.

where X_j and Z_j are vectors of personal and household characteristics of individual j , respectively. Following estimation, workers in the agricultural sector are assigned a probability score based on their X and Z characteristics and the estimated vector of common determinants β_p . The workers are then ordered based on this probability score, and workers with higher probabilities to be in non-agricultural sectors are moved out of the agricultural sector up to a point where the predicted share of workers by sector (the macro constraint) is satisfied.

Once the agricultural workers with a highest likelihood of being in non-agricultural sectors have changed sector of employment, the next step is to adjust their labor remuneration. The first step in this process is estimating a Mincer equation for workers in agricultural (A) and non-agricultural (NA) sectors:

$$\ln(Y)_{j,s} = \mathbf{X}_j \beta_s + \varepsilon_{j,s} \quad s = (A, NA) \quad (5.2)$$

Migrants carry their personal endowments $\mathbf{X}_j$ and their residual ε_j from one sector to the other. Nevertheless once they arrive to the non-agricultural sectors, their vector of personal characteristics $\mathbf{X}_j$ will be rewarded with prices β_{NA} and their residuals will be re-scaled to take into account the differences in the distribution of unobservables between the agricultural and non-agricultural sectors. Hence assuming worker j is a migrant her income assignment function will be defined as:

$$\ln(Y)_{j,NA} = X_j \beta_{NA} + \varepsilon_j^* \quad (5.3)$$

where $\varepsilon_j^* = \varepsilon_{j,A} * \frac{\sigma_{\varepsilon,NA}}{\sigma_{\varepsilon,A}}$ and $\sigma_{\varepsilon,s}$ is the standard deviation of the distribution of residuals in sector s .

VI. Income Assignment

The final step in the GIDD microsimulation is to adjust factor returns by skill and sector, as well as the average income/consumption per capita, in accordance with the results of the CGE model. There are two potential difficulties in translating the price changes of the CGE model into the micro data. First, following the implementation of the re-weighting and migration routines certain changes have already taken place both in the average survey income and its distribution. Therefore, the macro constraints on changing returns to sector and skills $[y_{s,l}]$ as well as the average income $\bar{y}$ are imposed *net* of the changes that have already taken place up to this stage. Second, achieving full consistency between macro and micro data is often difficult if not impossible.¹² Since there is no guarantee

¹² See the discussion in [Bussolo et al 2007](#) for a more detailed statement of this consistency problem and some examples.

that the first period wages in the CGE model match the labor earnings in the micro data, directly passing the changes in factor returns from the former to the latter may result in inconsistent evolution of wage premiums in the two models. In extreme cases, wage gaps may even be reversed in one model but not in the other. In order to hedge against these potential complications while ensuring maximum consistency between the macro and micro outcomes, the GIDD adjusts the ratios between wage premiums rather than wages themselves.

Beginning with a distribution of earnings by sector and skill $[y_{s,l}]$ in the macro data, define a series of $(s+1)$ wage gaps as follows:

$$g_{s,l} = \frac{y_{s,l}}{y_{1,1}} - 1 \quad (6.1)$$

where $y_{1,1}$ is the average labor earnings of unskilled workers in agriculture. The micro data will have a set of wage premiums $[g'_{s,l}]$ which may or may not be consistent with the macro data. The counterfactual wage gaps in the GIDD will then be calculated as:

$$\hat{g}_{s,l} = g'_{s,l} \frac{\hat{g}_{s,l}}{g_{s,l}} \quad (6.2)$$

This implies that even if initial and final wages differ between the macro and micro models, the *percentage* change in the wage gaps (themselves expressed as percentage premiums over labor earnings of unskilled workers in agriculture) will be consistent across the two models. This eliminates the possibility of wage gap reversal and ensures that the distributional changes are consistently mapped from the macro to the micro data.

