FINMARK TRUST
Making financial markets work for the poor

COVID-19 tracking survey
Methodology

March 2020

Authors

Bobby Berkowitz and Shirley Jeoffreys-Leach

About FinMark Trust

FinMark Trust is an independent non-profit trust whose purpose is 'Making financial markets work for the poor, by promoting financial inclusion and regional financial integration'. We pursue our core objective of making financial markets work for the poor through two principle programmes. The first is through the creation and analysis of financial services consumer data to provide in depth insights on both served and unserved consumers across the developing world. The second is through systematic financial sector inclusion and deepening programs to overcome regulatory, supplier and other market level barriers hampering the effective provision of services. Together, these programmes unlock financial inclusion and sector development through a symbiotic relationship between rigorous data collection and research activities. Our work can be found in South Africa, throughout the SADC region and the global arena.

For more information:

Visit our website at www.finmark.org.za

Email Grant Robertson (Head Data and Analytics) on grantr@finmark.org.za

Call us on +27 11 315 9197

CONTENTS

1. OVERVIEW

This cross-country tracker measures the impact of COVID-19 on peoples' healthcare access and behaviour, earnings and finances, food security, and fears relating to their personal safety aims to bridge the gap for African policymakers. The reason for this is that the tracker can pinpoint what the virus means for the man or woman on the street. We are rapidly collecting information that can capture the impact of the virus on livelihoods and wellbeing. The evidence collected can be used to help policymakers, the private sector, and civil society know how best to respond to the pandemic in order to protect their citizens and economies.

The project was launched on 8 April 2020 and data collection is being repeated in seven markets five times, in approximately two to three week surveying intervals. This allows us to monitor trends in the mindsets, livelihoods, and coping strategies of people as they try to manage the virus and its impact over periods of time.

The sample per data collection period per market is:

Country	08.04 – 28.04	29.04 – 08.05	25.05 –	TBC	TBC	TBC	TBC	TBC
Ghana	-	-	-	~1000	~1000	~1000	~1000	~1000
Kenya	1034	1421	~1000	~1000	~1000	-	-	-
Nigeria	1806	2359	~1500	~1500	~1500	-	-	-
Rwanda	1134	1394	~1000	~1000	~1000	-	-	-
South Africa	1057	1099	~1000	~1000	~1000	-	-	-
Uganda	-	1078	~1000	~1000	~1000	~1000	-	-
Zambia	-	-	~1000	~1000	~1000	~1000	~1000	-

Table 1: Country-wave samples

2. DATA COLLECTION AND SAMPLING METHODOLOGY

Data is being collected using mobile computer-aided telephonic interviews (mCATI), in which call centre agents call and interview people. While face-to-face surveys typically provide a better representation of the overall population, using telephonic interviewing is recommended because of two critical factors:

1. In order to flatten the curve of infections, social responsibility requires that social distancing becomes a way of life as the virus spreads. Telephonic surveys can be conducted under social distancing protocols.
2. Data collection using mobile phones can be collected and reported on more quickly than traditional face-to-face in-person surveys. They do not require the logistical planning and time involved in getting interviewers out into the field.

A limitation of using a mobile method is the exclusion of people who do not have access to a mobile phone. There is a risk that the sample will not include many people who are at the very bottom of the pyramid. While these people are under-represented, there are people in this cohort of interest who do own and use mobile phones and the method can, therefore, reach them.

In order to account for the under-representation of poor people, particularly women, we are using a combination of random digit dialling and quota sampling. The purpose of this is to get a wide spread of people to boost the representation of under-represented cohorts – at about an 80/20 ratio for the sample.

The sampling approach has been improved for each new data collection period. For the first time period of the tracker (08.04 – 28.04), we used a quota sampling approach using a demographically profiled database of people in these markets. Quotas were set on gender, age, urbanicity, and region.

The databases used represent only a small portion of the potential universe of mobile phone owners and the resulting sample of the first time period of data collection in Nigeria, South Africa, and Rwanda skewed strongly to urban, middle-class people with higher levels of education and, therefore, required extreme weighting.

A further limitation in the first time period was driven by our decision to rush into field as COVID-19 was just beginning to spread in these African markets. In order to gather data as soon as possible, we opted to go ahead with our first time period of data collection in English only.

For the second time period, we changed the sampling strategy to collect approximately 80% of the sample with a random digit dialling approach. This broadens the universe well beyond the profiled database and we also included translation of the survey into local languages.

Overall, the second time period of data collection produced a better sample. However, it does still skew to people with higher levels of education and for the third wave we have, therefore, included education level as a quota variable.

Overall, the second time period of data collection produced a better sample. However, it does still skew to people with higher levels of education and for the third wave we have, therefore, included education level as a quota variable.

