



Tulane Economics Working Paper Series

Commitment to Equity Assessment (CEQ): Estimating the Incidence of Social Spending, Subsidies and Taxes Handbook

Nora Lustig

Sean Higgins

Working Paper 1219

October 2012

Abstract

This handbook presents a step-by-step guide to applying the incidence analysis used in the multi-country project CEQ. We define the pre- and post-net transfers income concepts, discuss the methodological assumptions used to construct them, explain how taxes, subsidies and transfers should be allocated at the household level, and suggest what to do when the information on taxes and transfers is not included in the household survey. We also describe the indicators that are used to assess the distributive impact, progressivity and effectiveness of social spending, subsidies and taxes. In addition, we present sample Stata code for producing some of the indicators. Keywords: handbook, taxes and transfers,

fiscal incidence, poverty, inequality

JEL: Keywords: H22, D31, D63, I32, I38



Tulane Economics Working Paper
CIPR Working Documents

Commitment to Equity Assessment (CEQ):
Estimating the Incidence of Social Spending, Subsidies and Taxes¹
Handbook

Nora Lustig and Sean Higgins²
Department of Economics
Tulane University
New Orleans, LA
nlustig@tulane.edu; shiggins@tulane.edu

October 2012³

Abstract

This handbook presents a step-by-step guide to applying the incidence analysis used in the multi-country project CEQ. We define the pre- and post-net transfers income concepts, discuss the methodological assumptions used to construct them, explain how taxes, subsidies and transfers should be allocated at the household level, and suggest what to do when the information on taxes

¹ Led by Nora Lustig (Tulane University) and Peter Hakim (Inter-American Dialogue), the *Commitment to Equity* (CEQ) project is designed to analyze the impact of taxes and social spending on inequality and poverty, and to provide a roadmap for governments, multilateral institutions, and nongovernmental organizations in their efforts to build more equitable societies. The CEQ uses incidence analysis and a specially designed diagnostic questionnaire to address three questions: How much redistribution and poverty reduction is being accomplished in each country through social spending, subsidies and taxes? How progressive are revenue collection and government spending? Within the limits of fiscal prudence, what could be done to increase redistribution and poverty reduction in each country through changes in taxation and spending? CEQ is the first framework to comprehensively assess the tax and benefits system in developing countries and to make the assessment comparable across countries and over time. Initially, CEQ has focused on Latin America. CEQ/Latin America is a joint project of the Inter-American Dialogue (IAD) and Tulane University's Center for Inter-American Policy and Research (CIPR) and Department of Economics. The project has received financial support from the Canadian International Development Agency (CIDA), the Development Bank of Latin America (CAF), the General Electric Foundation, the Inter-American Development Bank (IADB), the International Fund for Agricultural Development (IFAD), the Norwegian Ministry of Foreign Affairs, the United Nations Development Programme's Regional Bureau for Latin America and the Caribbean (UNDP/RBLAC), and the World Bank. <http://commitmenttoequity.org>

² Nora Lustig is Samuel Z. Stone Professor of Latin American Economics and Sean Higgins is a doctoral student in the Department of Economics, Tulane University. The authors are very grateful to Samantha Greenspun and Emily Travis for their excellent assistance in the preparation of this document. Earlier versions of this handbook received very valuable comments from Jim Alm, Nancy Birdsall, Ludovico Feoli, Francisco Ferreira, Ariel Fiszbein, Peter Hakim, Miguel Jaramillo, Luis F. Lopez-Calva, Mario Marcel, Santiago Levy, Tamara Ortega Goodspeed, Carola Pessino, Jeffrey Puryear, David Roodman, Jaime Saavedra, John Scott and other members of CEQ's Advisory Board.

³ An earlier version of this handbook was published as Tulane University Economics Department Working Paper No. 1119 in April 2011. This version was most recently revised in September 2012.

and transfers is not included in the household survey. We also describe the indicators that are used to assess the distributive impact, progressivity and effectiveness of social spending, subsidies and taxes. In addition, we present sample Stata code for producing some of the indicators.

Keywords: handbook, taxes and transfers, fiscal incidence, poverty, inequality

JEL: H22, D31, D63, I32, I38

Table of Contents

1. Introduction

2. Income Concepts and Methodological Assumptions

- i. Income Concepts: Definitions**
- ii. Constructing the Income Concepts: Methodological Assumptions**
- iii. When Information on Taxes and Transfers is not in the Survey**
- iv. Benchmark Case and Sensitivity Analyses**
- v. Constructing Market Income**
- vi. Constructing Net Market Income, Disposable Income, Post-fiscal Income and Final Income**

3. Definitions of Progressivity

4. Incidence Results and Indicators

- i. Sheet 1 – Reduction in Inequality and Poverty**
- ii. Sheet 2 – Effectiveness Indicators**
- iii. Sheet 3 – Measures of Progressivity and Horizontal and Vertical Inequality**
- iv. Sheet 4 – Incidence by Decile and Socioeconomic Groups**
- v. Sheet 5 – Concentration Shares by Decile and Socioeconomic Groups**
- vi. Sheet 6 – Income Distribution by Decile and Socioeconomic Groups**
- vii. Sheet 7 – Fiscal Incidence Curves and Fiscal Mobility Profiles by Deciles**
- viii. Sheet 8 – Concentration Coefficients and Budget Shares for Social Spending and by Program**
- ix. Sheet 9 – Coverage and Leakages by Program**
- x. Sheet 10 – Fiscal Mobility Matrices**
- xi. Sheet 11 – Probit of the Before and After Transfers Poor**
- xii. Sheet 12 – Needs vs. Resources**
- xiii. Sheet 13 – Cumulative Distribution Functions of Income**
- xiv. Sheet 14 – Lorenz Curves**
- xv. Sheet 15 – Inequality of Opportunity**

xvi. Sheet 16 – Progressiveness of Pensions

xvii. Sheet 17 – Comparison with Other Studies

References

1. Introduction

The Commitment to Equity Assessment (CEQ) uses standard incidence analysis⁴ to address the following three questions: How much redistribution and poverty reduction does a country accomplish through social spending and taxes? How progressive are revenue collection and government spending? What could be done to further increase redistribution and improve redistributional effectiveness? CEQ is among the first efforts to comprehensively assess the tax/benefit system in developing countries (including indirect subsidies and taxes and in-kind benefits in the form of free education and health care) and to make the assessment comparable across countries and over time. Applications of CEQ can be found in, for example, Bucheli et al. (2012) and Lustig et al. (2012).

The purpose of this handbook is to present a step-by-step guide to applying the incidence analysis used in CEQ and completing the Master Workbook Template, a spreadsheet file that contains all the information used and produced by CEQ.⁵ The handbook is organized as follows. Section 2 explains how to construct the income concepts used in the incidence analysis. Section 3 presents the definitions of progressivity. Section 4 describes the indicators that are used to assess the distributive impact, progressivity and effectiveness of social spending, subsidies and taxes. In addition, it explains how to complete the accompanying Workbook, and includes sample Stata code for producing some of the sheets.

2. Income Concepts and Methodological Assumptions

i. Income Concepts: Definitions

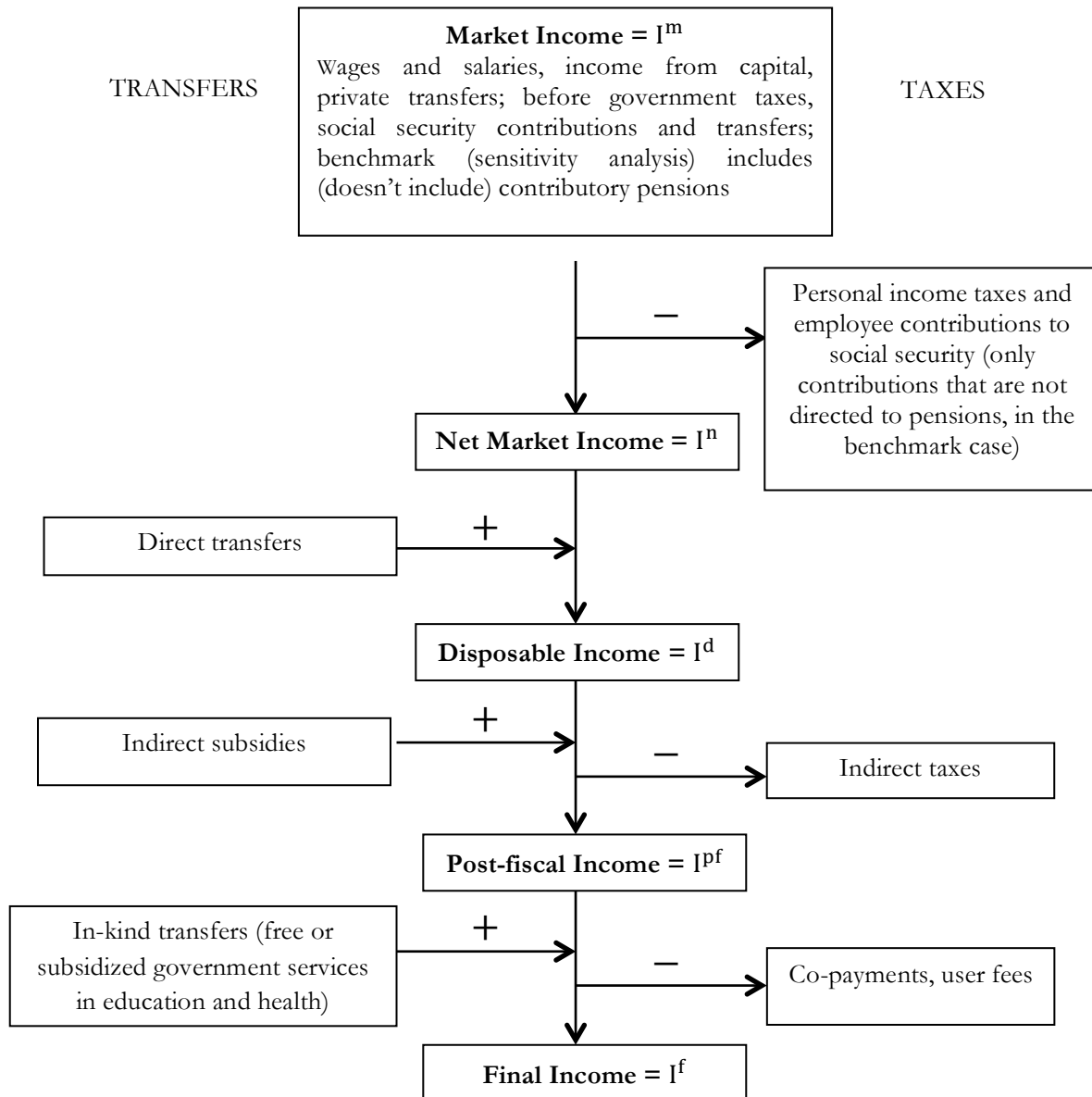
As usual, any incidence study must start by defining the basic income concepts. In our study we use five: market, net market, disposable, post-fiscal and final income. The categories included in each concept are shown in Diagram 1 and described in more detail below. One area in which there is no agreement is how pensions from a pay-as-you-go contributory system should be treated. Arguments exist in favor of both treating contributory pensions as part of market income because they are deferred income (Breceda et al., 2008; Immervoll et al., 2009) or as a government transfer especially in systems with a large subsidized component (Goñi et al., 2011; Immervoll et al., 2009; Lindert et al., 2006; Silveira et al., 2011). Since this is an unresolved issue, in our study we defined a benchmark

⁴ For a description, applications and limitations of standard incidence analysis see, for example, Adema and Ladaique (2005), Alleyne et al. (2004), Atkinson (1983), Bergh (2005), Bourguignon and Pereira da Silva (2003), Barr (2004), Barros et al. (2009), Birdsall et al. (2008), Breceda et al. (2008), Dilnot et al. (1990), Ferreira and Robalino (2010), Fiszbein et al. (2009), Grosh et al. (2008), Goñi et al. (2011), Kakwani (1977), Lambert (2002), Lora (2006), Morra et al. (2009), Lustig (2000), O'Donnell et al. (2008), Shah (2003), Suits (1977), van de Walle and Nead (1995), World Bank (2000/2001, 2006, 2009b, 2011).

⁵ The Master Workbook Template is under the proprietorship of CEQ.

case in which contributory pensions are part of market income. We also include a sensitivity analysis in which pensions are classified under government transfers.⁶

Diagram 1 – Definitions of Income Concepts: A Stylized Presentation



Note: in some cases we also present results for “final income*” which is defined as disposable income plus in-kind transfers minus co-payments and user fees.

More detailed definitions of the income concepts are as follows.

⁶ Immervoll et al. (2009) do the analysis under these two scenarios as well.

Market income is defined as:

$$I^m = W + IC + AC + IROH + PTran + SSP \text{ (benchmark)}$$

$$I^{ms} = W + IC + AC + IROH + PTran \text{ (sensitivity analysis)}$$

Where,

I^m, I^{ms} = market income⁷ in benchmark and sensitivity analysis, respectively.

W = gross (pre-tax) wages and salaries in formal and informal sector; also known as earned income.

IC = income from capital (dividends, interest, profits, rents, etc.) in formal and informal sector; excludes capital gains and gifts.

AC = autoconsumption; also known as self-production.

IROH = imputed rent for owner occupied housing; also known as income from owner occupied housing.

PTran = private transfers (remittances and other private transfers such as alimony).

SSP = retirement pensions from contributory social security system.

Net Market income is defined as:

$$I^n = I^m - DT - SSC \text{ (benchmark)}$$

$$I^{ns} = I^{ms} - DT - SSC^s \text{ (sensitivity analysis)}$$

Where,

I^n, I^{ns} = net market income in benchmark and sensitivity analysis, respectively.

DT = direct taxes on all income sources (included in market income) that are subject to taxation.

SSC, SSC^s = respectively, all contributions to social security except portion going towards pensions⁸ and all contributions to social security without exceptions.

Disposable income is defined as:

$$I^d = I^n + GT \text{ (benchmark)}$$

$$I^{ds} = I^{ns} + GT + SSP \text{ (sensitivity analysis)}$$

Where,

I^d, I^{ds} = disposable income in benchmark and sensitivity analysis, respectively.

GT = direct government transfers; mainly cash but can include transfers in kind such as food.

SSP = retirement pensions from contributory social security system.

⁷ Market income is sometimes called primary income.

⁸ Since here we are treating contributory pensions as part of market income, the portion of the contributions to social security going towards pensions are treated as 'saving.'

Post-fiscal income is defined as:

$$I^{pf} = I^d + \text{IndS} - \text{IndT} \text{ (benchmark)}$$

$$I^{pfs} = I^{ds} + \text{IndS} - \text{IndT} \text{ (sensitivity analysis)}$$

Where,

I^{pf} , I^{pfs} = post-fiscal income in benchmark and sensitivity analysis, respectively.

IndS = indirect subsidies (e.g., lower electricity rates for small-scale consumers).

IndT = indirect taxes (e.g., value added tax or VAT, sales tax, etc.).

Final income is defined as:

$$I^f = I^{pf} + \text{InkindT} - \text{CoPaym} \text{ (benchmark)}$$

$$I^{fs} = I^{pfs} + \text{InkindT} - \text{CoPaym} \text{ (sensitivity)}$$

Where,

I^f , I^{fs} = final income in benchmark and sensitivity analysis, respectively.

InkindT = government transfers in the form of free or subsidized services in education and health; urban and housing.

CoPaym = co-payments, user fees, etc., for government services in education and health.⁹

Because some countries do not have data on indirect subsidies and taxes, we also defined *Final income** = $I^{fe} = I^d + \text{InkindT} - \text{CoPaym}$.

ii. Constructing the Income Concepts: Methodological Assumptions

To construct the income concepts using the above definitions, one must have access to micro-data from a recent household survey with data on income and, ideally, consumption. The information from this data set will be combined with data on taxes and the transfer programs from public sector accounts. When constructing the income definitions, we make the following methodological assumptions.

Definition of Household

We adopt the definition of a household used by SEDLAC, which excludes external members of the household: boarders (*inquilinos* in Spanish and *pensionistas* in Portuguese), and domestic servants and their families are not considered part of the household, and must be dropped from the data set.

NOTE: This definition of household is used to calculate household income. It is important to note, however, that the poverty and inequality calculations will be in terms of individuals (for example, the incidence of poverty will equal the proportion of individuals whose income is below the poverty line), unless otherwise specified.

⁹ One may also include participation costs, such as transportation costs or foregone incomes because of use of time in obtaining benefits. In our study, they were not included.

Adult Equivalence and Economies of Scale

CEQ uses household per capita income, and thus does not adjust for adult equivalence or economies of scale within households. For each income concept, total household income for the respective concept is divided by the total number of members in the household.

Missing or Zero Incomes

When a survey respondent reports receiving a certain income source but does not report the value or reports a value of zero as their income from that source, we adopt the convention used by SEDLAC: Missing and zero incomes are regarded as zero, unless the household head's primary income source is missing or zero, in which case the household is excluded from the data (CEDLAS and World Bank, 2012).