Note that equation (6.2) does not change the average earnings of unskilled workers in agriculture and only operates on labor income. In order to adjust the micro data such that the percentage change in the per capita income/consumption y' matches the change in real consumption per capita y in the CGE model, a final adjustment is carried out:

$$\hat{y}' = y' \frac{\hat{y}}{y} \quad (5.3)$$

The adjustment of equation (6.3) implicitly accounts for changes in land, natural resource, and capital prices because these enter the household budget constraint in the CGE model and thus have an income effect on consumption. Therefore, the income adjustment process described in equations (6.1) and (6.3) allows the changes in labor remuneration to affect the income distribution of a given country, but the change in welfare at the national level is determined by the changes in all factor prices, including land and capital.

This approach conveniently avoids the issue of identifying sources of household income different from labor, but is justifiable on several grounds. First, it avoids the difficulties involved in estimating the contribution of capital to household earnings.¹³ Second, movements in skilled wage and returns to capital are often correlated, so the GIDD is able to capture the distributional impacts of changing returns to capital through equation (6.2). Third, the empirical literature on decomposing changes in the income distribution over time (e.g., Ferreira et al 2004) is usually able to explain much of the change in total inequality without resorting to estimation of capital incomes.

VI. Applications

The GIDD framework has been used in various studies. To provide some examples of what kind of findings can be generated from a GIDD-based analysis, this section briefly summarizes the results from three recent applications. The first considers what will happen to global income distribution in the next couple of decades; the second application highlights the role of China and India as engines of global growth and redistribution; and the third addresses the changes of global income distribution and global poverty due to the elimination of all the distortions to international trade of agricultural commodities.

In the first application the GIDD in conjunction with a global computable general equilibrium model is used to generate a new income distribution for the year 2030 (see Bussolo, De Hoyos and Medvedev, 2008b). This study then identifies the drivers of the expected distributional changes by means of two complementary approaches. The analysis is initially conducted in terms of the convergence and dispersion components, i.e. changes in income disparities between and within countries. Results show that the reduction in global income inequality between 2000 and 2030 is the outcome of two opposing forces: the inequality-reducing convergence effect and the inequality-enhancing dispersion effect (Table 1).

Table 1: Global Income Inequality

Index			Dispersion	Convergence
	2000	2030	Only	Only
Gini	0.672	0.626	0.673	0.625
Theil	0.905	0.749	0.904	0.749
Mean Log Deviation	0.884	0.764	0.893	0.759

Data source: Authors' own calculations using data from GIDD

Three main findings have emerged: first, even with significant changes of within-country inequality levels, all the potential reduction of global inequality can be accounted for by the projected convergence in growth rates of average incomes across countries. Second,

¹³ Most econometric solutions to the problem of imputing capital earnings ignore the selection bias in the self-employment decision. Furthermore, it is questionable whether it is possible even in principle to extract information on capital income from surveys that are generally not designed to capture this information and where definitions of “capital” may vary widely between micro data and national accounts.

the aggregate impact of the changes of the within-countries component of inequality appears to be minor; however specific countries, and specific households' types within countries, may experience large distributional shifts. Third, a main cause of local inequality changes is the adjustments of factor rewards.

To translate these results into a more practical and policy relevant perspective, this study considers what happens to a specific income group during the 2000-2030 time period. The group under consideration is labeled "global middle class" (GMC) and comprises people whose income levels are between the average incomes of Brazil and Italy, in purchasing power parity terms.¹⁴ The combination of the convergence and divergence components described earlier drive a dramatic increase in the size of the global middle class and its profound compositional change in favor of developing country nationals. A key conclusion asserts that developing country members of the global middle class are likely to become an increasingly important group within their own countries, will increase their political influence and possibly provide continued momentum for policies favoring global integration.

