****

**Part IV- The Commitment to Equity Assessment Tools**

**[Commitment to Equity Handbook](http://www.commitmentoequity.org/handbook/)**

**[Estimating the Impact of Fiscal Policy on Inequality and Poverty](http://www.commitmentoequity.org/handbook/)**

**[Brookings Institution Press and CEQ Institute, Tulane University](http://www.commitmentoequity.org/handbook/)**

**[Nora Lustig (editor)](http://www.commitmentoequity.org/handbook/)**

March 1, 2018

**Data and Software Requirements**

**CEQ Team**

The Commitment to Equity (CEQ) Handbook is a step-by-step guide to applying the fiscal incidence analysis that is at the heart of CEQ Assessments. This chapter will present, in detail, the elements necessary for the completion of the first steps in a CEQ Assessment: information and data gathering, and software preparation.[[1]](#footnote-2)

**I. Data Requirements**

A CEQ Assessment requires three types of data: microeconomic, fiscal-administrative, and (some) macroeconomic. The necessary microeconomic data is typically available in a household consumption expenditure or income survey (“household survey”, or HS, for short) dataset. When the HS data is enumerated in, or recorded for, the same year as the fiscal-administrative data; and when it is representative at the national level (at least), fewer pre-analysis adjustments will need to be made.

The fiscal-administrative data provides information necessary for allocating the different fiscal interventions in appropriate or commensurate magnitudes to individuals and households represented in the household survey dataset. The fiscal-administrative data, when reliable, can also provide a check on whether household survey dataset records – on program eligibility, for example - are consistent with administrative records.

Macro-economic data is used as a third source of information to calibrate magnitudes for fiscal elements that are allocated across the individuals and households in the household survey. For example, when fiscal-administrative data together with macroeconomic data reveal that total VAT revenues were equivalent to 7 percent of total final household consumption expenditure, this ratio will be used to calibrate the amount of VAT allocated within the household survey.

The household survey dataset together with the fiscal-administrative data will be the foundation for allocating the fiscal interventions and for constructing the eight CEQ Core Income Concepts that are the key intermediate inputs into a CEQ Assessment; they are discussed in greater detail below.

**Microdata:**

The (preferably) nationally-representative household survey dataset can be an expenditure and income survey (HIES), a household income (employment) survey, or a household budget survey (HBS). Below the sections of the HS are discussed:

1. The household roster and the income and/or expenditures module - hopefully the official cleaned data, failing this, in the raw or semi-cleaned, item-by-item form are required.
2. The health and education modules are somewhere in between necessary and very desirable. When health and/or education are not covered, reference to a secondary survey that does capture utilization of those services (the Demographic and Health Surveys[[2]](#footnote-3), for example) is useful.
3. The remaining modules are often useful - taxpayer status can be determined from questions in the labor module, for example - and if they are available it would be good to have them.
4. If there are any official or generally accepted practices, methods, or algorithms for calculating the value of household expenditures, household size, per-adult equivalent scales, and the national poverty line, it is highly desirable to have those from the beginning.

**Administrative Data:**

A CEQ Assessment requires administrative data for the year of the analysis. Specifically, it requires three main components: 1) an audited or executed budget, number of beneficiaries/ tax payers, and the amount received or liable to be paid for each fiscal intervention to be included in the analysis; 2) a detailed description of each fiscal intervention; and 3) the audited or executed budget of the General Government, or, failing this, information for the Central Government.

1. An audited or executed budget together with the administratively-recorded number of beneficiaries (or tax-payers, where relevant) and administratively-recorded amounts received (or paid, where relevant) for each fiscal intervention are essentials. Disaggregation of those quantities by region, by sub-program, or by sector are also essential. In the case of education, for example, expenditures and enrollment should be broken down by level of schooling: pre-primary, primary, secondary and tertiary (at least). If any public spending ends up in the private system (civil servant and teachers, usually), then it would be desirable to have some way to estimate that spending. In the case of health, please provide whatever general breakdown of the spending is available: expenditures absorbed by hospitals versus clinics, or by inpatient services versus outpatient services[[3]](#footnote-4) hospitalized patients versus outpatients, or spending on wages versus goods and services.

With the fiscal-administrative data, the CEQ analyst can construct Table 4-1 columns B, D and F (shaded in blue). Sheet C1 of the CEQ Master Workbook (MWB) also asks for the information in Table 4-1 columns A through E, so starting a preliminary table outside of the MWB will cut down on MWB archival time later.

Table 4-1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fiscal Intervention (A) | Audited or executed budget (B) | Total amount according to the survey (C) | Number of beneficiaries/payers according to administrative accounts (D) | Number of beneficiaries/payers according to the survey (E) | Benefit according to administrative accounts (F). This applies only to direct transfers and subsidies. | Benefit reported in the survey if applicable (G). This applies only to direct transfers and subsidies. |
| Guardian Angel | 15,000,000 | 12,000,000 | 16,696 | 12,542 | 80 | 80 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

The Table 4-1 shows the example of a Program named Guardian Angel. The explanation for each column is found below.