Income Underreporting (Scaling Up)

It is well-known that household income surveys tend to understate “true” income. This has several possible causes: people might underreport their own incomes (on purpose or by accident), surveys might fail to ask adequate questions to capture certain categories of income or might have too long of a recall period, and society's richest members are usually not captured by household surveys (especially under conditions of high inequality when a large share of national income is concentrated on a small fraction of the population). For this reason, some studies scale up household survey income to match a comparable definition of income in national accounts before estimating poverty. However, Deaton (2005) argues that the methodologies of computing income in national accounts should not be used when estimating poverty because they are upward-biased and not designed to generate poverty statistics. Thus, we do not scale up income by national accounts when estimating poverty indicators. In the Master Workbook Template these include the poverty measures (Incidence Results and Indicators Sheet 1), coverage and leakages (Incidence Results and Indicators Sheet 9), mobility matrices (Incidence Results and Indicators 10), and probit (i.e., to determine whether someone is poor, which determines the value of the dependent dummy variable) (Incidence Results and Indicators Sheet 11).

However, when calculating inequality indicators or the incidence of public transfers over the whole distribution, failing to adjust for income underreporting would necessarily overestimate the redistributive effect of in-kind transfers, as the monetary value of the transfers received by households is obtained from the budgetary cost of providing these transfers as reported in the public component of national accounts. In countries where direct taxes are imputed to households by applying the prevailing tax law (adjusted for tax evasion when the survey allows identification of informal employment) rather than directly reported in the household survey questionnaire, failing to adjust for income underreporting would also overestimate the redistributive effect of direct taxes. Thus, a second “scaled up” vector of income variables should be generated for each household, which scales up reported market income to national accounts. This scaling up is done by identifying the closest equivalent definition of income in national accounts, then aggregating the total population's market income according to the household survey. The ratio of aggregate income in

national accounts is used to aggregate income in the survey as a multiplier for each household's market income in the survey. From this scaled up market income, direct taxes and employee contributions to social security are subtracted to arrive at net market income. If taxes are reported directly in the household survey, they should also be multiplied by the multiplier; if they are imputed to households, no multiplier adjustment should be made. When direct monetary transfers are added to arrive at disposable income, the same criteria is applied: if the transfer is reported on the survey, it is multiplied by the multiplier, but if it is imputed to households based on national accounts totals, it is not adjusted. The same criteria is applied to indirect subsidies, indirect taxes, in-kind transfers, in-kind taxes, co-payments, and user fees as they are added and subtracted from disposable income to arrive at the subsequent definitions of income: anything reported on the survey should be adjusted using the multiplier, while anything imputed to households based on national accounts should not be adjusted. The scaled up vector of income definitions should be used for all inequality and distribution-related indicators, such as the Gini coefficients, Theil index, 90/10, income distribution by deciles, Lorenz curves, concentration curves and concentration coefficients, incidence of transfers and taxes, anonymous and non-anonymous fiscal incidence curves, Kakwani index of progressivity, Reynolds-Smolensky index, redistributive effect, and impact on inequality of specific programs.

In sum, the original, non-scaled-up, vector of income definitions should be used for all poverty estimations, while a second, scaled up, vector of income definitions should be used for all estimations related to income distribution, including estimations of inequality and progressivity.

Behavioral, Indirect, and Spillover Effects

In general (for now), CEQ does not account for behavioral, indirect, or spillover effects.

Intertemporal Effects, General Equilibrium Effects, Marginal vs. Average Incidence

In general (for now), CEQ is a static incidence analysis that does not account for intertemporal effects or general equilibrium effects.

iii. When Information on Taxes and Transfers is not in the Survey

Unfortunately the information on direct and indirect taxes, transfers in cash and in-kind, and subsidies cannot always be obtained directly from household surveys. Thus, one of the most important aspects of CEQ is a detailed description of how each component of income is calculated (for example, directly drawn from the survey or simulated) and the methodological assumptions that are made while calculating them. When taxes and transfers can be obtained directly from the household survey, we call this the *Direct Identification Method*. When the direct method is not feasible, one can use the inference, simulation, imputation or alternate survey methods (described in more detail below). As a last resort, one can use secondary sources: e.g., incidence or concentration shares by quintiles or deciles that have been calculated by other authors as is done by Goñi et al. (2011) for

instance. Finally, if none of these options can be used for a specific category, the analysis for that category will have to be left blank.

Direct Identification Method

On some surveys, questions specifically ask if households received benefits from (paid taxes to) certain social programs (tax and social security systems), and how much they received (paid). When this is the case, it is easy to identify transfer recipients and taxpayers, and add or remove the value of the transfers and taxes from their income, depending on the definition of income being used.

Inference Method

Not all surveys have the information necessary to use the direct identification method. In some cases, transfers from social programs are grouped with other income sources (in a category for “other income,” for example). In this case, it might be possible to infer which families received a transfer based on whether the value they report in that income category matches a possible value of the transfer in question.

Simulation Method

In the case that neither the direct identification nor the inference method can be used, transfer benefits (taxes) can sometimes be simulated, determining beneficiaries (taxpayers) and benefits received (taxes paid) based on the program (tax) rules. For example, in the case of a conditional cash transfer that uses a proxy means test to identify eligible beneficiaries, one can replicate the proxy means test using survey data, identify eligible families, and simulate the program’s impact. However, this method gives an upper bound, as it assumes perfect targeting and no errors of inclusion or exclusion. In the case of taxes, estimates usually make assumptions about informality and evasion.

Imputation Method

The imputation method is a mix between the direct identification and simulation methods; it uses some information from the survey, such as the respondent reporting attending public school or receiving a direct transfer in a survey that does not ask for the amount received, and some information from either public accounts, such as per capita public expenditure on education by level, or from the program rules.

The four methods described above rely on at least some information directly from the household survey being used for the analysis. As a result, some households receive benefits, while others do not, which is an accurate reflection of reality. However, in some cases the household survey analyzed lacks the necessary questions to assign benefits to households. In this case, there are two additional methods.

Alternate Survey

When the survey lacks the necessary questions, such as a question on the use of health services or health insurance coverage (necessary to impute the value of in-kind health benefits to households), an alternate survey may be used by the author to determine the distribution of benefits. In the alternate survey, any of the four methods above can be used to identify beneficiaries and assign benefits. Then, the distribution of benefits according to the alternate survey is used to impute benefits to all households in the primary survey analyzed; the size of each household’s benefits depends on the quantile to which the household belongs. Note that this method is more accurate than the secondary sources method below, because although the alternate survey is somewhat of a

“secondary source,” the precise definitions of income and benefits used in CEQ can be applied to the alternate survey.

Secondary Sources Method

When none of the above methods are possible, secondary sources that provide the distribution of benefits (taxes) by quantile may be used. These benefits (taxes) are then imputed to all households in the survey being analyzed; the size of each household’s benefits (taxes) depends on the quantile to which the household belongs.

NOTE: It is very important to specify which identification method is used for each transfer program, tax, etc. This information should be explicitly mentioned in the accompanying Master Workbook Template.

iv. Benchmark Case and Sensitivity Analyses

As mentioned above, there is no agreement on whether to consider contributory social security pensions as part of market income or as a government transfer. Hence, we opted for doing it both ways to check how sensitive results to the treatment of contributory pensions are. We define a benchmark case in which contributory pensions are part of market income. We also do a sensitivity analysis where pensions are classified under government transfers. A second sensitivity analysis is country-specific (i.e., some countries may want to check the implications of classifying other transfers as market income, etc.), and a third uses non-scaled income for the inequality and incidence indicators that used scaled income in the benchmark case (this has to be done up to post-fiscal income since for final income estimates scaling-up is unavoidable). Details are as follows.

Benchmark Case

As mentioned under the definitions of income, all pensions except pensions received from the non-contributory system should be included in market income. Pensions received from the non-contributory system (sometimes called “minimum pensions”) are social assistance, thus they are not included in market income in the benchmark case; they are treated as a government transfer.

Including all pensions except those from the non-contributory system as part of market income is a simplification. In countries with a “pay-as-you-go” pension system, employee and employer contributions into the social security system can be smaller than the amount paid out by the system, which results in a social security deficit that is financed by the government. In this case, a portion of pensions should technically be considered a subsidy; however, there is no way to identify from the household surveys whose pensions are coming from the subsidized portion of the social security system, and whose pensions are coming from the contributory pool. As a result, all pensions except those from the non-contributory system will be considered part of market income in the benchmark case.

Since we are considering pensions a form of intertemporal savings, we do not subtract contributions to the social security system that are directed to pensions from income when moving from market to net market income in the benchmark case. (In the case where income reported on the survey is net of contributions directed to pensions, the latter must be imputed and added into income.) We do subtract out direct taxes and contributions that are not directed to pensions.

Sensitivity Analysis 1

The main sensitivity analysis is to treat social security pensions as a government transfer. Thus they are not included in market or net market income. Contributions to social security directed to pensions are subtracted out of income when moving from net market income to disposable income in the sensitivity analysis. As a result, benchmark case disposable income and sensitivity analysis disposable income are slightly different, even though in both cases contributory pensions have been added in by this point: the difference is that in the benchmark case contributions directed to pensions are never subtracted out of income so they are included in benchmark case disposable income.

Sensitivity Analysis 2

This is reserved for country-specific sensitivity analyses. For example, in Brazil there are special circumstances pensions which are considered a transfer in the benchmark case, but which are part of the contributory system and thus Sensitivity Analysis 2 considers them part of market income, along with regular contributory pensions. As another example, one might wish to simulate a new or proposed reform to a transfer program in Sensitivity Analysis 2.

Sensitivity Analysis 3

For indicators that use scaled income (see above), Sensitivity Analysis 3 calculates the same indicator using non-scaled income.

v. Constructing Market Income

Market income begins with gross (pre-tax) wages and salaries from the formal and informal sectors (also known as earned income) and income from capital (rents, profits, dividends, interest, and so on). It also includes private transfers (remittances and other private transfers such as alimony), imputed rent for owner occupied housing; also known as income from owner occupied housing and autoconsumption (also known as self-production). In the benchmark case, market income also includes retirement pensions from the contributory social security system. Most of these components can be directly extracted from the household survey data; the potentially more involved methodological details for imputed rent for owner occupied housing and autoconsumption are discussed below. Also, since many surveys report post-tax wages, pre-tax wages must be constructed by simulating personal income taxes based on the prevailing tax code.

Imputed Rent for Owner Occupied Housing

There are multiple methodologies to impute the value of owner-occupied housing. In some countries, survey questionnaires ask families who own their homes to report the amount they think they would be paying in rent for the same dwelling, or for how much they would rent it out. In the case where there is no such question, or if the authors feel that survey respondents do not have sufficient information about housing markets to answer this question accurately, the regression methodology described below can be used instead.

A second methodology uses a regression to impute the value of owner-occupied housing. The regression methodology requires a survey question to those who are renting their homes, asking how much they pay per month in rent. For the subset of households that rent, (the log of) their monthly rent is the dependent variable in the regression. Potential independent variables include any characteristics about the dwelling, as well as income per capita of the household. See Higgins (2011) for more detail. After exploring a number of potential independent variables, Higgins (2011) ends up using the following variables for the case of Brazil: number of bedrooms, number of bathrooms, log household income per capita, rural dummy, state dummies, interaction terms between state dummies and the rural dummy, sets of dummies for whether the dwelling is a house, apartment, or room in a shared building, the material of the walls, type of sewage, presence of piped water, floor material, roofing material, and an intercept. Alternatively, Jiménez, Paz and Yáñez (2012) perform three separate regressions for houses, apartments, and other housing types, using similar dependent variables as those described from Higgins (2011). After obtaining a vector of coefficients for the dependent variables from the regression for those who own their home, this vector of coefficients is applied to those variables for owner-occupiers (i.e., the sub-matrix X containing an intercept and those variables as columns, is post-multiplied by the vector of coefficients β from the regression). This generates the predicted value of imputed rent for owner occupiers, \hat{y} . If you use the regression methodology, describe your variable selection process (see Higgins, 2011 for an example) and provide the regression results in the Master Workbook Template on the sheet titled Valuation of Imputed Rent for Owner's Occupied Housing and Autoconsumption, in the Methodological Aspects and Assumptions section.

Note that the first method, where the response to a survey question about the value of owner-occupied housing is used, requires such a question in the survey at hand, and the second method requires that families who rent their dwellings are asked to report how much they pay in rent. If neither piece of information is available, we resort to the methodology used by SEDLAC for countries in this scenario, which only requires a question as to whether households rent their homes. By this methodology, the incomes of families who own their own homes is increased by 10%, which according to SEDLAC is a value that is “consistent with estimates of implicit rents in the region” (CEDLAS and World Bank, 2012, p.18).

Autoconsumption

The method used to determine the value of autoconsumption depends on the survey data available. Surveys with consumption data often ask whether that item was produced or purchased. The value of items that were produced by the household, taken from the household's own business inventory, or donated to the household (by someone other than the government) are included in market income as autoconsumption. Other surveys simply ask the value of production for own consumption; in that case this value is added to market income.

vi. Constructing Net Market Income, Disposable Income, Post-fiscal Income and Final Income

Constructing Net Market Income: Start from Market Income

Subtract Direct Taxes

Direct taxes are personal income taxes. Corporate taxes and other forms of direct taxes that are not captured by the household survey are not included in this analysis. When personal income taxes are not reported in the survey, they should be simulated based on the prevailing tax code and tax evasion assumptions. When the tax incidence is obtained by the simulation method, the latter should be described with detail, including the evasion assumptions. As a last resort, the incidence of taxes could be obtained from other studies on tax incidence for the same country.

Subtract Contributions to Social Security

As discussed in the above section, contributions to social security are treated as follows:

Benchmark case: Because we are considering pensions a form of intertemporal savings, we do not subtract contributions to the social security system that are directed to pensions from income when moving from market to net market income in the benchmark case. (In the case where income reported on the survey is net of contributions directed to pensions, the latter must be imputed and added into market income.) We do subtract out contributions that are not directed to pensions.

Sensitivity Analysis 1: Contributions to social security directed to pensions are subtracted out of income when moving from net market income to disposable income in the sensitivity analysis. As a result, benchmark case disposable income and sensitivity analysis disposable income are slightly different, even though in both case contributory pensions have been added in by this point: the difference is that in the benchmark case contributions directed to pensions are never subtracted out of income so they are included in benchmark case disposable income.

Constructing Disposable Income

Add Direct Government Transfers

Direct government transfers includes, but is not limited to, conditional cash transfer programs, non-contributory pensions, scholarships, and other direct transfers (which may or may not be targeted to the poor). Food transfers, although not cash, are considered a direct transfer because they have a well-defined market value and are close substitutes for cash. Unemployment benefits and other benefits that might be part of the contributory system but are intended to deal with idiosyncratic shocks are also counted as direct transfers (and should therefore not be included in social security pensions which are part of market income in the benchmark case).

Add Contributory Pensions in Sensitivity Analysis 1 ONLY

In Sensitivity Analysis 1, contributory pensions were not a component of market income and are considered a government transfer, and are thus added into income when moving from net market to disposable income.

Constructing Post-fiscal Income

Subtract Indirect Taxes

Ideally, indirect taxes should be imputed using consumption data. If the survey being used contains both income and consumption data, tax rates for the prevailing indirect taxes (such as consumption taxes in the form of a value-added tax) are applied to each household's reported consumption of the corresponding items. Informality and evasion should be taken into account. For example, many surveys ask where each consumption item was purchased, or at least the main location where a household purchased its goods. In the former case, goods purchased from informal markets, such as ambulant vendors or flea markets, are assumed to have been purchased on the informal market, and thus no indirect tax was paid. In the latter case, if a household identifies usually purchasing their goods in an informal market, it is assumed that no indirect tax was paid. An exception will be made when taxes are applied to the inputs of production rather than the final good, and when it is not feasible to evade these taxes on the inputs to production.

When the above is not possible due to a lack of consumption data, secondary sources may be used. For example, a secondary source might provide the incidence by decile of indirect taxes with respect to market income. This incidence by decile is then applied to the disposable income of each individual in the corresponding decile from the CEQ analysis to obtain her spending on indirect taxes. The implicit assumption being made when one uses indirect taxes by decile is that everyone in that decile pays the same proportion of their income (equal to the average over the decile) in indirect taxes.

When scaling up indirect taxes, the total from national accounts used to perform the scaling up should include only the taxes that were accounted for in the author's imputation or in the secondary sources.

Add Indirect Subsidies

Indirect subsidies must often be imputed using data on consumption of subsidized items. Indirect subsidies are analogous to indirect taxes and the same method applies.

Constructing Final Income

Add In-kind Transfers

The value of in-kind transfers is based on the use of public services as reported in the survey. Details, by category of in-kind transfer, are given below.

Education

From national accounts, obtain public spending per student by level (pre-school, primary [lower and upper if applicable], secondary [lower and upper if applicable], tertiary [university and technical if applicable]). Provide the definition of each level (i.e., the corresponding grade levels and age groups). For students who report attending public school, depending on the level they report attending, use the average public spending per student for that level as the valuation of their in-kind benefit from public education, which is added into income when moving from post-fiscal to final income, or from disposable to final income*. In addition to having a variable for in-kind education benefits, the researcher should create separate variables for benefits at each level (i.e., a variable for pre-school education benefits, another for primary education benefits, etc.) for the more disaggregated analysis that is required on some sheets of the Master Workbook Template.