The second GIDD application mentioned above considers the role of China and India in shaping the future evolution of the global income distribution and in particular these countries' contribution to the emergence of the global middle class (see Bussolo et al. 2007). According to our baseline, in 2030, 16.1 percent of the world population will belong to what can be called a global middle class, up from 7.6 percent in 2000. That is, in 2030 more than a billion people in developing countries will buy cars, engage in international tourism, demand world-class products, and require international standards for higher education. Compare that with only 250 million people in developing countries who had access to these kinds of living standards in 2000. By assigning an individual to the global middle class according to his or her income, Table 2 shows the evolution of this income group and contrasts it with the groups of the poor and the rich.¹⁵ This table also shows that the great majority of the global middle class entrants are citizens of developing countries; hence tomorrow's global middle class will be formed, primarily, by today's citizens from poor countries. The total increase in the global middle class is explained by (1) population growth rates of cohorts within this class that are above the world average, and (2) by higher economic growth rates in developing countries which pull their citizens out of *poverty* and into the global middle class. The population growth rates of households within the global middle class (as classified in 2000) was relatively low with an average rate of 18 percent over the entire period, as opposed to the world average of 32 percent. Therefore, the great majority of the increase in the global middle class is explained by high economic growth rates taking place in developing countries.

¹⁴ In 1993 PPP prices, the lower threshold is 303 dollars per person per month, while the upper threshold is 611 dollars per person per month. This means that per capita earnings of members of the global middle class are 10 to 20 times above the international poverty line of 1 dollar a day. These income thresholds are due to the global middle class definition proposed by Milanovic and Yitzhaki (2002).

¹⁵ Notice that the definition of poor used here is far from being comparable to the standard 1 dollar-a-day definition.

Table 2 The global middle class is growing, and its composition is changing

	Shares				Growth Rates (% 2000-2030)	
	2000		2030			
	Pop.	Income	Pop.	Income	Pop.	Income
Poor	82.0	28.7	63.0	17.0	2	29
Middle class, of which:	7.6	13.8	16.1	14.0	178	0
<i>Developed country nationals</i>	3.5	6.8	1.2	1.0	-52	-2
<i>Developing country nationals</i>	4.1	7.0	14.6	12.9	363	3
Rich	10.5	57.5	20.9	69.0	163	28
Total	100	100	100	100	32	109

Notes: (1) totals may not sum to 100 because of rounding.

(2) Poor are defined as individuals with an income below the average of Brazil; the middle class was defined as individuals with an income between the per capita incomes of Brazil and Italy; rich are those individuals with incomes at or above the average income in Italy.

(3) Thresholds of Brazil and Italy are annual per capita incomes (2000 PPP) of US\$3,914 and US\$16,746.

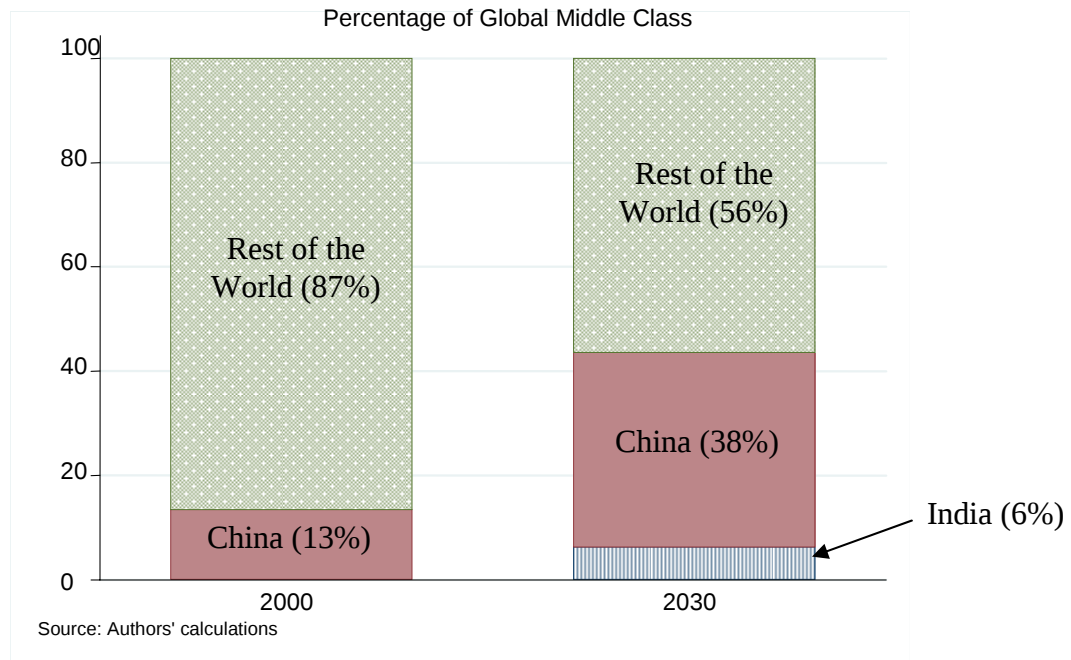
Source: Authors' calculations.