3. MODELLING AND WEIGHTING APPROACHES

3.1 Overview of statistical adjustment and inference

Traditional design-based survey analysis requires knowledge of the population as well as a level of oversight and control over how the population will be selected. These design decisions are used to calculate probabilities of selection and, therefore, weights used to represent the population. With the design-based approach, at least theoretically, every part of a population would need to be accommodated in the sample frame and the probability for selecting anyone would need to be calculated. For telephone interviewing, this would require describing what each telephone number represents according to descriptive demographic variables. Knowing the probability, at targeted population level, that that number will be randomly selected in accordance with these simultaneous demographic descriptions will also be required.

That is not the case with our samples. For every number dialled, we cannot calculate the probability that that person will be randomly selected in the population before we interview them. This indicates that we do not know how to appropriately structure the phone numbers of our research suppliers beforehand to create a design-based weighting approach. Samples always contain some biases which need to be corrected. This study's requirement for remote interviewing means that biases will be even more pronounced and will require greater correction. This is because, across these countries, the population who answer telephone surveys skews towards urban, male, younger, and middle-income populations. These skews differ both across countries and time.

Because design-based inference is not possible in our context, we implement two model-based alternatives to obtain population inferences from samples. These are raking, and multilevel regression with poststratification (MRP). In the sections below, we describe each procedure and how it was implemented in the context of these surveys. Before doing so, we first describe the reference data that underpins each procedure.

3.2 Reference data

A necessary component for an accurate statistical adjustment is to have a clear picture of the composition of the target (national) population, through a reference dataset. This dataset will ideally come from a recent credible study, preferably a country's census or large-scale household survey administered by a country's national statistical authority.

A reference demographic dataset is useful when it:

- contains relevant variables to stratify the population against;
- is accurate;
- has a large and well-distributed sample across the stratification variables.

To provide for relative consistency across markets, we have used the Demographic and Health Surveys (DHS) as our reference datasets, as these datasets are available and coded similarly across each market.

The COVID survey contains two types of outcome indicators; those that refer to the individual respondent, and those that refer to the respondent's household. The variables used for the statistical adjustment will differ between the two types of indicators. Table 3 describes the variables that are used to adjust both the individual-level and the household-level indicators:

Individual-level adjustment variables	Household-level adjustment variables
Region	Region
Urban/rural	Urban/rural
Language	Language
Age category	Household size
Gender	Number of children in the household
Education	Number of elderly in the household
Relationship status	

Table 2: Variables used for statistical adjustment in each market

We also include two additional adjustment variables that are meant to reflect household assets at both the individual and household level. These variables differ for each market, as listed in Table 3:

Market	Household asset variables
Kenya	Primary cooking fuel; television ownership
Nigeria	Primary cooking fuel; television ownership
Rwanda	Primary cooking fuel; primary water source
South Africa	Type of toilet
Uganda	Primary cooking fuel; type of toilet
Zambia	Primary cooking fuel; television ownership

Table 3: Market-specific variables used for statistical adjustment

All of the variables listed above are used to perform survey adjustment with both procedures (raking and MRP). In the next sections, we describe these procedures.

3.3 Raking

Raking, also known as iterative proportional fitting, is probably the most common model-based approach for survey adjustment. The procedure generates a set of weights, one for each survey respondent, so that the weighted survey reflects the reference population as closely as possible on key demographic characteristics. For example, once weights are generated the weighted proportion of women in the sample should be similar to their distribution in the target population. The weighted proportion of people that fall into each education category in the sample should be similar to the distribution in the target population, and so on.

Raking, also known as iterative proportional fitting, is probably the most common model-based approach for survey adjustment.

3.3.1. Trimming and the bias/variance trade-off

With unconstrained raking, certain observations may be assigned a weight that is abnormally high or low. If an observation receives a very high weight, our reported results will be highly dependent on this single observation. This is a situation we want to avoid because it leads to less stable results. In order to mitigate this problem, we implement a weight trimming procedure for each raked dataset.

The maximum allowable weight is chosen separately for each dataset, in order to find a balance between the accuracy of the weighting procedure (bias), and the dependence on a small number of high-weight observations (variance). More restrictive weights (stronger trimming) will reduce variance by making the results less dependent on individual observations. This also has the potential to increase bias because the weighted dataset is less reflective of the target population. For datasets that were relatively representative of the target population before weighting, we used more restrictive weights to reduce variance because bias is less of a concern. However, for samples that are less representative prior to weighting, we had to allow larger weights (and more variance) in order to reduce the bias to acceptable levels. We generally tried to keep the relative bias for all poststratification variables (defined as the mean for the variable in the target population). These were divided by the weighted mean of that variable in the sample) to below 10% although; in some of the earlier datasets, we did not rigidly stick to this rule.