* Fiscal Intervention: this refers to the official name of the fiscal intervention. For example, Guardian Angel.
* Audited or executed Budget: this column refers to how much was spent on the specific program or the total revenue produced by a specific tax during the year of the analysis in local currency units (LCU). For example, the Table 4-1 shows that the government spent 15 million on Guardian Angel in the year of the analysis. This information is generally published by the Ministry of Finance, the ministry or institution in charge of the program, or by the Tax Authority in the case of taxes or by the National Institute of Statistics.
* Total amount according to the survey: this column refers to the weighted sum of the amount of all individual benefits or tax payed in annual terms. For example, the Table 4-1 shows that the sum of all individual benefits is 12 million, i.e. 80 LCU monthly x 12,542 beneficiaries x 12 months. It is important to note that in this case we multiply by 12 months because the benefit is received by month and from January to December. Have in mind that this is not always the case. Some benefits, such as scholarships are received by 10 months instead of 12, for example.
* Number of beneficiaries/payers according to administrative accounts: this refers to the number of people or families that, according to the administrative records, appeared as receiving the benefit or paying the tax. This information is generally published or provided by the ministry or institution in charge of the program, by the tax authority in the case of taxes, the National Institute of Statistics, or the National Audit Office. For example, the Table 4-1 shows that according to the ministry in charge of Guardian Angel, the year of the analysis, 16,696 individuals were beneficiaries of the program.
* Number of beneficiaries/payers according to the survey: this column refers to the total number of individuals (using sample weights) who report receiving the benefit or paying taxes in the household survey. Alternatively, it can refer to the number of individuals designated as beneficiaries or tax payers in the survey. For example, 12,542 individuals report receiving benefits from Guardian Angel in the household survey.
* Benefit according to administrative accounts if applicable: this column refers to the benefit amount that each beneficiary should receive based on the program rules. For example, the rules of Guardian Angel establish that each beneficiary of the program should receive 80$ monthly.
* Benefit reported in the survey if applicable: this column refers to the benefit amount reported in the survey. Alternatively, it can refer to the benefit amount imputed in the survey. For example, every individual who reported being a beneficiary of Guardian Angel reported receiving a benefit of 80 monthly.

When administrative data is judged to be reliable and judged to reflect the *de facto*, on-the-ground delivery of the fiscal item in question, the information collected in this section will be useful for: A) identifying the fiscal interventions to be included in the analysis; B) assessing whether you can trust the information reported in the survey for a particular fiscal intervention; C) knowing how many beneficiaries need to be simulated in the survey for both cases when the survey does not report who are the beneficiaries and when the number of beneficiaries reported in the survey are fewer than the number of beneficiaries who really are receiving the benefit; and D) checking consistency between the survey and administrative data in order to assess if the allocation of a particular fiscal intervention needs to be adjusted.

1. The second component is detailed descriptions of each fiscal intervention. This information is necessary for the fiscal interventions that will be allocated by using the simulation method and desirable for those that will be allocated by applying any other allocation method. The information needed by category of fiscal intervention is the following:
   * 1. Direct taxes
        1. A list of existing tax regimes.
        2. Whether an individual can choose a specific tax regime or the individual is obliged to pay under a specific tax regime.
        3. The characteristics that must be met for someone to fall under a specific tax regime.
        4. Income tax rates and bands, and existing allowances by tax regime are necessities.
        5. Information on tax evasion, such as the characteristics of individuals who do not pay taxes. For example, one common characteristic of individuals who do not pay taxes is that they work in the informal sector.
        6. Estimation of Tax Expenditures.
     2. For social contributions:
        1. A list of existing social security systems.
        2. Whether an individual can choose a specific social security system or the individual is obliged to be affiliated with a specific social security system.
        3. The characteristics that must be met for someone to fall under a specific social security system.
        4. Income rates that affiliated must pay by service covered by the social security system. Contributions to pensions need to be estimated separately from health, unemployment, etc.
     3. For direct transfers and subsidies:
        1. A list of existing direct transfers and subsidies.
        2. The conditions that need to be met for someone to be beneficiary.
        3. Identify whether the beneficiary is an individual or a household.
        4. The targeting mechanisms used to select the beneficiaries.
        5. Amount of the benefit if it is fixed or how the amount is determined if it is variable.
        6. In the case of subsidies, it is important to know if it subsidizes the supply or the demand.
     4. Indirect taxes
        1. Value added and excises tax rates.
        2. Existing exemptions.
        3. Estimation of tax expenditures.
        4. Information on tax evasion.
        5. If the main survey used for constructing the CEQ Income concepts does not have information on consumption you will need a secondary source. Please see Chapter 6 of this handbook.
        6. Input-output table, SAM (Social Accounting Matrix), or SUT (Supply and Use table) for estimating indirect effects. For more details about Indirect Effects see Chapter 7 of this Handbook.
     5. Education
        1. Information on what type of schools receive resources from the government, if it is only public schools, semi private schools, and/or private schools.
        2. Information on co-pays.
        3. If the survey does not report who attends public schools, then individual characteristics that allow the identification of which individuals are more likely to attend a public school is necessary.
     6. Health
        1. Information on existing health systems.
        2. Information on what health systems receive resources from the government.
        3. Information on usage.
        4. Information on co-pays or other payments from households that are required to access public health services. Additionally, information on spending channeled through health insurance schemes, including the payments by households to participate in these schemes are desirable.
2. The third component is an audited or executed budget for the General Government, or, failing this, information for the Central Government[[4]](#footnote-5). Aggregated information about revenues and expenditures for the year of the analysis are necessary. It is important to make clear if this information refers to the General, Central or other government unit. The disaggregation of the information should be done as shown in Table 4-2. The information in columns of Table 4-2 is the same as the information asked for in sheet B3 of the Master Workbook Section ABC so this will save you time later on the completion of MWB ABC. It is advisable to compare the second column of the Table 4-2 to the Information published by the IMF in the Article IV Consultation Report of each country.