*Health*¹⁰

To impute the value received from public health services, the household survey must have information about the use of health services, and distinguish between public care (which is usually services received from the public health system or paid for by public health insurance schemes) and private care. In the absence of information about whether the care received was subsidized by government health spending, a survey question about whether the patient is covered by private insurance can be used as a proxy; i.e., patients who received health care and report having private health insurance are considered to have received private care, and thus received no in-kind transfer, and patients who report *not* having private health insurance are considered to have received public care. Ideally, the survey will also contain one or more questions about the type of service received.

If this information is not available in the survey being used, another survey that has information on both income and utilization of public health services – such as a health survey – should be used. In this case, to calculate final income one must then treat the results from the alternate survey similarly to a secondary source and impute values by quantiles (e.g., ventiles [groups of 5% of the population]) back into the original micro-data.

¹⁰ This section is largely based on O'Donnell et al. (2008), Chapter 14.

However, for the concentration coefficient of health spending (Incidence Results and Indicators Sheet 8 of the Master Workbook Template) or its coverage and leakages (Incidence Results and Indicators Sheet 9 of the Master Workbook Template), one should calculate these directly in the alternate micro-data, without imputing these results back into the original data set.

In addition to data on the use of public health services and the type of services received, data on total government spending on each of the different types of health services in the household survey is required. Some level of disaggregation by type of service received (at a minimum, distinguishing between in-patient and out-patient care) is required, in order to account for the fact that the value of a medical check-up is different from the value of a hospitalization. This data should also be disaggregated by region or state when possible to account for differences in the quality of health services across regions. Data that is disaggregated as described above is generally not available in the main source of public accounts (e.g., from the treasury or ministry of development), but can be obtained instead from national health accounts (e.g., from the health ministry).

In the event that the care received is partially but not fully subsidized, the amount paid for care by the individual or by private health care providers should be subtracted from the total benefit received by that individual. If public health care in the country being studied is, in general, not fully subsidized (for example, there is not a universal free health care system) but the household survey does not ask how much each individual paid for the service they received or how much was not covered by the public health insurance scheme, each individual's payment can be calculated as the average payment for that service; i.e., it is calculated as the total payment from individuals and private health insurers to the state for that service (available in national health accounts) divided by the total number of individuals receiving that service according to the household survey.

The total annualized health benefits received by an individual are thus defined as

$$h_i = \sum_k \alpha_k \left[q_{ki} \left(\frac{S_{kj}}{\sum_{i \in j} \omega_i \alpha_k q_{ki}} \right) - f_{ki} \right]$$

where q_{ki} indicates the number of times that individual i received care type k during the recall period, S_{kj} is the total spending (according to national health accounts) on service k in the region j where i resides, $i \in j$ indicates that we are summing over all individuals in region j , ω_i is the sampling weight corresponding to observation i , and α_k is the “annualization factor”: for services that have a recall period of one year on the questionnaire (e.g., “How many times in the last year did you receive service k ?”), $\alpha_k = 1$; for services that have a recall period of four weeks, $\alpha_k = 13$, etc.

Finally, f_{ki} is the user fee paid by individual i for service k . In the case of a health system with no user fees, we normally use $f_{ki} = 0$ (regardless of whether the system is fully or partially subsidized, because the level of subsidization would already be captured by the term $S_{kj}/\sum_{i \in j} \omega_i \alpha_k q_{ki}$) unless other costs such as waiting times are being incorporated in the analysis. When user fees exist, if the survey asks individuals how much they paid for that particular service or has information (sometimes found along with other consumption questions) about how much they paid in health costs, f_{ki} can be determined from the survey. Note that f_{ki} could still equal zero for some i , for example for poor individuals if there are fee exemptions for the poor. In the absence of such survey information, one can determine the average health user fee per visit, \bar{f} , as $\bar{f} = \frac{N_k}{\sum_k \sum_i \omega_i \alpha_k q_{ki}}$, where N_k is total user fee revenue, reported in public accounts or national health accounts. In other words, \bar{f} is total user fee revenue divided by the total number of times all individuals in the country utilized any type of public health service. To complete the calculation of total annualized health benefits received by an individual, one would then replace f_{ki} in the above equation with \bar{f} .

In countries with a contributory public health insurance scheme, we are also interested in knowing the concentration of coverage, so the concentration coefficients and coverage and leakages sheets of the Master Workbook Template (Incidence Results and Indicators Sheets 8 and 9, respectively) include a row for Contributory Public Health Insurance in addition to the row for Health Spending. The latter is based on use, using the total annualized health benefits, h_i , calculated as explained above. The former is calculated using a variable equal to zero for individuals not covered by the contributory public health insurance schemes and equal to the value of a basic health package for covered individuals.

Housing

Impute the in-kind value received by those who live in publicly (fully or partially) subsidized housing. Ideally, the survey will include information on who lives in subsidized housing and, if it is only partially subsidized, how much they paid in rent. The market value of their subsidized housing can be determined using a regression methodology (similar to the regression methodology described to impute the value of owner-occupied housing under the section *Imputed Rent for Owner Occupied Housing*). If housing is only partially subsidized, the amount they pay in rent should be subtracted from this total. For the observations for which this method results in a negative value, it should be replaced by zero; however, if a negative value results for many observations, this could be an indication that the linear model used to predict housing values is not a good fit and should be revisited.

Complementary Analysis: Tax Expenditure and Subsidized Portion of Social Security

Tax Expenditures

Tax expenditures result in people paying less indirect taxes, so they should not be added to income (because that would be double-counting). Nevertheless, many of the output tables include tax expenditures; for example, in the incidence table there is a column to the right where the incidence of tax expenditures should be estimated, while the concentration coefficients table has a row for the concentration coefficient and budget size (i.e., forgone revenue) of tax expenditures.

Subsidized Portion of Social Security

Although contributory pensions are not split up into a subsidized and non-subsidized portion in the main analysis (they are considered part of market income, in their entirety, in the benchmark case, and as a government transfer, in their entirety, in the sensitivity analysis), we do separate them into a subsidized and non-subsidized component for a complementary analysis. We propose two methods to impute how the subsidized portion of social security is distributed among households (calculations should be provided using both methods): a) divide the total social security deficit, defined as total social security payments minus total contributions to the social security system, by the number of people who receive pensions and assign the per capita value to each individual who received a pension; b) assign the subsidy in proportion to the pensions each household receives (i.e., the subsidized portion of pensions are distributed identically to total pensions; if forty percent of the social security system is subsidized, then for each individual who received a contributory pension, it is assumed that forty percent of that pension is subsidized). These two methods will give a lower and upper bound for the incidence of the subsidized portion of pensions.

3. Definitions of Progressivity

Since one of the criteria for evaluating the distributive impact of fiscal policy depends on the extent of progressivity of taxes and transfers, this is a good place to review the definitions used in the literature of what constitutes progressive taxes and transfers. To determine if a tax or transfer is progressive, concentration curves, concentration coefficients, and the Kakwani (1977) index are commonly used.

Concentration curves are constructed similarly to Lorenz curves but the difference is that the vertical axis measures the proportion of the tax (transfer) under analysis paid (received) by each quantile. Therefore, concentration curves (for a transfer targeted to the poor, for example) can be above the diagonal (something that, by definition, could never happen with a Lorenz curve). Concentration coefficients are calculated in the same manner as is the Gini; for cases in which the concentration coefficient is above the diagonal, the difference between the triangle of perfect equality and the area under the curve is negative, which cannot occur with the Gini for the income distribution by definition. The data used to generate concentration curves and coefficients are derived from incidence analyses.

In the literature there is no agreement on when to label a transfer as ‘progressive.’ The progressivity/regressivity of a transfer can be measured in absolute terms, by comparing the per capita transfers between quantiles, or in relative terms, by comparing transfers as a percentage of the income of each quantile. Some authors label a transfer as progressive only when the per capita

amount decreases with income (e.g., Scott, 2011). Others, when the proportion received (as a percentage of market income) decreases with income (e.g., Lindert et al., 2006; O'Donnell et al., 2008). Here, we have opted for the latter definition. This is consistent with an intuitively appealing principle: a transfer or tax is defined as progressive (regressive) if it results in a less (more) unequal distribution than that of market income.

In particular:¹¹

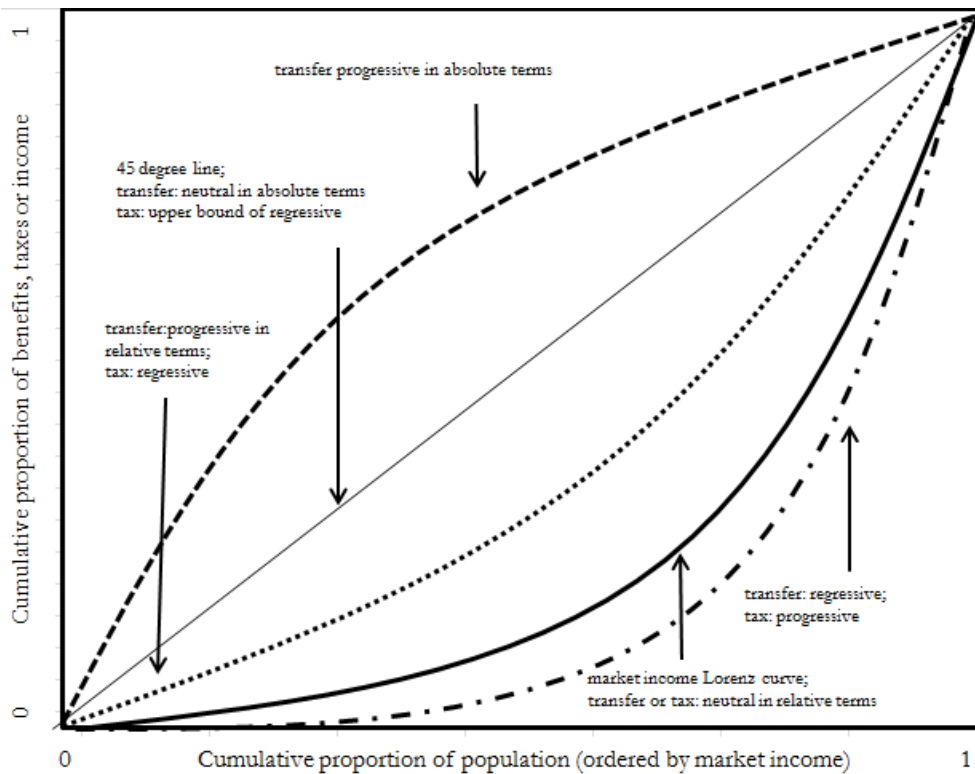
1. A tax is progressive (regressive) if the proportion paid – in relation to market income – increases (decreases) as income rises. The concentration coefficient is positive and larger (smaller) than the market income Gini. The Kakwani index, defined for taxes as the tax concentration coefficient minus the market income Gini, will be positive (negative) if a tax is progressive (regressive).
2. A transfer is progressive if the proportion received – in relation to market income – decreases as income rises. There are two types of progressive transfers: absolute and relative. A transfer will be progressive in absolute terms if the per capita amount received increases as income rises. The concentration coefficient will be negative. The Kakwani index, defined for transfers as the market income Gini minus the transfer's concentration coefficient, will be positive and higher than the market income Gini.¹² A transfer will be progressive in relative terms if the proportion received in relation to market income decreases as income rises but not so the per capita transfer. The concentration coefficient will be positive and lower than the market income Gini. The Kakwani index will be positive if a transfer is progressive in relative terms.
3. A transfer is regressive if the proportion received -- in relation to market income -- increases as income rises. The concentration coefficient will be positive and greater than the market income Gini. The Kakwani index will be negative.
4. A tax or transfer will be neutral (in relative terms) if the distribution of the tax or the transfer coincides with the distribution of market income. The concentration coefficient will be equal to the market income Gini. The Kakwani index will equal zero if a tax or transfer is neutral.

The four cases are illustrated graphically in Diagram Y.

Diagram 2. Concentration Curves for Progressive and Regressive Transfers and Taxes

¹¹ Our taxonomy of absolute and relative progressivity is similar to that used by Lindert et al. (2006). The taxonomy in O'Donnell et al. (2008) is slightly different: they refer to relative progressivity as “progressivity, or weak progressivity” and to absolute progressivity as “absolute or strong progressivity” (p. 185).

¹² The index originally proposed by Kakwani (1977) only measures the progressivity of taxes. It is defined as the tax's concentration coefficient minus the market income Gini. To adapt to the measurement of transfers, Lambert (1985) suggests that in the case of transfers it should be defined as market income Gini minus the concentration coefficient (i.e., the negative of the definition for taxes) to make the index positive whenever the change is progressive.



4. Incidence Results and Indicators

This section describes the informational requirements and indicators as well as the presentational format of CEQ in the Master Workbook Template. In addition to the information contained in this document and in the Master Workbook Template itself, authors can use the sample Stata code included below under each corresponding item as a guideline.¹³

The Master Workbook Template contains four main sections, each with various subsections (sheets). The first section is background information. This section both asks for background information about the country (such as the evolution of inequality and poverty over time and public accounts information) as well as descriptions of the various components of social spending, such as the tax system, flagship cash transfer programs, the health system, etc., in the country being analyzed. The second section asks for survey information, such as the exact survey questions used to construct each component of the income concepts. The third section, methodological aspects and assumptions, both defines many of CEQ's methodological assumptions (described in more detail in

¹³ In the future, we hope to develop computational software and tools (in the form of Stata programs) to systemize the process of producing CEQ outputs once the country authors have prepared the data following the second section of the handbook. The computational software would facilitate the process of carrying out a consistent and comparable analysis using our methodological framework and would generate publication-ready tables and graphs. The software would also facilitate sensitivity analyses and robustness checks which would test the sensitivity of the results to alternative taxonomical, imputational, and behavioral assumptions.

this document) and asks country authors to describe the specifics of applying that methodology to their country, or of additional assumptions they made. The fourth section contains the results and indicators of the incidence analysis. Here, we focus on the incidence results and indicators section. We begin by overviewing some additional definitions that are necessary to complete this section of the Template. We then go sheet-by-sheet through this section, describing the contents of each sheet and any additional definitions needed to complete the sheet. When appropriate, we include sample Stata code to produce the data that will be put in the sheet's tables.

General Definitions

Here we define some general concepts that are used throughout the Template.

Total CEQ Social Spending

In both the benchmark case and the sensitivity analyses, Total CEQ Social Spending includes government spending at all levels on health, education, and social assistance. In the benchmark case it does not include spending on social security pensions, but does include spending on programs such as unemployment benefits which may be part of the contributory system but are intended to smooth idiosyncratic shocks. In the Sensitivity Analysis 1 (the case where contributory pensions are considered a transfer), we define Total CEQ Social Spending plus Contributory Pensions. These numbers are presented both in absolute terms in local currency and as a percentage of GDP. The values for each component should be taken from public sector accounts for the same year as the year of the household survey being used. A breakdown of Total CEQ Social Spending, which shows the value of each of its components, must be provided on the “Macroeconomic and Public Accounts” sheet in the Background Information section of the Template (see Table 1 below for an example from Brazil).

CEQ Social Spending in Incidence Analysis

This number, which is by definition less than (or potentially equal to) Total CEQ Social Spending, only includes the components of Total CEQ Social Spending that are included in the analysis. In other words, if a particular component of social assistance is not captured by the survey and cannot be imputed, simulated, or otherwise incorporated into the analysis, then spending on that component is *not* included in CEQ Social Spending in Incidence Analysis. Table 1 below presents an example for Brazil, in which the rows highlighted in gray are not included in the analysis and thus spending on those items is not part of CEQ Social Spending in Incidence Analysis. For the case where pensions are considered a transfer (i.e., Sensitivity Analysis 1) we also define CEQ Social Spending plus Contributory Pensions Incidence Analysis.