How much of the expected increase in the global middle class is attributable to the economic performance of China and India? Figure 3 divides the global middle class into citizens from China, India and the rest of the World (RoW). In 2000 only 13.5 percent of the global middle class were Chinese nationals and no Indians belonged to this group.¹⁶ By 2030 citizens from China and India had a combined share of 44 percent of the global middle class, with the great majority (38 percent) being Chinese, in fact half of the total 740 million new entrants into the global middle class will be Chinese nationals.

The importance of China and India in the global middle class will depend on their economic and population growth rates and the changes in their within-country income inequality. For instance, in China, 56 million people belonged to the global middle class in 2000—each of them earning more than 90 percent of all Chinese citizens, i.e. they belonged to the richest decile. By 2030, assuming income inequality in China remains constant, there will be 361 million Chinese in the global middle class, and their earnings will range from the sixth to the ninth decile of the Chinese national income distribution. Chinese members of the global middle class will no longer be among the richest Chinese citizens but will probably be considered upper middle class in their country. On the other hand, if China manages to reduce income disparities, making middle income cohorts fatter, they would contribute even further to the global middle class.

¹⁶ It is quite likely that in reality some Indians are within the middle and high income ranges, nevertheless by the way the Indian Household Survey data is being collected, *outliers* (high income citizens) are not captured at all.

Figure 3 Chinese and Indian weight in the global middle class



These results highlight the fact that aggregate indicators of inequality of the global distribution of income depend heavily on changes between countries and much less on changes within countries. From a global inequality perspective, this is certainly true in a situation where very populous and initially poor countries (China and India) are growing at a rate above that of rich countries.

In the third application the GIDD is used to study the income distribution and poverty consequences of liberalizing agricultural trade (see Bussolo, De Hoyos and Medvedev, 2008a). The main results can be summarized as follows. With the elimination of all agriculture trade distortion, poverty – measured at the 1 dollar a day international poverty line – is reduced in all regions but in the Middle East and North Africa, where it is almost stable, and in South Asia where it increases considerably. This result needs to be qualified though: given that about 50 percent of all poor people live in South Asia the worsening of poverty in this region counterbalances all the gains in the other parts of the world and close to an additional 9 millions people fall into poverty. The situation is rather different at the higher poverty line of 2 dollars a day. In this case 14 million people escape poverty and most regions benefit from lower poverty incidence with the exclusion of Eastern Europe and Central Asia and Middle East and North Africa. Many non-agriculture households in South Asia are clustered below the 2 dollar a day poverty line and trade reform-related improvement in their incomes, versus the agricultural incomes' decline, explain the difference in global poverty results when the 1 dollar or the 2 dollar a day lines are used.

Although per capita income gaps between the agriculture-dependent households and the other households are marginally reduced by the removal of trade barriers, global inequality does not change. This is because the lower between agriculture and non-agriculture groups inequality is offset by a higher within group inequality which mainly originates by a widening of incomes within the agriculture sector.

