

Data Harmonization Brief

Harmonized Household Income and Expenditure Surveys (HHIES)

ECONOMIC
RESEARCH
FORUM



منتدى
البحوث
الاقتصادية

OPEN ACCESS MICRO DATA INITIATIVE (OAMDI)
*for the Arab Countries,
Iran and Turkey*

Data Harmonization Brief

Harmonized Household Income and Expenditure Surveys

(HHIES)

Introduction

The Economic Research Forum (ERF) has started the data harmonization project back in 2009 by securing a number of household income and expenditure surveys and harmonizing these data following its own set of harmonization standards guided by best practices pursued by other organizations working in the same field.

A first phase of the Open Access Micro Data Initiative (OAMDI) was launched in April 2013, to make micro data available to the data users, covering 17 Harmonized Household Income and Expenditure Surveys from Egypt, Palestine and Jordan, in addition to 4 Labor Market Panel Surveys from Egypt and Jordan, and data sets on Micro and Small Enterprises from Egypt, Lebanon, Turkey and Morocco.

ERF aiming at enriching household data disseminated through the first phase, has implemented a second phase of data harmonization/dissemination to include composite codes for essential variables considered, on the household as well as the individual level, following the Integrated Public Use Microdata Series (IPUMS) coding practices. The technique executed in the second phase builds on keeping the "standardized" version of the variable already generated in the first phase. Adding a "composite/detailed" version of some selected variables for which the coding was expanded to 3 or 4 digits to allow for more exhaustive responses.

What is New in the Second Phase?

As mentioned earlier, this phase consists of adding new/more detailed variables to the first phase of harmonization. The variables are constructed on the basis of composite codes. The aim of this expansion is to provide the researchers with a more comprehensive set of variables to avoid the loss of information given the heterogeneity of definitions between data sets across and within countries. This phase involves 12 and 10 household and individual characteristics, respectively. Detailed as follows:

Household level

Dwelling type [Composite/detailed codes]

Dwelling tenure [Composite/detailed codes]

Shared toilet facility [New]

Garbage disposal method [New]

More details for ownership of durables [New]

Marital status (Head) [Composite/detailed codes]

Nationality (Head/Spouse(s)) [Composite/detailed codes]

Highest level of education attained (Head/Spouse(s)) [Composite/detailed codes]

Main activity status (Head/Spouse(s)) [Composite/detailed codes]

Occupation (Head/Spouse(s)) [Composite/detailed codes]

Industry (Head/Spouse(s)) [Composite/detailed codes]

Sector of Employment (Head/Spouse(s)) [Composite/detailed codes]

Individual level

Relationship to head [Composite/detailed codes]

Marital status (Individual/Father/Mother) [Composite/detailed codes]

Nationality (Individual/Father/Mother) [Composite/detailed codes]

Highest level of education attained (Individual/Father/Mother) [Composite/detailed codes]

Received vocational training [New]

Type of vocational training received [New]

Main activity status (Individual/Father/Mother) [Composite/detailed codes]

Occupation (Individual/Father/Mother) [Composite/detailed codes]

Industry (Individual/Father/Mother) [Composite/detailed codes]

Sector of Employment (Individual/Father/Mother) [Composite/detailed codes]

Household Surveys Harmonization Procedures

The harmonization procedures of the household surveys consist of five long and tedious steps described as follows:

Step 1: Acquire

ERF staff negotiates with statistical agencies across the Arab countries, Iran and Turkey to acquire data sets. In addition to signing a memorandum of understanding that enables the use of data exclusively for teaching and scientific research purposes. Furthermore, it authorizes the publication of the harmonized data sets on the OAMDI portal. Acquiring micro data is typically a long process to develop a trust worthy relationship between the two parties nurtured with capacity building and technical assistance.

Step 2: Clean and harmonize

Once the data is acquired, the ERF data team starts by carefully checking the received data files and all provided relevant documentation; namely, code books, questionnaires, methodology, instructions documents, ...etc.

The data cleaning process then consists of identifying incorrect, inaccurate, or inconsistent parts of the data and modifying them. This is done in order to improve data quality and produce a cleaned/consistent data set for harmonization.