Table 1. Breakdown of CEQ Social Spending in Brazil, 2009

Program	Total Spending (millions of reais)	Notes and Source

Bolsa Família (CCT)	12,454.70	Total paid in transfers, Ministry of Social Development (MDS)
Child Labor Eradication	282.82	MDS
Bolsa Escola (CCT)	0.36	Total paid in transfers (essentially defunct in 2009), MDS
Benefício de Prestação Continuada (Non-contributory pension)	16,859.41	Total paid in transfers, MDS
Other elements of Basic Social Protection	2,397.55	Total from Relatório de Avaliação do Plano Plurianual, MDS, minus BPC transfers
Minimum Income Programs	128.45	Total for Renda Cidadã only; budget for other local programs not available
Auxílio Gás (cash transfer intended to pay for gas)	22.82	Total paid in transfers, MDS
Other auxílios	419.93	Not available in national accounts; this is the total according to the household survey
Abono PIS/PASEP	7,280.24	Ministry of Labor
Unemployment	18,599.90	Ministry of Labor
Professional Training Grant for Laid Off Workers	58.26	Ministry of Labor
Food for Workers Program	489.99	Ministry of Labor
Special Circumstances Pensions	72,564.11	Designed to cope with idiosyncratic shocks (does not include old-age or disability pensions), National Institute of Social Security
Other Scholarships	3,540.46	
Basic Food Basket	42.93	MDS
Food Access	583.66	Net of Basic Food Basket, MDS
Other Social Assistance	44,136.26	Includes Assistência ao Idoso, ao Portador de Deficiência, à Criança e Adolescente, Comunitária, and Demais Subfunções
Early Childhood Education (Educação Infantil)	9,566.32	Treasury
Primary Education (Ensino Fundamental)	75,081.87	Treasury
Secondary Education (Ensino Médio)	12,034.16	Treasury
Tertiary Education	26,006.03	Treasury

Other Education Programs for Children and Adults	1,340.36	Treasury
Special Education	1,007.96	Treasury
Education: Other Spending	44,153.78	Treasury
Health Spending	130,622.74	Treasury
Health: Other Spending	35,389.47	Treasury
CEQ Social Spending in Incidence Analysis	385,507.26	12.1% of GDP
Total CEQ Social Spending	515,064.55	16.2% of GDP
Contributory Pensions (INSS)	164,825.71	Total benefits from National Institute of Social Security net of special circumstances pensions
Contributory Pensions (other federal public)	53,689.43	Total previdência (federal) net of contributory pensions (INSS) and special circumstances pensions, Treasury
Contributory Pensions (state)	56,112.78	Total previdência (estados), Treasury
Contributory Pensions (local)	13,975.03	Total previdência (municípios), Treasury
CEQ Social Spending plus Contributory Pensions in Incidence Analysis	674,110.21	21.2% of GDP
Total CEQ Social Spending plus Contributory Pensions	803,667.50	25.2% of GDP ; this is similar to Social Spending according to CEPAL of 27% of GDP

Source: Pereira and Higgins (2012).

Notes: Dark gray rows indicate elements that were not included in the analysis; thus they are components of Total CEQ Social Spending but they are not components of CEQ Social Spending in Incidence Analysis.

Deciles

Each decile represents ten percent of the population. Individuals are ordered by income from poorest to richest, with the “first decile” referring to the poorest decile, and the “tenth decile” referring to the richest. Note that the division should be done such that the expanded population in each decile is equal (or approximately equal), rather than the number of raw observations in each decile. The expanded population refers to the number of individuals when the appropriate sampling weights are applied to each observation. Individuals in the same household should be kept in the same decile, whereas individuals in different households with the same income may be arbitrarily allocated to different deciles if they are near the cut-off, in order to keep decile sizes approximately equal. This is not possible with

Stata's built-in command `xtile`, and is best accomplished with the user-written command `quantiles` (to install, type `ssc install quantiles` in the command box of Stata). Let household per capita market income be saved as `ym`, the variable containing the identifying code for each household be called `hh_code`, and the variable containing the sampling weight be called `s_weight`. Then, the following command will create market income deciles following the instructions above, and create a new categorical variable called `ym_decile` containing the decile of each observation (i.e., the new variable will be an integer ranging from 1 to 10):

```
quantiles ym [iw=s_weight], gen(ym_decile) n(10) k(hh_code)
```

Some output tables are non-anonymous, in other words, they follow identified individuals, so for example the first decile always refers to the poorest ten percent of the population by initial income (i.e., market or net market income, as specified). For example, in the incidence table we are looking at the change in incomes caused by various taxes and transfers to the incomes of identified individuals: we want to know by how much the incomes of those who are initially in the poorest ten percent, etc. changed. On the other hand, other tables are anonymous so we allow re-ranking between income concepts. For example, in the income distributions table we are comparing the market incomes of the poorest ten percent of the population ranked by market income to the disposable incomes of the poorest ten percent of the population ranked by disposable income, even though these may not be the same individuals.

Since deciles must be created for each income definition, as well as for the benchmark case and each sensitivity analysis, it is wise to use loops in combination with the `quantiles` command. If `ym_BC` represents benchmark case market income, `ym_SA1` represents Sensitivity Analysis 1 market income, etc., the code would look like:

```
foreach x in BC SA1 SA2 SA3 {
  foreach i in m n d pf fstar f {
    quantiles y`i'_'x' [iw=s_weight], gen(y`i'_'x'_decile) n(10) k(hh_code)
  }
}
```

In the code for non-anonymous sheets such as the incidence analysis sheet, only the decile variables `ym*_decile` would be used (where `*` is a wildcard marker; `ym*_decile` indicates all variables in the data set beginning with `ym_` and ending with `_decile`) – in other words, only the market income deciles are used (or net market where specified). On the other hand, in the anonymous sheets such as the income distributions sheet, all the variables `y*_decile` would be used – in other words, income deciles change with each income concept.

Poverty Lines

All poverty lines are absolute and income-based. We use the following poverty lines: the standard international poverty lines of \$1.25 PPP per person per day (which we call "ultra-poverty"), \$2.50 PPP per person per day (extreme poverty), \$4 PPP per person per day (moderate poverty), the official poverty line, which preferably distinguishes between urban and rural areas and possibly by regions, and, for countries in Latin America, CEPAL's poverty lines. The latter are available from ECLAC (2010, Table A5).

To convert the international poverty lines in purchasing-power parity (PPP) adjusted US dollars into local currency poverty lines, the PPP conversion rate should be selected for the same year as the survey. The PPP conversion rate should be based on private consumption rather than GDP; if obtained from the World Development Indicators Databank (<http://databank.worldbank.org>), the series "PPP conversion factor, private consumption (LCU per international \$)" should be used. The yearly international poverty line in local currency is equal to the PPP per day poverty line times the PPP conversion factor (of local currency units per PPP dollar), times 365 days per year. For example, in the case of Brazil, the private consumption-based PPP conversion factor for 2009 (the same year as the household survey being used for Brazil) is 1.71 Brazilian reais = \$1 PPP, so the \$4 PPP per day international poverty line would be converted into local currency (reais) per year as follows:

$$\frac{\$4 \text{ PPP}}{1 \text{ day}} \times \frac{1.71 \text{ reais}}{\$1 \text{ PPP}} \times \frac{365 \text{ days}}{1 \text{ year}} = \frac{2502 \text{ reais}}{1 \text{ year}}$$

Thus, the \$4 PPP per day international poverty line is equivalent to 2,502 reais per year.

Note that if the survey is taken over many months, the data should first be deflated to a specific month. This is often facilitated by temporal deflating factors included as one of the data set's variables.

As mentioned in the previous section, non-scaled incomes are compared to poverty lines to determine if an individual is poor, following the recommendation of Deaton (2005).

Income Groups

We define a set of income groups, beginning with the three poor groups defined above: the ultra-poor (household per capita income less than \$1.25 PPP per day), the extreme poor (household per capita income greater than or equal to \$1.25 PPP per day and less than \$2.50 PPP per day), the moderate poor (household per capita income greater than or equal to \$2.50 PPP per day but less than \$4 PPP per day). The non-poor income groups are the vulnerable (household per capita income greater than or equal to \$4 PPP per day and less than \$10 PPP per day), the middle class (household per capita income greater than or equal to \$10 PPP per day but less \$50 PPP per day), and the rich (household per capita income greater than \$50 PPP per day). These income groups were formulated with middle income countries, particularly those in Latin America, in mind.

The \$1.25 PPP per day line approximately represents the average national poverty line of the bottom fifteen low-income, less-developed countries (Chen and Ravallion, 2010); thus in the context of middle-income countries we call those living on less than \$1.25 PPP per day the

“ultra-poor”. The \$2.50 and \$4 PPP per day poverty lines are commonly used as extreme and moderate poverty lines for Latin America, and roughly correspond to the median official extreme and moderate poverty lines in those countries (CEDLAS and World Bank, 2012). The \$10 PPP per day line is the upper bound of those vulnerable to falling into poverty in three Latin American countries, calculated by Lopez-Calva and Ortiz-Juarez (2011) and the lower bound of the middle class used by Kharas (2010) and Ferreira et al. (2012, forthcoming). The \$50 PPP per day line is the upper bound of the middle class proposed by Ferreira et al. (2012, forthcoming).

Non-scaled incomes are compared to these income lines to determine an individual’s income group, again based on the recommendation by Deaton (2005).

The following Stata code provides an example for converting this set of international lines in \$ PPP per day into local currency per year, saving these lines as local macros, and creating income groups by comparing market income to these lines. The example once again comes from Brazil, where the 2009 consumption-based PPP conversion factor was 1.71 Brazilian reais = \$1 PPP (in practice, more decimal places should be used for increased accuracy, as below).

A loop is used to loop over the benchmark case and sensitivity analyses. Note that comments on a single line are preceded by // or, if they are on their own line, by *. Comments on multiple lines are begun with /* and ended with */. The /// at the end of some lines tells Stata that the next line is a continuation of the code from the previous line; note that /// can only be used in do files, and not in the command prompt.

```
local PPP=1.7116 /* PPP conversion factor for Brazil 2009 (consumption-based)
                  from databank.worldbank.org */
local PPPyr=`PPP'*365 // Brazil data is annual; if monthly also divide by 12
local lines 125 250 400 1000 5000 // $1.25 etc. but won't work with decimals
foreach n in `lines' {
    local PL`n'=(`n'/100)*`PPPyr'
}
label define groups 1 "y<1.25" 2 "1.25<y<2.5" 3 "2.5<y<4" 4 "4<y<10" ///
                    5 "10<y<50" 6 "y>50"
foreach x in BC SA1 SA2 SA3 {
    generate ym_`x'_group=.
    replace ym_`x'_group=1 if ym_`x'<`PL125'
    replace ym_`x'_group=2 if ym_`x'>=`PL125' & ym_`x'<`PL250'
    replace ym_`x'_group=3 if ym_`x'>=`PL250' & ym_`x'<`PL400'
    replace ym_`x'_group=4 if ym_`x'>=`PL400' & ym_`x'<`PL1000'
    replace ym_`x'_group=5 if ym_`x'>=`PL1000' & ym_`x'<`PL5000'
    replace ym_`x'_group=6 if ym_`x'>=`PL5000'
    label values ym_`x'_group groups
}
```

Having outlined some relevant definitions, we now turn to each sheet of the Master Workbook Template's incidence results and indicators tables (i.e., the fourth section of the Template) individually.

i. Sheet 1 – Reduction in Inequality and Poverty

This sheet shows the change in inequality and poverty measures across the different income concepts, as well as the significance of these changes and the CEQ effectiveness indicators. The inequality indicators are the Gini, Theil, and 90/10 indices. The poverty indices included are the headcount index, poverty gap index, and squared poverty gap index. The CEQ effectiveness indicators will be defined under the section Sheet 2 – Effectiveness Indicators (even though they also appear on Sheet 1), since that output sheet is devoted entirely to effectiveness indicators.

Graphically, the Gini is represented by twice the area between the market income Lorenz curve and the line of equality. The market income Lorenz curve maps the cumulative share of market income on the vertical axis against the cumulative share of the population, ordered by market income, on the horizontal axis. It equals $2 \int_0^1 (p - L(p)) dp$, where p is the cumulative proportion of the total population when individuals are ordered in increasing income values using market income (graphically, p is also equivalent to the line of perfect equality) and $L(p)$ is the Lorenz curve. There are various user-written Stata commands to compute the Gini, including `concindexi` (calculates concentration coefficients and Ginis), `igini` (part of the Distributive Analysis Stata Package [DASP; Araar and Duclos, 2009]), and `ainequal`.

Sheet 1 also asks for the change in Gini and the statistical significance of that change from zero, measured by the p-value. This encounters computational nuances, since the p-value of the change is a function of the variance of $G(a) - G(b)$, where $G(a)$ is the Gini of income concept a (in the case of Sheet 1, market or net market income), and $G(b)$ is the Gini of income concept b . Thus one must calculate $\text{Var}[G(a) - G(b)] = \text{Var}(G(a)) + \text{Var}(G(b)) - 2\text{Cov}(G(a), G(b))$. The nuance lies in the fact that there are multiple ways to calculate $\text{Cov}(G(a), G(b))$, as described in Yitzhaki and Schechtman (forthcoming, Chapter 3).

For simplicity and to avoid error, we suggest using the user-written Stata command `digini` (part of the DASP package), which automatically calculates the significance of a change in Gini. The drawback of this command is that it is computationally burdensome in terms of the computer resources it requires, so using it can result in a slow do-file.¹⁴ The following sample Stata code uses `digini` to produce a matrix analogous to the Gini portion of the main table in Sheet 1, but without effectiveness indicators. Note that the `digini` command does not allow the incorporation of weights using the traditional `[weight]` syntax, but instead automatically uses the weights that are saved into the dataset using the `svyset` command; using `svyset` to save sampling weight and

¹⁴ Do-file refers to the text file containing Stata code, which is saved with the .do extension.

stratification information is explained under Sheet 11 – Probit of the Before and After Transfers Poor. The example below also loops through the benchmark case and sensitivity analyses, with income variables saved in Stata using the same names as in previous examples: e.g., ym_BC is benchmark case market income.

```
quietly {
* Preliminaries
count
local N = r(N) /* for degrees of freedom for t-test. weights not used to
                determine degrees of freedom */
local xlist BC SA1 SA2 SA3 // to loop through benchmark and sensitivities
local ylist m n d pf fstar f // income definitions
local i=1
foreach y in `ylist' {
    local `y'=`i'
    local i=`i'+1
} /* this loop makes locals m n d pf fstar f that equal 1 2 3 4 5 6 which I
    use later to put my results in the results matrix (i.e. this loop does
    the equivalent of a reverse tokenize for `ylist') */
* Calculations
foreach x in `xlist' {
    noisily display "`x'"
    digini ym_`x' yn_`x' // requires installation of DASP; use SCALED income
    scalar measurem = el(e(d1),1,1) // ineq for market income
    scalar measuren = el(e(d2),1,1) // ineq for net market income
    scalar diffmn = el(e(di),1,1) // difference
    scalar pvaluemn = 2*ttail(`N'-2, abs(el(e(di),1,1)/el(e(di),1,2)) )
        // pvalue of difference
    matrix results = J(5,6,.z)
    matrix results[1,1] = measurem
    matrix results[1,2] = measuren
    matrix results[2,2] = diffmn
    matrix results[3,2] = pvaluemn
    foreach y in d pf fstar f {
        foreach b in m n {
            digini y`b'_`x' y`y'_`x'
            scalar measure`y' = el(e(d2),1,1) // ineq for looped income measures
            scalar diff`b'`y' = el(e(di),1,1) /* difference between market or net
                market and looped income measure */
            scalar pvalue`b'`y' = 2*ttail(`N'-2, abs(el(e(di),1,1)/el(e(di),1,2)) )
                // pvalue of difference
        } // end of b-loop
    matrix results[1,`y'] = measure`y' /* note ``y'' uses the locals I
        created earlier with the reverse tokenize loop */
    matrix results[2,`y'] = diffm`y'
    matrix results[3,`y'] = pvaluem`y'
    matrix results[4,`y'] = diffn`y'
    matrix results[5,`y'] = pvaluen`y'
    } // end of y-loop
matrix colnames results = MARKET NET_MARKET DISPOSABLE POST-FISCAL ///
                        FINAL* FINAL
matrix rownames results = Gini Diff_wrt_Market (p-value) ///
                        Diff_wrt_Net_Mkt (p-value)
noisily matlist results, title("Sheet 1 - Gini") names(all) ///
    aligncolnames(ralign) nodotz underscore format(%11.4f) twidth(16) ///
    lines(columns) linesize(100)
```

```

    } // end of xlist-loop
} // end quietly

```

The Theil index, also known as the Theil's T index, is a member of the family of generalized entropy inequality measures, with the parameter $\theta = 1$. Hence, it is sometimes written as GE(1), and is defined as

$$GE(1) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\bar{y}} \ln \left(\frac{y_i}{\bar{y}} \right)$$

where y_i is individual i 's (household per capita) income, using whichever income concept the Theil is being calculated for, and \bar{y} denotes average income. Sampling weights are omitted from the above equation for notational simplicity, but they should of course be included in the calculation. A matrix analogous to the Theil portion of Sheet 1 can be produced using the same syntax as in the above Gini example, except replacing the line `digini ym_`x' yn_`x'` with `dientropy ym_`x' yn_`x', theta(1)` and similarly for the other `digini` line. The `dientropy` command is also a component of the DASP package.

The 90/10 measure how the relatively rich fare compared to the relatively poor. Specifically, after dividing the population into 100 income percentiles, it is calculated as the average income of those in the 90th percentile divided by the average income of those in the 10th percentile. A matrix analogous to the 90/10 portion of Sheet 1 can be produced using the same syntax as in the above Gini example, except replacing the line `digini ym_`x' yn_`x'` with `dinineq ym_`x' yn_`x', p1(.9) p2(.1)`. The `dinineq` command is also a component of the DASP package.