Table 4-2

|  |  |  |
| --- | --- | --- |
| **Government Revenue & Spending** | Currency Amounts in Administrative Accounts | Total (% of GDP) |
| **Total Revenue & Grants** |  |  |
| Revenue |  |  |
| Tax Revenue |  |  |
| Direct taxes *of which* |  |  |
| Personal Income Tax |  |  |
| Corporate Income Tax |  |  |
| Payroll Tax |  |  |
| Taxes on Property |  |  |
| Contributions to Social Insurance *of which* |  |  |
| From Employees |  |  |
| From Employers |  |  |
| From Self-Employed |  |  |
| Indirect Taxes *of which* |  |  |
| VAT |  |  |
| Sales Tax |  |  |
| Excise Taxes |  |  |
| Customs Duties |  |  |
| Taxes on Exports |  |  |
| Nontax Revenue |  |  |
| Grants |  |  |
| **Total Expenditure** |  |  |
| **Primary Government Spending** |  |  |
| Defence Spending |  |  |
| Social Spending |  |  |
| Social Protection |  |  |
| Social Assistance *of which* |  |  |
| Conditional or Unconditional Cash Transfers |  |  |
| Noncontributory Pensions |  |  |
| Near Cash Transfers (Food, School Uniforms, etc.) |  |  |
| Other |  |  |
| Social Insurance *of which* |  |  |
| Old-Age Pensions |  |  |
| Education *of which* |  |  |
| Pre-school |  |  |
| Primary |  |  |
| Secondary |  |  |
| Post-secondary non-tertiary |  |  |
| Tertiary |  |  |
| Health *of which* |  |  |
| Contributory |  |  |
| Noncontributory |  |  |
| Housing & Urban *of which* |  |  |
| Housing |  |  |
| Subsidies *of which* |  |  |
| Energy *of which* |  |  |
| Electricity |  |  |
| Fuel |  |  |
| Food |  |  |
| On Inputs for Agriculture |  |  |
| Infrastructure *of which* |  |  |
| Water & Sanitation |  |  |
| Rural Roads |  |  |
| Other Government Spending |  |  |
| Grants |  |  |

**II. Software Requirements**

In terms of software, the CEQ requires Stata 13 or newer and Microsoft Excel. You may also use another statistical package such as R or SPSS instead of Stata to construct the initial dataset --comprised of the fiscal interventions, the eight CEQ Core Income Concepts, and other variables. However, if you choose to do so, you will need to transform the final dataset to a file format suitable for Stata13 or newer to run the CEQ Stata package. Stata 13 (or newer) is required to run the CEQ Stata package which computes the CEQ Assessment analytical quantities, statistics, and indicators, and transfers them to the Master Workbooks Microsoft Excel files (MWBs). Please note that the CEQ Stata package contains a command named ceqgraph, which houses four subcommands (progressivity, conc, cdf, fi); each subcommand produce graphs that can be automatically transferred to their corresponding sheets of the MWB Section E if the subcommands are run using Stata 14.1 or newer. If you are using an older version of Stata, these graphs can be manually pasted into the MWBs.

A detailed description of the CEQ Stata Package can be found in either CEQ Handbook Chapter 8 or 9.

1. See also the introductory and first chapters of this Handbook for additional references to the steps covered in this chapter. [↑](#footnote-ref-2)
2. See <https://dhsprogram.com/data/> for more datasets available and more information on Demographic and Health Surveys. [↑](#footnote-ref-3)
3. Disaggregations by functional or economic classification are occasionally useful, but a disaggregation into sub-programs (from the point of view of a benefit recipient) is more desirable for CEQ Assessment purposes. [↑](#footnote-ref-4)
4. It is important to mention that, according to the System of National Accounts 2008 (2009), the General Government consists of central, state and local governments, together with the social security funds imposed and controlled by these units. In addition, it includes non-profit institutions engaged in non-market production that are controlled by government units or social security funds. [↑](#footnote-ref-5)