Sampling Design and Survey Weights

Impact of Russian Ukrainian War on Households in Egypt and Kenya 2024

Impact of Russian Ukrainian War on Households in Egypt and Kenya

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

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

Impact of Russian Ukrainian War on Households in Egypt & Kenya:
Sampling design and survey weights

Version: May, 2024

Summary

In this report, we outline the sampling methods, response rates, and weighting methodology employed for the Impact of Russian Ukrainian War on Households. Applying these weights in all analyses is crucial to ensure our sample's characteristics accurately reflect the target population within the country. While weighting helps address imbalances, it's important to acknowledge that it cannot fully account for potential biases arising from unobservable dimensions of non-response.

Sample

Impact of Russian Ukrainian War on Households survey targeted mobile phone users between the ages of 18 and 64. To achieve a representative sample, we employed random digit dialing (RDD) across valid mobile numbers. We made up to three attempts to reach individuals for the survey if the initial call was unanswered, disconnected, busy, or if the respondent requested a follow-up. Samples were stratified by country-specific market shares of mobile operators.

Response rates

Table 1 presents responses and response rates for both Egypt and Kenya. It's important to note that our response rates exclude phones out of service, those disconnected or busy after repeated attempts, and individuals' ineligible for the survey. Responses reflect successful completions, regardless of whether they occurred on the first, second, or third attempt.

	Kenya		Egypt	
	Freq.	Percent	Freq.	Percent
Phone disconnected/ busy	295	2.06	566	2.94
Not in service	4384	30.59	6531	33.96
Did not answer	1192	8.32	1274	6.62
Picked up and refused	4070	28.4	6523	33.92
Incomplete and refused	1384	9.66	1297	6.74
Incomplete and call returned	63	0.44	104	0.54
Complete	2030	14.17	2002	10.41
Not eligible	912	6.36	934	4.86
Total	14330	100	19231	100
Response Rate		23		18

Initial Weights

To reduce bias in a number of observable dimensions, inverse probability weighting was applied. The weights were created at two levels: Individual and household. The weights had the following inputs:

- Phone operators and their market shares provided by the data collection company

- Number of phones by operator for individuals (weighting for individuals) and household members (weighting for households).
- Representative data with comparable demographic and household characteristics for weighting non-response.

Denote individuals as *i* (ranging from 1 to N) and households as *h* (from 1 to N). Denote the number of phones of a particular operator, *o*, as t_o (operators in the range from 1 to M). Denote the total number of phones that should have been in the sample of *o*, given the total number of observed phones and market shares, as T_o.

We then generated initial market-share individual weights as:

$$w_i = 1/\sum_1^M [(\{\sum_1^N t_{o,i}\}/T_o) * t_{o,i}]$$

With these individual weights, we then pooled the phone surveys with representative in-person surveys and used a probit model weighted with survey weights (for the representative survey) and w_i (for the Impact of Russian Ukrainian War on Households data) to estimate the probability that an individual with certain characteristics is included in the phone survey data. [1]The predicted probability from that model, p_i, was used to generate individual weights for the Impact of Russian Ukrainian War data on households:

$$w'_i = w_i * (1 - p_i)/p_i$$

We likewise generated initial market-share household weights as:

$$w_h = 1/\sum_1^M [(\{\sum_1^N t_{o,h}\}/T_o) * t_{o,h}]$$

This takes into account the number of phones in the household for all members, so that the probability of a sample is higher in a household with more members or more phones. The predicted probability from the individual level model was combined with the household market share weight to generate a household weight as follows:

$$w'_h = w_h * (1 - p_i)/p_i$$

The household and individual weights were all winsorized at the 95th percentile to ensure that no outlier weight influenced the statistics. The weights were then normalized by dividing them by the mean weight.

General Notes:

- Individual weights should be used for all analyses where the outcome is at the individual level. For household-level results (e.g. household income), the weighting used to generalize to households (e.g. X% of households are food insecure).
- the same procedure follows the weight calculation for the ERF COVID-19 HH MENA Monitor sample weight.

[1] In the case of Kenya, there was not a publicly available household survey with individual phone ownership data, so we had to use the household level phone ownership instead

Representative in-person national surveys

The representative in-person national survey samples used to generate weights were as follows:

- Egypt: Egypt Labor Market Panel Survey 2018 (OAMDI, 2019).
- Kenya: Kenya Continuous Household Survey (KCHS) – 2021 (Kenya National Bureau of Statistics, 2023)

These were selected as the most recent publicly available data with individual phone ownership[2] and relevant demographic and labor market characteristics.

Table 2 below shows the covariates included in the weighting models by country.

Covariate	Egypt	Kenya
Sex	X	
Age group	X	
Education level	X	
Household size (categorically)	X	X
Labor mkt. status	X	
Administrative geography	Region	Region
Urban v. rural		X
Marital status	X	
Gender of the head of household		X
Presence of kids 0-5	X	X
Presence of kids in school	X	X
Percentage of females in the household		X

References

1- OAMDI, 2019. Labor Market Panel Surveys (LMPS), http://erf.org.eg/data-portal/. Version 2.0 of Licensed Data Files; ELMPS 2018. Egypt: Economic Research Forum (ERF).

2- **Kenya National Bureau of Statistics.** Kenya Continuous Household Survey (KCHS) - 2021. https://statistics.knbs.or.ke/nada/index.php/catalog/123/related-materials

3- **Krafft, C.** (2021, November 17). ERF COVID-19 MENA Monitor sampling, response rates, and weights [Report]. ERF Data Portal. http://www.erfdataportal.com/index.php/catalog/CMMHH

[2] In the case of Kenya, there was not a publicly available household survey with individual phone ownership data, so we had to use the household level phone ownership instead