The type of errors checked, following the ERF template for data cleaning, are:

- Duplicated records
- Variable ranges
- Not applicable cases
- Inconsistent cases

When the cleaning process on the newly acquired data is achieved, the data team then harmonizes it into a common cross-national template, pre-defined by ERF, and guided by best practices pursued by other organizations working in the same field. The list of harmonized variables, including the name, label, content and formulas used to generate each variable, is defined in the data dictionary published with each dataset.

Step 3: Re-check after harmonization

ERF data team double checks the quality of the harmonized data produced through a post-harmonization check and cleaning process, to ensure a good quality and reliability of the ERF produced data. All corrections and modifications processed for cleaning purposes are documented for each data set.

The post harmonization cleaning process involves checking the quality of all harmonized variables generated. A consistency check is also performed within variables and across survey years.

Step 4: Document

A comprehensive documentation is created for each data set using NESSTAR (the Software System for data publishing adopted by ERF), by recording all survey information and data characteristics, and attaching all documentation materials provided by the Statistical Agency or produced by ERF.

Step 5: Disseminate

The last step involves uploading micro data in both SPSS and STATA formats with all relevant documentation on its data portal. The goal is to make it available to researchers upon a quick and simple registration process, aiming at controlling the access to the micro data while protecting its confidentiality.

To encourage the responsible use of micro data, ERF acknowledges the use of data for scholarly research, or educational purposes only, and prohibits access if the data is used in the pursuit of any commercial or private ventures.

Documentation Materials Description

A set of documentation materials is published with each survey in PDF/XLS format, subject to availability, in all available languages. This set includes all documentation available for the original survey provided by the Statistical Agency, and for the harmonized data sets produced by ERF. In addition to a copy of all international classifications of expenditures, occupations and economic activities used during the harmonization process.

Materials produced by ERF and available for all surveys

- List of harmonized variables

The list of harmonized variables is a documentation material produced by ERF, in XLS format, presenting the full standardized list of the household-level and the individual-level harmonized variables.

- Data dictionary

The data dictionary is a documentation table produced by ERF for each data set. It presents the ERF set of harmonized variables and harmonization standards/concepts on the household as well as the individual levels.

The table's fields indicate the variables' names and labels, as appearing in the data file, a description of the content of each variable, general and country specific comments, and the formula used to generate the variable if computed by ERF.

In addition, the data dictionary provides a variable by variable mapping between the original and the standardized values to indicate how each variable was generated, followed by a column illustrating the variable universe and a key column denoting the availability of the variable in the harmonized data set.

- Expenditure aggregates

The expenditure aggregates is a documentation table produced by ERF for each data set. It presents a list of products' (expenditure items) codes/labels categorized under each major group of the Classification of Individual Consumption According to Purpose (COICOP). It also illustrates all general comments on different expenditure groups and formulas used to compute the expenditure aggregates.

Materials provided by the statistical offices

This set of resources varies among and within countries depending on the available information at the statistical agency. The ERF data staff is typically working to provide all the following documentation in Arabic, English, and/or French.

- Survey Questionnaire(s)
- Survey Coding Book
- Survey Methodology
- Survey Manual/Instructions Manual
- Survey User Guide

International Standard Classifications

Relevant documentation of various classifications standards are made available whenever needed. They mainly include the following:

- Classification of Individual Consumption According to Purpose (COICOP)

This is the international classification used to compose expenditure aggregates in all household surveys. It is published with each survey to serve as a guide on how expenditure aggregates were grouped and computed.

- International Standard Classification of Occupations (ISCO)

In most household surveys, the occupation classification follows the 1988 or 2008 ISCO. Whenever, this is the case, a document indicating the 2, 3 and 4 digits coding of the relevant year of occupation classification used in the survey data is available in the reference documents section.

- International Standard Industrial Classification of All Economic Activities (ISIC)

In many household surveys, the economic activity classification follows the Rev.3, Rev.3.1 or Rev.4 of the ISIC. Whenever, ISIC is used, a document indicating the 2, 3 and 4 digits coding of the economic activity classification version used in the survey data is available in the reference documents section.