Within country inequality varies within a wide interval ranging from increases of up to 3 Gini points to reduction of 2 Gini points. The majority of countries, around 60% of those included in the sample, experiences an increase of inequality. Most of the distributional change is due to changes of the average incomes between the agriculture and non-agriculture group of households, a direct consequence of the trade shock.

In summary, allocative efficiency gains combined with distributional shifts originating from the removal of agriculture trade restrictions are not enough to significantly alleviate poverty at the 1 dollar a day threshold nor at a higher poverty line. The pattern of global incomes change triggered by such trade reform, as simulated by the model used in this paper, is complex and cannot be simplistically reduced to being equivalent to a boost in growth rates of agriculture. The latter remains an essential component in the strategy for poverty eradication and trade liberalization can only play a constructive but somewhat limited role.

Conclusions

In an increasingly globalize world, many domestic policies have an impact that goes beyond the country's own frontiers, similarly, several economic policy proposals have a global nature, e.g. trade liberalization agendas, policies mitigating climate changes, etc. The GIDD is the first global CGE-Microsimulation model and it therefore closes an important gap in the policy evaluation empirical literature. The paper shows a new methodology to evaluate the global welfare effects of global and domestic macroeconomic policies. The GIDD is able to incorporate, in an ex-ante fashion, changes in demographic composition, sectoral re-allocation of labor, shifts in relative wages and overall growth and it thus represents an important step towards a more integrated global Macro-Micro evaluation framework. The paper develops the methodology in detail and then illustrates its usefulness by showing three recent applications if the GIDD: (a) how the global income distribution of income will look like in 2030, (b) how changes in the global growth, income distribution, and global middle class will be determined by China and India, and (c) an evaluation of the global poverty and distributional consequences of eliminating all present agricultural distortions. Although the GIDD represents an important contribution to our understanding of the global welfare effects of macro policies, more research is needed to update the GIDD's data and expand its modeling capabilities.

References (to be completed)

- Bourguignon, François and Luiz Pereira da Silva (ed.) (2003), “The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools”, World Bank and Oxford University Press.
- Bussolo, Maurizio, Jann Lay and Dominique van der Mensbrugghe. (2006) “Structural change and poverty reduction in Brazil : the impact of the Doha Round”. *World Bank Policy Research Working Paper* 3833.
- Bussolo, M., De Hoyos, R., Medvedev, D., and van der Mensbrugghe, D. (2007) “Global Growth and Distribution: Are China and India Reshaping the World?”, *World Bank Policy Research Working Paper*, 4392, Washington, DC
- Bussolo, M., De Hoyos, R. and Medvedev, D. (2008a) “Global Income Distribution and Poverty in the Absence of Agricultural Distortions”, in Anderson, K. (editor) “Effects of Distortions to Agricultural Prices on Growth, Income Distribution and Poverty”, *forthcoming*
- Bussolo, M., De Hoyos, R. and Medvedev, D. (2008b) “The Future of Global Income Inequality”, in Estache, A and Leipziger, D. (editors) “The Global Middle Class”, *forthcoming*
- Chen, Shaohua, and Martin Ravallion. (2003) “Household Welfare Impacts of China’s Accession to the World Trade Organization.” *Policy Research Working Paper* 3040, World Bank, Washington, DC.
- Ferreira, Francisco H. G., and Phillippe G. Leite. (2003) “Meeting the Millennium Development Goals in Brazil: Can Microsimulations Help?” *Economía* 3 (2): 235–79.
- . (2004) “Educational Expansion and Income Distribution: A Microsimulation for Ceará.” In *Growth, Inequality and Poverty*, ed. Anthony Shorrocks and Rolph van der Hoeven. London: Oxford University Press
- van der Mensbrugghe, Dominique (2006) “The LINKAGE Model Technical Documentation.” Washington, DC: World Bank.