The poverty indices included in Sheet 1 are members of the FGT class of poverty measures (Foster, Greer, and Thorbecke, 1984). Let households be ranked by y_i , household per capita income for the income variable for which poverty is being measured, from poorest to richest. Let the poverty line being used be denoted z . Then, following Foster, Greer, and Thorbecke (1984), denote $g_i = z - y_i$ the income shortfall of individual i (i.e., the increase in income that would be required for individual i to no longer be poor), and let q denote the number of poor individuals and n the total number of individuals. Then the FGT class of poverty measures is a function of the population's ordered income vector $\mathbf{y} = (y_1, \dots, y_n)$ and the poverty line z , and is defined as follows:

$$P_\alpha(\mathbf{y}; z) = \frac{1}{n} \sum_{i=1}^q \left(\frac{g_i}{z} \right)^\alpha.$$

The headcount index, or the proportion of the population that has income below the poverty line, is equal to the above equation with parameter $\alpha = 0$. The poverty gap, which measures the average shortfall (over the whole population, where non-poor individuals are assigned a shortfall of zero) as a proportion of the poverty line, is equal to the above equation with the parameter $\alpha = 1$. Finally, the squared poverty gap is distribution-sensitive, giving a higher weight to those who are poorer by weighting each individual's shortfall relative to the poverty line by itself (i.e., squaring it). It is equal to the above equation with parameter $\alpha = 2$. Sampling weights are omitted from the above equation for notational simplicity, but should of course be included in the calculation.

There are many user-written Stata programs to calculate the FGT indices, such as `apoverity`. However, since Sheet 1 also asks for the significance of changes in poverty over different income concepts, we again recommend using the DASP package. The syntax is somewhat different than before due to the way results are stored for poverty measures as opposed to inequality measures in DASP, as illustrated by the sample code below. The example below produces results for the international poverty lines, which are saved in locals ``PL125'`, ``PL250'` and ``PL400'`, which can be done using the sample code from the poverty lines section. As usual, it loops over the benchmark case and sensitivity analyses.

```
quietly {
foreach x in BC SA1 SA2 SA3 { // loop over benchmark etc.
  noisily display "Scenario: `x'"
  foreach n in 125 250 400 { // loop over poverty lines
    scalar n_as_decimal = `n'/100
    noisily display "Poverty Line: $" n_as_decimal " PPP"
    foreach alpha in 0 1 2 { /* alpha parameters for headcount ratio, poverty
                             gap, squared poverty gap */
      difgt ym_`x' yn_`x', alpha(`alpha') pline1(`PL`n'') pline2(`PL`n'')
      // use NON-SCALED income
      scalar measurem = r(est1) // pov for market income
      scalar measuren = r(est2) // pov for net market income
      scalar diffmn = r(est3) // difference
      scalar pvaluemn = 2*ttail(`N'-2, abs(r(est3)/r(std3))) // pvalue of dif
      matrix results = J(5,4,.z)
      matrix results[1,1] = measurem
      matrix results[1,2] = measuren
      matrix results[2,2] = diffmn
      matrix results[3,2] = pvaluemn
      foreach y in d pf {
        foreach b in m n {
          difgt y`b'_`x' y`y'_`x', alpha(`alpha') ///
            pline1(`PL`n'') pline2(`PL`n'') // NON-SCALED income
          scalar measure`y' = r(est2) // pov for looped income measure
          scalar diff`b`y' = r(est3) /* difference between market or net
                                     market and looped income measure */
          scalar pvalue`b`y' = 2*ttail(`N'-2, abs(r(est3)/r(std3)))
          // pvalue of difference
        } // end of b-loop
        matrix results[1,`y''] = measure`y' /* note ``y'' uses the locals I
created earlier (in previous sample code) with the reverse tokenize loop */
        matrix results[2,`y''] = diffm`y'
        matrix results[3,`y''] = pvaluem`y'
        matrix results[4,`y''] = diffn`y'
        matrix results[5,`y''] = pvaluen`y'
      } // end of y-loop
      matrix colnames results = MARKET NET_MARKET DISPOSABLE POST-FISCAL
      matrix rownames results = P`alpha' Diff_wrt_Market (p-value) ///
                               Diff_wrt_Net_Mkt (p-value)
      noisily matlist results, title("`alpha'") names(all) ///
        aligncolnames(ralign) nodotz underscore format(%11.4f) twidth(16) ///
        lines(columns) linesize(100)
    } // end of alpha-loop
  } // end n-loop for poverty lines
} // end quietly
```


Sheet 1 also asks for GDP, total non-scaled disposable income according to the household survey, direct transfers (only those included in the analysis) according to national accounts, direct and in-kind transfers (again, only those included in the analysis) according to national accounts, non-scaled direct transfers according to the household survey, non-scaled direct and in-kind transfers according to the survey, and for the latter four, their counterparts when pensions are added in, for the sensitivity analysis. These are used for the effectiveness indicators, which are explained in the next section.

ii. Sheet 2 – Effectiveness Indicators

The CEQ Effectiveness Indicator can be defined for any inequality or poverty measure of interest. In Table 1, we provide effectiveness indicators for the Gini coefficient and headcount index. The indicator is defined as the redistributive effect or effect on poverty of the transfers being analyzed divided by their relative size. Specifically, it is defined as follows for the Gini. Note that it would be similarly defined for any other inequality or poverty measure by replacing the word Gini with the appropriate measure. For direct transfers, the effectiveness indicator is the fall between the net market income and disposable income Ginis as a percent of the net market income Gini, divided by the size of direct transfers (only those included in the analysis) as a percent of GDP. For direct and in-kind transfers, the effectiveness indicator is the fall between the net market income and final income* Ginis as a percent of the net market income Gini, divided by the size of the sum of direct transfers (only those included in the analysis), education spending, health spending, and (where it was included in the analysis) housing and urban spending, as a percent of GDP.

In mathematical notation, let $X(I^j)$ be the inequality or poverty measure of interest (e.g., the Gini coefficient or headcount index), which is defined at each benchmark case income concept $j = m, n, d, pf, f, f^*$ (market income, net market income, disposable income, post-fiscal income and final income) and each sensitivity analysis income concept $j = ms, ns, ds, pfs, fs, f^*s$. Let S^D be total public spending on the direct transfer programs captured by the survey or otherwise estimated by the authors, measured by budget size in national accounts (note that in the sensitivity analysis this concept includes spending in social security pensions), and let S^H, S^E and S^U be total public spending on health, education, and (where included) housing programs, respectively. Then the effectiveness indicator for direct transfers is defined as:

$$\frac{(X(I^n) - X(I^d))/X(I^n)}{S^D/GDP}$$

and the effectiveness indicator for direct and in-kind transfers is defined as:

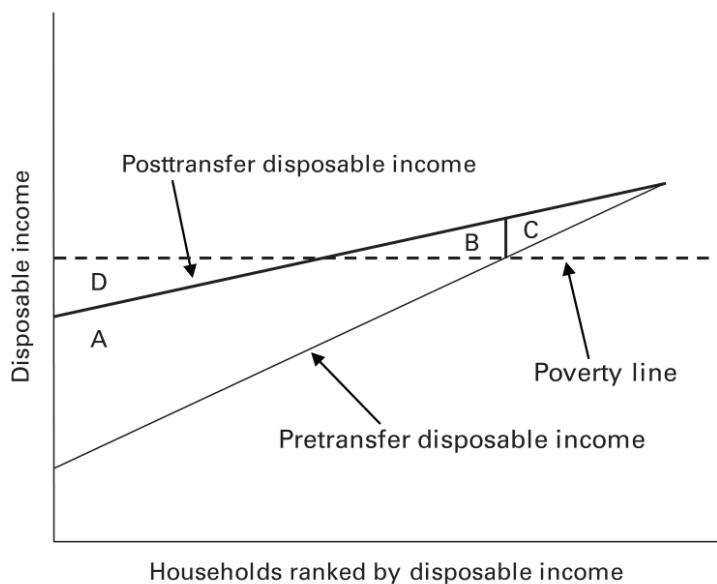
$$\frac{(X(I^n) - X(I^{f*}))/X(I^n)}{(S^D + S^H + S^E + S^U)/GDP}$$

Note that in the sensitivity analysis, when contributory pensions are considered a government transfer, they are not part of net market income but are part of disposable income, thus some of the change between $X(I^{ns})$ and $X(I^{ds})$ is attributable to contributory pensions, and therefore in the sensitivity analysis S^D must include spending on contributory pensions. In the benchmark case, however, contributory pensions are already included in net market income, so S^D does not include any spending on contributory pensions. Also note that S^U should only be included in the denominator of the effectiveness indicator for direct and in-kind transfers if housing programs are included in the analysis.

The above describes the effectiveness indicator identified that we call “(national accounts)” because we use transfer budget sizes from national accounts and GDP for the denominator. The alternative effectiveness indicator, which we denote “(household survey)”, uses non-scaled transfer sizes from the survey and, in place of GDP, non-scaled total disposable income according to the survey.

In addition to the CEQ Effectiveness Indicators, Sheet 2 includes a number of other effectiveness indicators developed by other authors. We use the three poverty-based effectiveness indicators from Beckerman (1979): vertical expenditure efficiency, poverty reduction efficiency, and spillover index, as well as poverty gap efficiency, which was added by Immervoll et al. (2009). Non-scaled income should be used for these effectiveness indicators since they are poverty-based. Diagram 3 is intended to aid the explanation of these indicators.

Diagram 3. Additional Efficiency Indicators



Source: Adapted from Beckerman (1979).

The diagram is not to scale, nor are the income curves necessarily straight. In the diagram, total direct transfers (not scaled) is $A+B+C$, direct transfers reaching the net market income poor is $A+B$,

the total net market income poverty gap is A+D, and the total disposable income poverty gap is D. Beckerman (1979) then defines:

$$\text{Vertical expenditure efficiency} = (A+B)/(A+B+C)$$

$$\text{Spillover index} = B/(A+B)$$

$$\text{Poverty reduction efficiency} = A/(A+B+C).$$

Immervoll et al. (2009) additionally define:

$$\text{Poverty gap efficiency} = A/(A+D).$$

In more technical notation, we have:

$$\text{Vertical expenditure efficiency} = \frac{\sum_{\{i | y_i^n < z\}} \omega_i (y_i^d - y_i^n)}{\sum_i \omega_i (y_i^d - y_i^n)}$$

$$\text{Spillover index} = \frac{\sum_{\{i | y_i^n < z \leq y_i^d\}} \omega_i (y_i^d - z)}{\sum_{\{i | y_i^n < z\}} \omega_i (y_i^d - y_i^n)}$$

$$\text{Poverty reduction efficiency} = \frac{\sum_{\{i | y_i^d < z\}} \omega_i (y_i^d - y_i^n) + \sum_{\{i | y_i^n < z \leq y_i^d\}} \omega_i (z - y_i^n)}{\sum_i \omega_i (y_i^d - y_i^n)}$$

$$\text{Poverty gap efficiency} = \frac{\sum_{\{i | y_i^d < z\}} \omega_i (y_i^d - y_i^n) + \sum_{\{i | y_i^n < z \leq y_i^d\}} \omega_i (z - y_i^n)}{\sum_{\{i | y_i^n < z\}} \omega_i (z - y_i^n)}$$

where ω_i is the sampling weight corresponding to observation i , y_i^n is individual i 's household per capita net market income, y_i^d is individual i 's household per capita disposable income, and z is the poverty line. In the case of national poverty lines that vary by region, or the CEPAL poverty lines which vary depending on whether an individual lives in a rural or urban area, z in the above equations would simply be replaced by z_i , where the latter varies across individuals depending on their location. These effectiveness indicators can be calculated as follows for moderate poverty:

```
quietly {
local PPP = 1.7116
local PPPyr = `PPP'*365 // divide by 12 if monthly
scalar z = 4*`PPPy` // $4 PPP poverty line in local currency per year
gen difference = yd_BC - yn_BC
gen pre_shortfall = z - yn_BC
gen post_shortfall = z - yd_BC
summarize difference [aw=s_weight] if yn_BC < z
scalar AB = r(sum)
summarize difference [aw=s_weight]
scalar ABC = r(sum)
summarize post_shortfall [aw=s_weight] if yn_BC < z & yd_BC >= z
scalar B = -r(sum)
summarize difference [aw=s_weight] if yd_BC < z
```

```

scalar A1 = r(sum)
summarize pre_shortfall [aw=s_weight] if yn_BC < z & yd_BC >= z
scalar A2 = r(sum)
scalar A = A1 + A2
summarize pre_shortfall [aw=s_weight] if yn_BC < z
scalar AD = r(sum)
scalar VEE = AB/ABC
scalar Spillover = B/AB
scalar PRE = A/ABC
scalar PGE = A/AD
noisily display "Vertical Expenditure Efficiency: " %7.6f VEE
noisily display "Spillover index: " %7.6f Spillover
noisily display "Poverty Reduction Efficiency: " %7.6f PRE
noisily display "Poverty Gap Efficiency: " %7.6f PGE
}

```

Finally, we include commonly used tax productivity indicators from Gallagher (2005). These are as follows: (i) the number of taxes that comprise the top seventy-five percent of receipts; (ii) percentage of total taxpayers that provide seventy-five percent of tax receipts; (iii) total number of tax rates; (iv) VAT rate; (v) indirect as percentage of total taxes; (vi) VAT collections as a percent of total tax collections; (vii) tax ratio, defined as the ratio of actual tax collections to GDP; (viii) administrative cost of taxation; (ix) gross compliance ratio, which is the actual VAT collection divided by potential VAT collection; and (x) VAT productivity, defined as the ratio of VAT collections to GDP divided by the nominal VAT rate.

iii. Sheet 3 – Measures of Progressivity and Horizontal and Vertical Inequality

A useful summary statistic to measure progressivity is the Kakwani index. For taxes, the Kakwani (1977) index of progressivity can be thought of graphically as twice the area between the market income Lorenz curve and the tax concentration curve. If the tax concentration curve is above the Lorenz curve, the Kakwani index will be negative, which indicates that taxes are regressive in relative terms. Equivalently, the Kakwani index can be calculated as the tax's concentration coefficient (with the population ranked by market income) minus the market income Gini. Recall that the scaled-up version of market income must be used. In other words, $K^{\text{tax}} = D_m^{\text{tax}} - G^m$, where D_m^{tax} represents the concentration coefficient of a particular tax when the population is ranked by market income.

To adapt to the measurement of transfers, Lambert (1985) suggests that in the case of transfers it should be defined as market income Gini minus the concentration coefficient (i.e., the negative of the definition for taxes) to make the index positive whenever the change is progressive. Thus, we have $K^{\text{transfer}} = -(D_m^{\text{transfer}} - G^m)$, where D_m^{transfer} represents the concentration coefficient of a particular transfer when the population is ranked by market income.

Sheet 3 asks for the Kakwani index for direct transfers, direct and in-kind transfers, direct taxes, indirect taxes, and all taxes. To capture the progressivity of direct transfers and taxes and indirect subsidies and taxes all combined, Sheet 3 also asks for the Reynolds-Smolensky index of post-fiscal income with respect to market income. Graphically, the Reynolds-Smolensky of post-fiscal income

with respect to market income is twice the area between the market income Lorenz curve and the concentration curve of post-fiscal income with respect to the market income distribution. Note that the concentration curve of post-fiscal income is not the same as the Lorenz curve for post-fiscal income, as the concentration curve does not re-rank the population (population is still ranked by market income), whereas the Lorenz curve does re-rank the population (population would be re-ranked by post-fiscal income). Equivalently, the Reynolds-Smolensky can be calculated as the market income Gini minus the concentration coefficient of post-fiscal income when the population is ranked by market income. In other words, $RS = G^m - D_m^{pf}$, where D_m^{pf} represents the concentration coefficient of post-fiscal income when the population is ranked by market income. Recall that scaled up incomes must be used.

The following sample Stata code, which uses the user-written command `concindexi` to calculate both Ginis and concentration indices, can be used to calculate the Kakwani and Reynold's Smolensky indices. Let the variable `directtransfers` represent scaled household per capita direct transfers, `transfers` scaled direct and in-kind transfers, `directtax` scaled direct taxes, `indirecttax` scaled indirect taxes, and `alltax` all (direct plus indirect) taxes. The following code is for the benchmark case only, but could easily be looped over, as in the previous examples, to also produce results for the sensitivity analyses. In this example `ym_BC` represents scaled benchmark case market income.

```
quietly {

* market income Gini
concindexi ym_BC [aw=s_weight], welfarevar(ym_BC) clean // clean option nec.
matrix Gini = r(CII)
scalar gini = Gini[1,1]

* kakwani: transfers
noisily display "KAKWANIS"
foreach tran in directtransfers transfers {
    concindexi `tran' [aw=s_weight] if i==1 , welfarevar(ym_BC) clean
    matrix CC`tran' = r(CII)
    scalar D_`tran' = CC`tran'[1,1]
    scalar kakwani_`tran' = -(D_`tran' - gini)
    noisily display "`tran'"
    noisily display kakwani_`tran'
} // end transfers loop

* kakwani: taxes
foreach tax in directtax indirecttax alltax {
    concindexi `tax' [aw=s_weight], welfarevar(ym_BC) clean
    matrix CC = r(CII)
    scalar D_`tax' = CC[1,1]
    scalar kakwani_`tax' = D_`tax' - gini
    noisily display "`tax'"
    noisily display kakwani_`tax'
} // end of taxes loop

* reynolds-smolensky: post-fiscal
concindexi ypf_BC [aw=w], welfarevar(ym_BC) clean
matrix RS = r(CII)
scalar rs = RS[1,1]
scalar reynoldssmolensky = gini - rs
noisily display "REYNOLDS SMOLENSKY"
```

```
noisily display "of post-fiscal income wrt market income"
noisily display reynoldssmolensky

} // end quietly
```

In addition to measuring progressivity, Sheet 3 decomposes a change in inequality into vertical and horizontal equity components. Vertical equity is concerned with the extent to which a policy equalizes incomes, and is thus closely linked to the measures of progressivity also on this output sheet. Horizontal equity, on the other hand, is concerned with how pre-policy equals are treated, postulating – in the classical definition of horizontal equity – that they should be treated equally. The re-ranking definition of horizontal equity differs slightly, postulating that the pre-policy income ranking should be preserved; for example, if individual A was poorer than individual B before policy C, but receives enough transfer benefits from policy C that she becomes richer than individual B after policy C, there is horizontal *inequity*.¹⁵

Using the reranking definition of horizontal equity, the change in Gini between pre-taxes and transfers, and post-taxes and transfers can be decomposed into the Kakwani vertical effect (Kakwani, 1984) and Atkinson-Plotnick index of reranking (Atkinson, 1980; Plotnick, 1981), as follows. Let X denote income before taxes and transfers (market income) and N indicate income after direct taxes (net market income), or income after taxes and transfers (disposable, post-fiscal, final*, or final income). Let $G(I)$ denote the Gini for income definition I , where $I \in \{X, N\}$. Let $D(N)$ denote the concentration coefficient for income after taxes when individuals are ranked by income before taxes (market income). Then the change in Gini, $G(X) - G(N)$, can be decomposed into $[G(X) - D(N)] - [G(N) - D(N)]$, where the first term, $G(X) - D(N) = V$, the Kakwani vertical effect, and the second term, $G(N) - D(N) = R$, the Atkinson-Plotnick index of reranking. Thus the observed (percentage point) change in Gini equals $V - R$, where V measures the change that could have occurred in the absence of reranking.¹⁶ These can be easily computed using the `concindexi` command, as illustrated below. The following code produces a matrix similar to the benchmark case portion of the V-R decomposition table in Sheet 3.

```
quietly {
matrix results = J(2,5,.z) // create empty matrix for results

* Ginis
foreach y in m n d pf fstar f { // loop over income concepts
concindexi y`y'_BC [aw=w], welfarevar(y`y'_BC) clean // Ginis
matrix Gini`y' = r(CII)
scalar gini`y' = Gini`y'[1,1]
} // end loop over income concepts

* Concentration coefficients of post-tax income wrt market income, V and R
foreach y in n d pf fstar f { // loop over post-tax income concepts
concindexi y`y'_BC [aw=w], welfarevar(ym_BC) clean // post-tax CCs
matrix D`y' = r(CII)
scalar d`y' = D`y'[1,1]
scalar V`y' = ginim - d`y'
matrix results[1,`y'-'1] = V`y' /* note ``y'' uses the locals I created
```

¹⁵ A more thorough summary of vertical and horizontal equity can be found in Duclos (2008).

¹⁶ For details on the history of this decomposition method and alternative formulations, see Urban (2009).

```

    earlier (in previous sample code) with the reverse tokenize loop */
    scalar R`y' = gini`y' - d`y'
    matrix results[2,`y'`-1] = R`y' /* note ``y'' uses the locals I created
    earlier (in previous sample code) with the reverse tokenize loop */
} // end loop over post-tax income concepts

* Display results
matrix colnames results = m_to_n m_to_d m_to_pf m_to_f* m_to_f
matrix rownames results = V R
noisily matlist results, title("Sheet 2 - VR Decomposition") names(all) ///
    aligncolnames(ralign) nodotz underscore format(%11.4f) twidth(6) ///
    lines(columns) linesize(100)

} // end quietly

```

Using the classical definition of horizontal equity, the change in Gini can be decomposed into a vertical equity component and the horizontal inequity component, using non-parametrics to determine pre-tax "equals." Following Duclos and Lambert (2000), the horizontal inequity component equals the per capita gain in revenue that would come from substituting the tax system by a welfare-neutral horizontally equitable tax system. It can be calculated using the DAD software (Duclos and Araar, 2006), which is free stand-alone software. Although DAD is stand-alone and thus not used within Stata, it can load Stata data sets.¹⁷

Although Duclos, Jalbert, and Araar (2003) develop a decomposition methodology that decomposes a change in inequality into vertical equity, classical horizontal inequity, and re-ranking components (rather than just vertical equity and one of the two types of horizontal inequity), its implementation “has not yet been widely employed in empirical research,” possibly because it “requires a certain expertise related to data smoothing and curve-fitting methods” (Urban, 2012, p. 1-2). Although the aforementioned DAD software includes a module to calculate this decomposition, Urban (2012) identifies problems in the module’s weighting scheme and irregularities in the results. As a result, Sheet 3 does not include the Duclos-Jalbert-Araar decomposition.

iv. Sheet 4 – Incidence by Decile and Socioeconomic Groups

The incidence sheet asks for the totals (in local currency per month or year) by decile and income group (defined above) of various categories of income components, as well as for each income definition. These totals are used to automatically generate incidence by decile or income group, which is defined as the total value of a particular income component or income concept received by a certain market income decile divided by the total market income of that same decile. The calculation is non-anonymous, meaning that we do not re-rank the population: the totals by decile that we are comparing are always by market income decile: in other words, we are measuring how much the incomes of identified individuals change when we add in certain income components.

Because many countries lack household survey data on direct transfers and have difficulty simulating them with accuracy, we also have a “Part B” of Sheet 4 that uses net market income, rather than

¹⁷ DAD can be downloaded from <http://dad.ecn.ulaval.ca/>

market income, to define deciles and income groups, to ensure strict comparability across countries. Nevertheless, all countries should fill out both Part A and Part B if possible.

For the incidence tables, scaled income should be used (except in the case of Sensitivity Analysis 3).

After creating deciles and income groups as explained in Section 2, the totals table can be generated using the `table` command with the `contents()` option, as shown below. Let the benchmark case market income deciles be saved as `ym_BC_decile`, household per capita benchmark case market and net market income `ym_BC` and `yn_BC`, respectively, household per capita direct taxes as `directtax`, and contributions to Social Security (only those directed towards things other than pensions, as explained in Section 2) as `contrib_to_SS_excl_pensions`. To display the first few columns of the incidence table, the command would be:

```
table ym_BC_decile [pw=s_weight], contents(sum ym_BC sum directtax ///
      sum contrib_to_SS_excl_pensions sum yn_BC) row format(%16.0f)
```

Similarly, the table that is separated by income groups rather than by deciles can be generated by replacing `ym_BC_decile` with the income group variable, e.g., `ym_BC_group`. The `row` option is used to provide a total for the population at the bottom of each column, and the `format()` option is used to ensure that the totals do not default to being expressed in scientific notation, since they are often large numbers.

Columns to the right of the main portion of the incidence table are included for tax exemptions and population. Tax exemptions are a form of subsidy, but they are already being accounted for in income in the form of lower taxes paid, so imputing them into people's incomes as a subsidy would be double-counting. Nevertheless, we want to know how the amount received by each decile compares with their market incomes, so we add a separate column at the end for tax exemptions. The population totals serve two purposes: (i) they ensure that all deciles are approximately the same size, and (ii) for income groups, they allow us to see how much each group receives in proportion to their population size.

v. Sheet 5 – Concentration Shares by Decile and Socioeconomic Groups

This sheet shows the concentration shares by decile or income group of the income components and concepts that were included in Sheet 4. In fact, it uses the data entered into the totals table on Sheet 4, so no additional data need be entered on Sheet 5; the sheet is automatically generated using Excel formulas.

Sheet 5 is again non-anonymous, meaning that deciles and income groups are always defined by one's market income (or net market income in the case of Part B, for the reasons explained under Sheet 4).

vi. Sheet 6 – Income Distribution by Decile and Socioeconomic Groups

In contrast with Sheet 5 in which individuals are ranked by initial household per capita market income (non-anonymous), in Sheet 6 the population is re-ranked at each income concept before we measure the total income of each decile or income group. When we use numbers from this sheet to measure changes in income, we are comparing, for example, the total income of the poorest *final income* decile with the market income of the poorest *market income* decile, even though these deciles are not necessarily composed of the same individuals. Standard measures of inequality, such as the Gini coefficient, and the FGT poverty indicators, are anonymous: i.e., we do not care about the previous rank or income level (sometimes called the ‘reference point’) of individuals. The information on this table can again be calculated using the `table` command, looping over income concepts:

```
foreach y in m n d pf fstar f {
    display "Income concept: `y'"
    table y`y'_BC_decile [pw=s_weight], contents(sum y`y'_BC) row ///
    format(%16.0f)
}
```

vii. Sheet 7 – Fiscal Incidence Curves and Fiscal Mobility Profiles by Deciles

Sheet 7 provides a graphical representation of some of the information contained in Sheets 4 and 6. It compares the incidence of transfers and taxes with “post-fisc” incomes, both without re-ranking (nonanonymous; Sheet 4) and with re-ranking (anonymous; Sheet 6). The former are analogous to the Income Mobility Profiles proposed by Van Kerm (2009) and will be called *Fiscal Mobility Profiles* (FMP). The anonymous fiscal incidence curves shall be called *Fiscal Incidence Curves* (FIC); they measure the anonymous redistribution induced by fiscal policy along the entire income distribution. Sheet 7 is based entirely on information entered in Sheets 4 and 6, and the graphs are generated automatically, so no additional information need be entered on Sheet 7.

viii. Sheet 8 – Concentration Coefficients and Budget Shares for Social Spending and by Program

Sheet 8 provides the concentration coefficients of individual transfer programs as well as aggregate categories such as Total Direct Transfers and CEQ Social Spending in Incidence Analysis. Let p be the cumulative proportion of the total population when individuals are ordered in increasing income values using market income, and let $C(p)$ be the concentration curve, i.e., the cumulative proportion of total program benefits (of a particular program or aggregate category) received by the poorest p percent of the population. Then, the concentration coefficient of that program or category is defined as $2 \int_0^1 (p - C(p)) dp$. As discussed in Section 3, a program that is progressive in absolute terms will have a concentration curve above the line of perfect equality, and thus the area $2 \int_0^1 (p - C(p)) dp$ will be negative, implying a negative concentration coefficient.

Note that concentration coefficients of transfers are calculated with respect to an income definition (market income in Part A and net market income in Part B) that does not include the transfer as one of its components. This makes theoretical sense, as we are aiming to judge how benefits are distributed with respect to people's ordering before taxes and transfers, not their ordering once they have received the benefit. Contributory pensions, however, *are* included in benchmark case market income, and we thus do not calculate their concentration coefficient with respect to benchmark case market income. To more clearly illustrate why, we note that Immervoll et al. (2009) show that the concentration coefficient for contributory pensions with respect to an income definition that includes contributory pensions can be higher than the market income Gini, signaling that they are regressive, whereas the Gini of market income net of pensions is higher than the Gini of market income with pensions, indicating that they are progressive. This seemingly contradictory occurrence is avoided by calculating all concentration coefficients of transfers with respect to pre-transfers income. Hence, we do not calculate the concentration coefficient of contributory pensions or CEQ Social Spending in Incidence Analysis plus Pensions with respect to benchmark case market income; instead, we include an additional column to calculate their concentration coefficients with respect to Sensitivity Analysis 1 market income, which does not include contributory pensions as one of its components.

Concentration coefficients and their standard errors (which also must be entered in Sheet 8) can be calculated using the user-written program `concindexi`, which allows a variable list, so the concentration indices for all variables can be calculated in one command. The following example loops over the benchmark case and sensitivity analyses (for the market income variable with respect to which the concentration coefficients are calculated). It additionally loops over using market and net market income as the pre-transfer income with respect to which concentration coefficients are calculated, in order to fill out Parts A and B of the Master Workbook Template, respectively.

```
local m "Part A" // (this is just for organizational purposes so it is
local n "Part B" // easy to tell what corresponds to what)
foreach y in m n { // to loop over part A and part B
    display ""
    display "`y'"
    display ""
    foreach x in BC SA1 SA2 SA3 { // to loop over benchmark and sensitivity
        display ""
        display "`x'"
        concindexi varlist [aw=s_weight], welfarevar(y`y'_'x') clean
    } // end of x-loop
} // end of loop over parts A and B
```

where `varlist` would be replaced by the list of variables indicating the amount of benefits received from particular programs as well as aggregate categories. Scaled income is used, except in the case of

Sensitivity Analysis 3 which is non-scaled by definition. This means that both the income variable entered as the `welfarevar` should be scaled, as well as the variables in `varlist`.

Because the redistributive effect of a particular program is a function of both how progressive it is (measured by the concentration coefficient) and how large it is, Sheet 8 also asks for the budget sizes of each program. It asks for the budget size according to national accounts, which should be given net of administrative costs whenever possible, as well as the non-scaled total income from that source, according to the household survey.

In an effort to be self-contained, Sheet 8 also asks for brief descriptions of all programs in the table, so that someone who is looking at the table can easily look up the details of a particular program, such as what type of transfer it is, its target population, and its conditions.

ix. Sheet 9 – Coverage and Leakages by Program

Sheet 9 measures the coverage of the poor and those in other income groups, leakages to the non-poor, and average benefits per capita, per individual in a beneficiary household, and per transfer recipient. The distinction between the latter two deals with the question of how the “average transfer” should be calculated: because the transfer is added to aggregate household income which is then shared by everyone in the household, an economist would most likely measure the average transfer size among a particular income group as the total benefits received by that group divided by the number of individuals in that group who live in households that received the transfer. On the other hand, when the government reports the average transfer size, it usually reports the total spent on transfers divided by the number of transfer recipients, where a transfer recipient is defined as the individual who physically receives the transfer, and not individuals who live in the same household as a transfer recipient.

The majority of Sheet 9 is filled out automatically using formulas; the researcher must only fill out the (non-scaled) total benefits received by group, number of individuals in beneficiary households by group, and number of recipients by group (individuals who report receiving the transfer, not other members of their household) for each individual program and broad category, as well as the market income totals and population totals by group. The individual programs and broad categories included in Sheet 9 are only examples; they should be replaced by the actual program names, and should be expanded upon. For example, if there are more than one conditional cash transfer (CCT) program, they should be listed in separate rows rather than under one aggregate row for CCTs. In addition, the researcher must enter the PPP conversion factor (row 120) and whether the totals are monthly or yearly (row 121) so that the average benefits can also be calculated in \$PPP per day.

From the total benefits by group and population totals entered by the researcher, the following are automatically calculated in the Master Workbook Template: share of benefits going to each income

group (which can be used to determine what percent of benefits are leakages to the non-poor), share of individuals in beneficiary households in each income, percent of individuals in each group who live in beneficiary households (which can be used to determine coverage of the poor), average per capita benefits among beneficiary households by group, average benefits per capita by group, and average benefits per transfer recipient by group. The average benefits are calculated both in local currency and in \$PPP per day.

Total benefits received by group can be calculated in the same way that was demonstrated for broader income categories in Sheet 4 using the `table` command. Number of individuals in beneficiary households by group and number of recipients by group can be calculated multiple ways; one of the possibilities is discussed here. When constructing household aggregate income from each component, create a dummy variable that equals one for the individuals that report receiving income from that source, and zero for everyone else (including other members of their household that do not report that income source). Suppose for a particular program, this variable is called `dummy_program1`. The next step is to create a second dummy variable that equals one for each member of a household in which someone received a transfer. This can be accomplished as follows:

```
bysort hh_code: egen hh_dummy_program1 = max(dummy_program1)
```

where `hh_code` is the household identifier variable (i.e., it is a variable that has the same value for all members of the same household, and a different value for members of different households). The `max()` function looks for the maximum value of the input variable. When it is combined with `bysort`, it outputs the maximum value of the input variable among individuals with the same value for the variable listed after `bysort`. In this case, it outputs the maximum value of `dummy_program1` for individuals within the same household, so households in which no one is a program recipient (everyone has `dummy_program1 = 0`) will receive a value of zero for `hh_dummy_program1`, and households in which at least one member is a recipient (and thus has `dummy_program1 = 1`) receive a value of one for `hh_dummy_program1`. Then, these dummy variables can be used in combination with the `table` command to count the number of transfer recipients and members of beneficiary households by group for each program as follows:

```
table ym_BC_group [pw=s_weight], contents(sum dummy_program1 ///
      sum hh_dummy_program1) row format(%16.0f)
```

Another measure of interest is the coverage and leakages of these programs among their target population. Thus, Sheet 9 also asks for the total benefits received by group of CCTs among households with children, non-contributory pensions and contributory pensions in households with a member over age 65, and education by level in households with children of the corresponding age. These totals are used to calculate the same measured listed above, except for the target population only.

x. Sheet 10 – Fiscal Mobility Matrices

The fiscal mobility matrix is a transition matrix that measures the proportion of individuals that move from a before taxes and transfers income group (e.g., non-poor) to another income group (e.g., poor) after their income is changed by taxes and transfers. Note that taxes and transfers can cause individuals to move up or down the income categories. The matrix in percents is row-stochastic, where rows represent before taxes and transfers income groups and columns represent after taxes and transfers income groups. Lustig and Higgins (2012) illustrate the importance of the fiscal mobility matrix: standard measures fail to identify downward fiscal mobility among the poor caused, for example, by high consumption taxes. As usual, non-scaled income is used to determine the income group to which an individual belongs. There are multiple matrices for the different possible definitions of post-tax income: for example, there is a mobility matrix for market to disposable income, as well as a mobility matrix for market to post-fiscal income. The mobility matrices have additional rows and columns concatenated to them to show the population shares by income group and the mean market income of that income group, for ease of reference.

To generate the fiscal mobility matrix, the researcher is asked for the total number of individuals in each (i, j) pair, where i is a pre-tax income group and j is a post-tax income group. The matrix of totals can be easily generated in Stata using the `tabulate` command. For example, for the benchmark case mobility matrix from market to post-fiscal income, the command would be:

```
tabulate ym_BC_group ypf_BC_group [iw=s_weight]
```

Mean market income by market and post-fiscal income groups (for the concatenated column and row, respectively) could be calculated with the following code.

```
foreach y in m pf {  
    table y`y'_BC_group [pw=s_weight], contents(mean ym_BC) row format(%16.0f)  
}
```

While the fiscal mobility matrix measures the proportion of the population that loses and gains enough to move to a higher income group, it does not capture the amount lost or gained (except to the extent that the amount lost or gained might be large enough to move more than one income group). Thus, the fiscal mobility matrix is complemented by income loss and income gain matrices, which measure the amount lost by those who lose and the amount gained by those who gain, respectively. One version of the loss and gain matrices is in average local currency lost or gained, and the other shows the average loss or gain as a proportion of before taxes and transfers income. The matrix also shows the average market income of the losers in pre-taxes and transfers income group i and post-taxes and transfers income group j , which serves as a useful reference point. The average loss in currency, average proportional loss, and average market income of losers in cell ij (as well as their counterparts in the gain matrices) can be calculated using the `table` command with the

contents option, this time used with two variables after the command rather than one to create a two-way table. An if-condition determines who is included in the calculation: for the income loss (gain) matrix, only those who have lost (gained) income are considered, i.e., those who have post-taxes and transfers income that is lower (higher) than pre-taxes and transfers income. Thus, the income loss (gain) matrix will be lower (upper) triangular by definition. The following code illustrates an example, where pre-taxes and transfers income is market income and post-taxes and transfers income is post-fiscal income.

```
/* first create a local for the PPP factor since mean market incomes need to
be in $PPP per day, per the Master Workbook Template's instructions */
local PPP=1.7116 /* PPP conversion factor for Brazil 2009 (consumption-based)
                  from databank.worldbank.org */
local PPPyr=`PPP'*365 // Brazil data is annual; if monthly also divide by 12
foreach y in m pf {
    gen y`y'_BC_PPP = y`y'_BC/`PPPyr' // creates income variables in PPP/day
}

* variables for income loss or gain and proportional loss or gain
gen change = ypf_BC - ym_BC
gen change_proportional = (ypf_BC - ym_BC)/ym_BC

* average loss in local currency
table ym_BC_group ypf_BC_group if change<0 [iw=s_weight], contents( ///
    mean change)

* average proportional loss, mean market income in PPP
table ym_BC_group ypf_BC_group if change<0 [iw=s_weight], contents( ///
    mean change_proportional mean ym_BC_PPP)
```

xi. Sheet 11 – Probit of the Before and After Transfers Poor

Two probit regressions are used to measure the correlates of poverty (i) before taxes and transfers and (ii) after taxes and transfers, conditional on being poor before taxes and transfers. The probit regressions are restricted to household heads, and the \$2.50 PPP per day line is used to determine poverty. The first probit measures the probability of being poor before transfers; the dependent variable is a dummy for being market income poor and the independent variables are household characteristic variables. The second probit measures the probability of being poor after transfers conditional on being poor before; i.e., the regression is restricted to the market income poor household heads. The dependent variable is a dummy for being disposable income poor and the independent variables are household characteristic variables.

Since most surveys use a complex sample design with clustering and stratification, if you do not take the sample design into account when you run the probit, your standard errors will be incorrect. Note

that using the sampling weights is *not sufficient*. If the survey you are using has a three-stage sample design, it will have, in addition to the commonly used variable for each observation's sampling weight, a variable for the primary sampling unit and the strata. For a survey with a two stage sample design, it will have a variable for the sampling weight and primary sampling unit only (which is sometimes confusingly called strata in the data sets). In Stata, the survey sample design variables (sampling weight, strata, and primary sampling unit) can be saved with the data set using the `svyset` command (followed by the `save` command so that the next time the data set is opened, Stata will remember the survey sampling design). Once the survey sample design is saved in the data set, commands that are designed to produce standard errors that account for stratification and clustering can be told to account for them using the `svy:` prefix. In addition, some user-written commands such as those that are part of the DASP package (e.g., `digini`, `dientropy`, `dinineq`, and `difgt` used for Sheet 1) automatically use the information about sampling weights, strata, and primary sampling units. However, for programs not in the DASP package, the user should never assume that the command automatically incorporates the survey sampling design information.

Let the sampling weight variable in our data set be saved as `s_weight`, the strata be saved as `s_strata`, and the primary sampling unit be saved as `s_unit`. Then the syntax for saving the sampling information would be

```
svyset s_unit [pw=s_weight], strata(s_strata)
```

In the case of a survey with a two-stage sample design rather than three-stage, the `strata()` option would not be included. After saving, closing, and re-opening the data set, one can make sure that the survey sampling design is saved in the data set by typing `svydes`.

When running the `probit`, it is necessary to use the `svy:` prefix to obtain correct standard errors. For example, the syntax of the `probit` used in Higgins (2012) was as follows¹⁸ for the first `probit` (note that the sampling weights do not need to be included in the command below because they are already saved in the survey sampling design which is invoked by the `svy:` prefix).

```
svy: probit ym_BC_extremepoor ///
      number_children ///
      North_dummy NE_dummy SE_dummy South_dummy ///
      rural ///
      male ///
      age25_40 age41_64 age65_ ///
      pri_incomplete pri_complete sec_complete ter_complete ///
      informal selfemployed unemployed ///
      single divorce_separated widow ///
      black pardo indigenous asian_other ///
if hh_head==1
```

¹⁸ It is slightly adapted here for illustrative purposes.

The first line includes the `svy:` prefix explained above, the `probit` command, and the dependent variable: a dummy variable equal to one for being extremely poor, which can be created with the command `generate ym_BC_extremepoor = (ym_BC_PPP < 2.5)` where `ym_BC_PPP` is non-scaled benchmark case market income converted into \$PPP per day. The `if`-condition in the last line restricts the probit to household heads. The independent variables are: number of children in the household, region in which the household is located, a dummy variable for rural areas, and the following characteristics of the household head: gender, age (as a set of dummy variables for different age ranges), education level (again as a set of dummy variables), labor status, marital status, and self-reported race. For each set of dummy variables, one of the categories must be omitted from the probit. This group serves as the reference group to which the other groups can be compared: it has an implicit coefficient of zero. In the example above for Brazil, the omitted dummies for each group of dummies (which are each on their own line in the code) are: Midwest region, urban, female, age less than 25, never attended school, employed in the formal sector, married, and white. For ease of comparability across countries, please omit similar categories; in other words, omit household heads that are less than 25 rather than including them and instead omitting household heads greater than 65.

The second probit's syntax would be as follows. The differences between its syntax and the syntax of the first probit are in bold.

```
svy: probit yd_BC_extremepoor ///
      number_children ///
      North_dummy NE_dummy SE_dummy South_dummy ///
      rural ///
      male ///
      age25_40 age41_64 age65_ ///
      pri_incomplete pri_complete sec_complete ter_complete ///
      informal selfemployed unemployed ///
      single divorce_separated widow ///
      black pardo indigenous asian_other ///
if hh_head==1 and ym_BC_extremepoor==1
```

In addition to the coefficients from the probit regression, Sheet 11 also asks for the marginal effects for each variable and their standard errors (calculated using the delta method), which can be computed immediately after running each probit regression using the `margins` command. Here we briefly describe the theoretical underpinnings of marginal effects for a probit. In a linear regression, we would have $E[y|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$; the marginal effects are $\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x}} = \boldsymbol{\beta}$. By contrast, in a probit regression, which is non-linear, we have $E[y|\mathbf{x}] = \Phi(\mathbf{x}'\boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \varphi(z)dz$ where $\Phi(\cdot)$ is the cumulative distribution function for the normal distribution and $\varphi(\cdot)$ is its probability distribution function; the marginal effects are $\frac{\partial E[y|\mathbf{x}]}{\partial \mathbf{x}} = \varphi(\mathbf{x}'\boldsymbol{\beta})\boldsymbol{\beta}$. Unlike in the case of linear regression, the marginal effect is dependent on \mathbf{x} , which in turn varies by observation. Thus, we calculate the *average marginal effect* which consists of calculating the individual marginal effect for each observation that was included in the probit (in this case, household heads), using that observation's values for \mathbf{x} and

the probit-generated values for β , and then averaging the individual marginal effects to obtain the average marginal effect. To calculate the average marginal effect, one must simply type the command `margins` after running the probit, which by default incorporates the sampling weights and if-condition that were included in the original probit regression, and calculates the average marginal effect rather than an alternative formulation of the marginal effect (do not use the options `at()` or `atmeans`, which calculate the *marginal effects at a representative value* and *marginal effect at the mean*, respectively).

xii. Sheet 12 – Needs vs. Resources

This sheet compares the amount of money that would be required to eliminate income, health, and education poverty (assuming perfect coverage and targeting) compared with the amount of resources available. It can be very useful when answering the Diagnostic Questionnaire which asks a number of questions that compare needs to resources. The table consists of a “full table” and a “summary table”; the summary table is filled out automatically using Excel formulas based on inputs entered into the full table. Since needs are measured before and after taxes and transfers, there are four scenarios considered: before transfers is market income and after transfers is disposable income; before transfers is market income and after transfers is post-fiscal income; before transfers is net market income and after transfers is disposable income; and before transfers is net market income and after transfers is post-fiscal income. The various elements of Sheet 12 are defined in turn. For simplicity, the definitions below are for the first scenario (before transfers is market income and after transfers is disposable income), but they can be easily adapted to the additional scenarios. Non-scaled incomes are used to determine who is poor and calculate the shortfalls for the income poverty gap.

Income Poverty Gap

This is equal to the total shortfall of the poor’s incomes below the poverty line. This total is *not* normalized by the poverty line or divided by the population, as it was in Sheet 1 to calculate the poverty gap *index*. Continuing the notation used under Sheet 1, denote $g_i = z - y_i$ the income shortfall of individual i (i.e., the increase in income that would be required for individual i to no longer be poor), q denote the number of poor individuals (using whichever income concept is being used), and let the population be ranked by the income concept being used from poorest to richest. Then the income poverty gap is defined as $\sum_{i=1}^q g_i$. The before transfers income poverty gap uses market or net market income, and the after transfers income poverty gap uses disposable or post fiscal income.

Before Transfers Education Poverty Gap

The before transfers education poverty gap is defined as the total annual cost of educating the poor. It is calculated by dividing the annual public spending on education at level l from national accounts by the number of students at level l from national accounts (or,

equivalently, obtaining spending per student by level from national accounts), multiplying that by the number of children who are in the age range that corresponds to education level l and are market (or net market) income poor, then summing over all l , where l are the education levels, e.g., primary, lower secondary, upper secondary. Poor children who are not enrolled in school are included in the calculation of total demand for education among the market income poor, and treated as belonging to the level to which their age corresponds. Poor children who are enrolled in school but are behind their age level are treated as belonging to the level that corresponds to their age, not the actual level in which they are enrolled. The critical ages for schooling are from six to eighteen years old, so children under six years old or over eighteen years old are not included in the calculation. The critical level of schooling is twelve years, so individuals who have already completed twelve years of schooling are not included in the calculation.

After Transfers Education Poverty Gap

The after transfers education poverty gap is calculated similarly to the before transfers education coverage gap, but instead of multiplying the spending per student by level by the total number of market (or net market) income poor students at that level, it is multiplied by the number of market (or net market) income poor students at that level *who are not enrolled in school*.

Before Transfers Health Poverty Gap

The before transfers health poverty gap is defined as the total cost of providing basic health coverage to the poor. From public accounts or national health accounts, obtain the cost of a basic health package. Multiply this by the total number of market (or net market) income poor.

After Transfers Health Poverty Gap

The after transfers health poverty gap is calculated similarly to the before transfers health poverty gap, except instead of multiplying the cost of a basic health package times the number of market (or net market) income poor, it is multiplied by the number of market (or net market) income poor who are not covered by the public health insurance scheme. If data on health coverage is not available or if the country does not have a public health insurance scheme, the after transfers health poverty gap can be defined based on use; in that case, the after transfers health poverty gap is defined as the before transfers health poverty gap minus total in-kind health benefits received by the market (or net market) income poor. The latter is calculated as described in the Income Concepts and Data Requirements section.

Human Capital Poverty Gap

The before (after) transfers human capital poverty gap is the sum of the before (after) transfers education poverty gap and the before (after) transfers health poverty gap.

Overall Poverty Gap

The before (after) transfers overall poverty gap is the sum of the before (after) transfers income poverty gap and the before (after) transfer health poverty gap.

Total Government Spending

Total government spending according to public sector accounts. It should include all social spending, all administrative spending, spending on housing, water, sanitation, etc., spending on economic subsidies, servicing external debt, military spending, etc. It should include both recurrent spending and investment spending (e.g., in education, health, and infrastructure). If you are including subnational spending and taxes in your study, it should include subnational spending. Write down a specific definition of total government spending used in your study and specify whether it is federal/central only or the latter plus subnational. Remember to document the source/s with specific locations of from where the data came.

Primary Government Spending

Primary government spending is equal to total government spending net of domestic and external debt servicing.

Total Government Revenue

Total government revenues include the total budgetary income of the federal/central government: tax and non-tax revenue plus income generated by direct budgetary controlled entities or public enterprises. In countries where revenue collected at the provincial or state level is important, the total should include the revenues obtained by governments at the subnational level if possible. Specify whether subnational revenue is included.

Targeted Anti-Poverty Spending

Targeted anti-poverty spending includes direct transfers programs that by design use a mechanism to target benefits to the poor.

Reaching the poor

Resources reaching the poor is calculated using the household survey, since this information is not available in national accounts.

xiii. Sheet 13 – Cumulative Distribution Functions of Income

This set of graphs shows the cumulative distribution functions of non-scaled benchmark case market, net market, disposable, and post-fiscal income. The cumulative distribution function (CDF) of income is then defined as $\int f(y^j)dy$ where $f(y^j)$ is the probability density function of income concept j . Hence, the CDF is anonymous by definition: the underlying distribution is ranked by whatever income concept is being measured, rather than maintaining the original market income

ranking. Five income CDF graphs should be generated with various levels of “zooming in” along the x-axis, which is the axis measuring income, with the following domains: (i) \$0 to \$2.50 PPP; (ii) \$0 to \$4 PPP; (iii) \$0 to \$10 PPP; (iv) \$0 to \$50 PPP; (v) \$0 to \$100 PPP.

The graphs of income CDFs can be generated using the user-written `incomecdf` available from the CEQ Website.¹⁹ There are two possible syntaxes: the first is to convert the income variables in local currency into \$PPP per day, as described under Sheet 10, and use those variables in the variable list, as follows:

```
incomecdf ym_BC_PPP yn_BC_PPP yd_BC_PPP ypf_BC_PPP
```

The second option is to use income in local currency (which could be in daily, monthly, or yearly terms), then to specify the consumption-based PPP conversion factor and whether the data is daily, monthly, or yearly as options, as follows.

```
incomecdf ym_BC yn_BC yd_BC ypf_BC, ppp(1.7116) yearly
```

where `yearly` can be replaced by `monthly` or `daily`.

xiv. Sheet 14 – Lorenz Curves

This graph shows the Lorenz curves for each income definition. The Lorenz curve maps the cumulative share of income (using whichever income concept the curve corresponds to) on the vertical axis against the cumulative share of the population, ordered by income (using whichever income concept the curve corresponds to), on the horizontal axis. Because the horizontal axis is re-ranked with each income concept, the Lorenz curve is an anonymous measure by definition; its non-anonymous analog would be the concentration curve of each income definition with respect to the market income rankings. For income concept j , the Lorenz curve is defined as

$$L_j(p) = \frac{1}{\bar{y}^j} \int_0^p y^j dF(y^j) \text{ for } p \in [0,1]$$

where \bar{y}^j is mean income, $F(y^j)$ is the cumulative density function of income, and p is the proportion of the population.

In Stata, the command `glcurve` can be used to graph Lorenz curves; however, the command does not allow users to graph multiple Lorenz curves on the same graph without a bit of programming.

¹⁹ To download this program from within Stata, type:
`net from http://www.commitmenttoequity.org`
`net install incomepdf`

As illustrated below, one can use `glcurve` with the `nograph` option to generate a set of coordinates for each observation in the data set which marks where that observation lies on the Lorenz curve. After obtaining coordinates corresponding to each income definition and normalizing the y-axis to obtain the Lorenz rather than generalized Lorenz curves, one can graph all curves on the same graph using the Stata graphing command `twoway`, as illustrated below.

```
* for Lorenz curves:
foreach y in m n d pf f {
    quietly glcurve y`y' _BC [aw=s_weight], pvar(`y'x) glvar(`y'yaxis) ///
        nograph replace
    * normalizing 0 to 1:
    quietly summ `y'yaxis
    scalar max`y'yaxis = r(max)
    gen `y'y = `y'yaxis/max`y'yaxis
}

* for 45 degree line:
gen diagline = my

* for legend:
label var my "Market"
label var ny "Net Market"
label var dy "Disposable"
label var pfy "Post-fiscal"
label var fy "Final"
label var diagline "45 Degree Line"

* graph:
set scheme slcolor // sets the graph formatting scheme
twoway ///
    (line my mx [pw=s_weight], sort) ///
    (line ny nx [pw=s_weight], sort) ///
    (line dy dx [pw=s_weight], sort) ///
    (line pfy pfx [pw=s_weight], sort) ///
    (line fy fx [pw=s_weight], sort) ///
    (line diagline diagline [pw=s_weight], sort clcolor(gray)) ///
    , ///
    ytitle("Cumulative proportion of income") ///
    xtitle("Cumulative proportion of the population") ///
    legend(ring(0) pos(11)) ///
    xscale(range(0 1)) ///
    yscale(range(0 1))
```

xv. Sheet 15 – Inequality of Opportunity²⁰

²⁰ This section is based on a brief description of inequality of opportunity sent to the authors by Norbert Fiess.

Sheet 15 measures ex-ante inequality of opportunity based on circumstances sets.²¹ First, circumstances sets are identified: for example, one circumstances set could be {female, black, parents were college graduates, urban}: all individuals with those four traits are grouped together in that circumstances set. Circumstances are pre-determined factors that are not dependent on an individual's effort, such as race, gender, and parents' education or parents' income. Once each individual's circumstances set has been identified, the mean income of each circumstances set (i.e., the mean income of all individuals in that circumstances set) is calculated for each income concept. Scaled benchmark case income is used for each income concept. Let s_i^j indicate the mean income for income concept j of everyone in individual i 's circumstances set. Each individual is attributed the mean income of their circumstances set, and this income distribution is called the smoothed income distribution. Inequality measured over the smoothed income distribution for each income concept uses the mean log deviation, which gives the measure of inequality of opportunity in levels by income concept. Dividing the resulting measure by the mean log deviation for the original income distribution measures the ratio of inequality due to inequality of opportunity as opposed to inequality of effort. The latter, called inequality of opportunity in ratios on Sheet 15, traces out how each redistributive step affects inequality of opportunity. For example, if the proportion of inequality explained by unequal opportunities decreases from net market to disposable income but increases from disposable to post-fiscal income, this would indicate that direct transfers have an equalizing impact on ex ante opportunities, but indirect taxes and subsidies have an unequalizing effect.

The mean log deviation of the smoothed distribution (for income concept j) is calculated as

$$\frac{1}{n} \sum_i \ln \left(\frac{\mu^j}{s_i^j} \right)$$

where μ^j is the mean income of the population for income concept j (either the original or smoothed distribution can be used to calculate μ^j since they have the same mean by definition), and s_i^j is defined above.

Inequality of opportunities in levels and trends can be calculated by the user-written `oppincidence` available from the CEQ Website,²² using the following syntax. Scaled incomes should be used.

```
oppincidence ym_BC yn_BC yd_BC ypf_BC yf_BC [aw=s_weight], ///
groupby(male race fathers_education mothers_education rural)
```

where the variable list immediately following `oppincidence` lists all the income concepts over which inequality of opportunities is being calculated, and the required argument `groupby()` gives the categorical variables used to determine circumstances sets. Note that, since the variables used to

²¹ See Checchi and Peragine (2010) and Ferreira and Gignoux (2011).

²² To download this program from within Stata, type:
net from <http://www.commitmenttoequity.org>
net install oppincidence

determine circumstances sets must be categorical rather than continuous, variables such as `fathers_education` and `mothers_education` have to be given as categories (i.e., 1 = never attended school, 2 = primary incomplete, 3 = primary complete, etc.) rather than as years of schooling completed.

xvi. Sheet 16 – Progressiveness of Pensions

This sheet summarizes the progressiveness of pensions and CEQ Social Spending, and will be used to construct multi-country summary tables. All of the elements of this sheet are generated automatically using Excel formulas based on inputs from other parts of the Master Workbook Template.

xvii. Sheet 17 – Comparison with Other Studies

The final sheet consists of comparing the results from the incidence analysis with the results from other studies (incidence analyses in particular) for the same country. A thorough comparison table is included as an example in the Master Workbook Template.

References

- Adema, Willem and Maxime Ladaïque. 2005. "Net Social Expenditure, 2005 Edition: More Comprehensive Measures of Social Support." OECD Social, Employment and Migration Working Papers 29.
- Alleyne, Dillon, James Alm, Roy Bahl and Sally Wallace. 2004. "Tax Burden in Jamaica." Georgia State University International Studies Program Working Paper 04-34.
- Araar, Abdelkrim and Jean-Yves Duclos. 2012. "DASP: Distributive Analysis Stata Package. User Manual, DASP version 2.2." http://dasp.ecn.ulaval.ca/modules/DASP_V2.2/DASP_MANUAL_V2.2.pdf
- Atkinson, Anthony B. 1980. "Horizontal Equity and the Distribution of the Tax Burden." in Aaron, H.J., Boskins, M.J. (eds), *The Economics of Taxation* (Washington D.C., Brookings), 3-18.
- _____. 1983. *Social Justice and Public Policy*. MIT Press.
- Barr, Nicholas. 2004. *Economics of the Welfare State: Fourth Edition*. New York: Oxford University Press.
- Barros, Ricardo, Francisco Ferreira, Jose Molinas Vegas and Jaime Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington DC: World Bank. Available online <http://siteresources.worldbank.org/BRAZILEXTN/Resources/322340-1222953505624/BookHOI.pdf>
- Beckerman, Wilfred. 1979. "The impact of income maintenance payments on poverty in Britain, 1975." *Economic Journal* 89, 261-279.
- Bergh, Andreas. 2005. "On the counterfactual problem of welfare state research: How can we measure redistribution?" *European Sociological Review* 21(4).
- Birdsall, Nancy, Augusto de la Torre and Rachel Menezes. 2008. *Fair Growth: Economic Policies for Latin America's Poor and Middle-Income Majority*. Washington DC: Brookings Institution Press.
- Bourguignon, François and Luiz A. Pereira da Silva, eds. 2003. *The Impact of Economic Poverty and Income Distribution*. Washington DC: World Bank.
- Breceda, Karla, Jamele Rigolini and Jaime Saavedra. 2008. "Latin America and the Social Contract: Patterns of Social Spending and Taxation." Policy Research Working Paper 4604. World Bank Latin American and Caribbean Region Poverty Department Poverty Reduction and Economic Management Division. Washington DC: World Bank. Available online <http://go.worldbank.org/BWBRP91A50>
- Bucheli, Marisa, Nora Lustig, Maximo Rossi and Florencia Amabile. 2012. "Social Spending, Taxes and Income Redistribution in Uruguay." Tulane University Economics Working Paper, New Orleans, Louisiana, April 2012.
- CEDLAS (Centro de Estudios Distributivos, Laborales y Sociales) and World Bank. 2012. "A Guide to the SEDLAC Socio-Economic Database for Latin America and the Caribbean." http://sedlac.econo.unlp.edu.ar/download.php?file=archivos_upload_items_metodologia/Guide_14_english.pdf
- Checchi, D. and Peragine, V. 2010. "Inequality of Opportunity in Italy." *Journal of Economic Inequality*, 8 (4), 429-450.
- Chen, Shaohua and Martin Ravallion. 2010. "The Developing World is Poorer than We Thought, but No Less Successful in the Fight Against Poverty." *The Quarterly Journal of Economics* 125(4): 1577-1625.
- Deaton, Angus. 2005. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." *The Review of Economics and Statistics* 87(1): 1-19.

- Dilnot Andrew, John Kay, and Michael Keen. 1990. "Allocating Taxes to Households: A Methodology." *Oxford Working Papers* 42(1): 210-230.
- Duclos, Jean-Yves. 2008. "Horizontal and Vertical Equity." *The New Palgrave Dictionary of Economics*. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.
- Duclos, Jean-Yves, Vincent Jalbert and Abdelkrim Araar. 2003. "Classical Horizontal Inequity and Reranking: An Integrating Approach." *Research and Economic Inequality* 10: 65-100.
- Duclos, Jean-Yves and Abdelkrim Araar. 2006. *Poverty and Equity: Measurement, Policy, and Estimation with DAD*. New York: Springer and International Development Research Centre.
- Duclos, Jean-Yves and Peter Lambert. 2000. "A normative approach to measuring classical horizontal inequity." *Canadian Journal of Economics* 33, 87-113.
- Economic Commission for Latin America and the Caribbean (ECLAC). 2010. *Social Panorama of Latin America*, 2009. Santiago, Chile: United Nations.
- Ferreira, F.H.G., Gignoux, J. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *The Review of Income and Wealth*.
- Ferreira, Francisco, Julián Messina, Jamele Rigolini and Renos Vakis. 2012 (forthcoming). *Socio-Economic Mobility and the Rise of the Middle Class in Latin America and the Caribbean*. World Bank Regional Flagship Report for Latin America and the Caribbean.
- Ferreira, Francisco H.G. and David Robalino. 2010. "Social Protection in Latin America: Achievements and Limitations." Policy Research Working Paper 5305. Washington DC: World Bank, Latin America and Caribbean Region Office of the Chief Economist, and Human Development Network Social Protection and Labor Unit. Available online http://www-wds.worldbank.org/servlet/WDSCContentServer/WDSP/IB/2010/05/10/000158349_20100510134942/Rendered/PDF/WPS5305.pdf
- Fiszbein, Ariel, Norbert Schady, Francisco Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington DC: World Bank.
- Foster, James, Joel Greer, and Erik Thorbecke. 1984. "A class of decomposable poverty measures." *Econometrica* 52(3), 761-766.
- Gallagher, Mark. 2005. "Benchmarking Tax Systems." *Public Administration and Development* 25: 125-144.
- Goñi, Edwin, J. Humberto López, and Luis Servén. 2011. "Fiscal Redistribution and Income Inequality in Latin America." *World Development* 39(9): 1558-1569.
- Grosh, Margaret, Carlo del Ninno, Emil Tesliuc, and Azedine Ouerghi. 2008. *For Protection and Promotion: The Design and Implementation of Effective Safety Nets*. Washington DC: World Bank. Available online http://siteresources.worldbank.org/SAFETYNETSANDTRANSFERS/Resources/For_Protection_and_Promotion_complete.pdf
- Higgins, Sean. 2011. "Predicting Rent to Impute the Value of Owner Occupied Housing." _____ . 2012. "The Impact of Bolsa Família on Poverty: Does Brazil's Conditional Cash Transfer Program Have a Rural Bias?" *Journal of Politics and Society* 23: 88-125.
- Immervoll, Herwig, Horacio Levy, José Ricardo Nogueira, Cathal O'Donoghue, and Rozane Bezerra de Siqueira. 2009. "The Impact of Brazil's Tax-Benefit System on Inequality and Poverty." *Poverty, Inequality, and Policy in Latin America*. Eds. Stephan Klasen, and Felicitas Nowak-Lehmann. Cambridge: Mass.: MIT Press. 271-302.
- Jiménez Pozo, Wilson, Verónica Paz Arauco and Ernesto Yañez Aguilar. 2012. "Fiscal Policy, Urban-Rural Inequality and Rural Poverty in Bolivia." Report prepared for the International Fund for Agricultural Development.

- Kakwani, N.C. 1977. "Measurement of Tax Progressivity: An International Comparison." *The Economic Journal* 87(345): 71-80.
- _____. 1984. "On the measurement of tax progressivity and redistributive effect of taxes with applications to horizontal and vertical equity." *Advances in Econometrics* 3: 149-168.
- Kharas, Homi. 2010. "The Emerging Middle Class in Developing Countries." OECD Development Centre Working Paper 285. Available online <http://www.oecd.org/dataoecd/12/52/44457738.pdf>
- Lambert, Peter. 1985. "On the redistributive effect of taxes and benefits." *Scottish Journal of Political Economy* 32(1): 39-54.
- _____. 2002. *The Distribution and Redistribution of Income: Third Edition*. Manchester United Kingdom: Manchester University Press.
- Lindert, Kathy, Emmanuel Skoufias, and Joseph Shapiro. 2006. "Redistributing Income to the Poor and Rich: Public Transfers in Latin America and the Caribbean." Social Protection Discussion Paper 0605. Washington, D.C.: The World Bank.
- Lora, Eduardo, ed. 2006. *The State of State Reforms in Latin America*. Washington DC: World Bank.
- Lopez-Calva, Luis F. and Ortiz-Juarez, Eduardo. 2011. "A Vulnerability Approach to the Definition of the Middle Class." Policy Research Working Paper 5902. Washington, D.C.: World Bank.
- Lustig, Nora, George Gray-Molina, Sean Higgins, Miguel Jaramillo, Wilson Jiménez, Veronica Paz, Claudiney Pereira, Carola Pessino, John Scott and Ernesto Yañez. 2012. "The Impact of Taxes and Social Spending on Inequality and Poverty in Argentina, Bolivia, Brazil, Mexico and Peru: A Synthesis of Results." Tulane University Economics Working Paper 1216, New Orleans, Louisiana, April 2012.
- Lustig, Nora and Sean Higgins. 2012. "Fiscal Incidence, Fiscal Mobility and the Poor: A New Approach." Tulane Economics Department Working Paper 1202. New Orleans, Louisiana: Tulane University. April 2012. <http://ideas.repec.org/p/tul/wpaper/1202.html>
- Lustig, Nora. 2000. "Crises and the Poor: Socially Responsible Macroeconomics." *Economía: The Journal of the Latin American and Caribbean Economic Association* 1(Fall): 1-45. Washington DC: Brookings Institution Press.
- Morra Imas, Linda G. and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington DC: World Bank.
- O'Donnell, Owen, Eddy van Doorslaer, Adam Wagstaff and Magnus Lindelow. 2008. "Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation." WBI Learning Resources Series. Washington, DC: The World Bank.
- Pereira, Claudiney and Sean Higgins. 2012. "Fiscal Policy, Urban-Rural Inequality and Rural Poverty in Brazil." Report prepared for the International Fund for Agricultural Development.
- Plotnick, Robert. 1981. "A Measure of Horizontal Equity." *The Review of Economics and Statistics* 63(2), 283-288.
- Scott, John. 2011. "Gasto Público y Desarrollo Humano en México: Análisis de Incidencia y Equidad." Working Paper for *Informe sobre Desarrollo Humano México 2011*. Mexico City: UNDP.
- Shah, Anwar, ed. 2003. *Handbook on Public Sector Performance Reviews*. Washington DC: The World Bank.
- Silveira, Fernando Gaiger, Jhonatan Ferreira, Joana Mostafa and José Aparecido Carlos Ribeiro. 2011. "Qual o Impacto da Tributação e dos Gastos Públicos Sociais na Distribuição de Renda do Brasil? Observando os Dois Lados da Moeda." *Progressividade da Tributação e Desoneração da Folha de Pagamentos Elementos para Reflexão*. Eds. José Aparecido Carlos Ribeiro, Álvaro Luchiezi Jr., and Sérgio Eduardo Arbulu Mendonça. Brasília: IPEA. 25-63.

- Suits, Daniel B. 1997. "Measure of Tax Progressivity." *The American Economic Review* 67(4): 747-752.
- Urban, Ivica. 2009. "Kakwani decomposition of redistributive effect: Origins, critics and upgrades." ECINEQ Working Paper 2009-148.
- Urban, Ivica. 2012. "Implementation Issues in the Duclos-Jalber-Araar Decomposition of Redistributive Effect." *Public Finance Review* 00(0): 1-23.
- Van de Walle, Dominique and Kimberly Nead, eds. 1995. *Public Spending and the Poor: Theory and Evidence*. Baltimore and London: Published for the World Bank by John Hopkins University Press.
- Van Kerm, Philippe. 2009. "Income Mobility Profiles." *Economics Letters* 102(2): 93-95.
- World Bank. 2000/2001. "World Development Report 2000/2001: Attacking Poverty." Washington, DC and New York: Published for the World Bank by Oxford University Press.
- _____. 2006. "Country Policy and Institutional Assessments." *Operations Policy and Country Services*. Available online <http://siteresources.worldbank.org/IDA/Resources/CPIA2006Questionnaire.pdf>
- _____. 2009. "The World Bank's Country Policy and Institutional Assessment, An Evaluation." Independent Evaluation Group. Washington, DC: World Bank. Available online http://siteresources.worldbank.org/EXTCPIA/Resources/cpia_full.pdf
- _____. 2011. "CPIA Public Sector Management and Institutions Cluster Average." Available online <http://data.worldbank.org/indicator/IQ.CPA.PUBS.XQ>
- Yitzhaki, Schlomo and Edna Schechtman. Forthcoming. "The Gini equivalents of the covariance, the correlation and the regression coefficient." *The Gini Methodology: a primer on statistical methodology*. Springer: 35-48.