Table 4 lists the maximum allowable weight that was used in the individual-level raking for each survey wave. As mentioned above, larger weights reflect a need to reduce bias to acceptable levels for samples that are less representative of the target population:

		Individual weight diagnostics		
		Maximum weight	Maximum relative bias	Design effect
Kenya	W1	15	1.05	4.38
	W2	15	1.06	4.27
	W3	25	1.05	4.99
Nigeria	W2	20	1.19	7.38
	W3	30	1.17	6.79
Rwanda	W2	20	1.02	4.76
	W3	20	1.02	5.22
South Africa	W2	20	1.00	2.83
	W3	40	1.05	7.18
Uganda	W1	25	1.07	10.38
	W2	40	1.12	13.6
Zambia	W1	15	1.00	2.46

Table 4: Rake weighting diagnostics

In this table, the maximum relative bias reflects how well the raked weights are able to adjust for differences in the survey adjustment variables. A maximum relative bias of 1 indicates that the weighted survey matches the target population exactly on all of these variables. The design effect summarises the influence of unequal weighting on the uncertainty of the resulting estimates. A design effect of 4 means that estimates using the weights have four times as much variance as unweighted estimates.

3.3.2. Properties of raking estimates

Raking is known to produce accurate estimates at the level of the target population. An advantage of raking is that we only need to produce two sets of weights; one for adjusting individual-level variables, and one for adjusting household-level variables. The appropriate set of weights can be used in the adjustment of all individual-level or household-level indicators.

Nonetheless, raking has some limitations. It has been shown to lead to biased estimates of outcomes among subgroups and for this reason, we do not recommend raking for subgroup-specific estimates. In addition, the uncertainty intervals around estimates based on raked weights do not typically reflect the uncertainty in the generation of the weights themselves. They, therefore, tend to overstate the accuracy of the estimates.

MRP is an alternative model-based method for obtaining estimates for a target population based on a non-representative sample. Although it is relatively new method when compared with raking, this method has gained popularity over the past 10-15 years.

3.4 Multilevel regression and poststratification (MRP)

MRP is an alternative model-based method for obtaining estimates for a target population based on a non-representative sample. Although it is relatively new method when compared with raking, this method has gained popularity over the past 10-15 years.

MRP begins by selecting a set of variables to use for statistical adjustment (Tables 2 and 3), and then subdividing the population into 'cells', based on their precise combination of these variables. The goal of MRP is to estimate the mean outcome in each of these cells (using a multilevel regression model), and then to take a weighted average of these cell-specific estimates according to how frequently each cell occurs in the target population. This is a procedure known as poststratification.

The model specification for the multilevel regression model is flexible; we fit a model that includes two-way interactions between all adjustment variables. The model is fit as a Bayesian hierarchical model. This model includes prior specifications that promote borrowing of strength between related cells. By borrowing information from other cells that have similar values of the adjustment variables, this allows us to obtain accurate estimates for cells that have little or no observed sample,.

MRP tends to provide similar estimates to raking at the level of the target population and provides better estimates for subgroups. It also provides more realistic uncertainty intervals, which take into account both the sampling error as well as the uncertainty in the statistical adjustment. However, its primary limitation is computational complexity; a separate model needs to be fit for each outcome of interest. Each of these models is more computationally complex than the simple raking model. Thus, it requires time and expertise for the setting up, testing, and running of each model.

3.5 Raking, MRP, and how to interpret the data

Given our extremely rapid turnaround time, we are using raking for national level statistics and larger disaggregated groups for analysis, e.g. gender. The data that is currently available on the dashboard (https://www.covid19tracker.africa/) is rake-weighted.

In addition, we are running MRP to validate the rake weighting results with some lag time, and MRP can be used for more detailed subgroup analysis. MRP results are not as yet available to the public.

.

Overall, raking and MRP produce similar results. However, MRP can adjust more easily to cross-sections between poststratification variables, while raking can have problems as a result of limited sample size. MRP does this through regularisation, i.e. borrowing strength across similar cells.

One strength of MRP is that standard error estimates on an MRP model take into account the uncertainty in the adjustment to the target population, whereas raking estimates take the weights as fixed. This makes a big difference in cases where the survey population varies greatly from the target.

One strength of MRP is that standard error estimates on an MRP model take into account the uncertainty in the adjustment to the target population, whereas raking estimates take the weights as fixed. This makes a big difference in cases where the survey population varies greatly from the target.

What this means for analysis is that with rake weighted data, subgroup analysis is more limited by the underlying sample. We recommend using the raked data to understand trends and differences over time but to treat the absolute percentages with some caution.

Another important element to note for analysis is that data of this nature is typically reported with estimations that state the results are accurate to a particular confidence (using frequentist statistical approaches) or credible (using Bayesian statistical approaches) level. Any level can be chosen but estimation intervals at 95% level are most common. It is common to see reporting referring to a 3% or 5% interval, as if this is standard across all reported estimates.

However, when weighting the data using raking or MRP, intervals simultaneously differ for every indicator, for different subgroups, and for responses in different percentiles. Statistical software packages can calculate and report on standard deviations and standard errors for rake-weighted data. Note that the assumptions underpinning these estimates and their associated reporting intervals are more limited than the MRP-derived equivalents we calculate and report. Therefore, users calculating their own estimation intervals should be cautious.

When using the rake weighted data for subgroup analysis, bigger estimation intervals are expected as levels of disaggregation increase. Our caution above is even more acute and we emphasise circumspection when reporting estimates for the rake-weighted data at sub-group levels. This is particularly the case for subgroups that disaggregate the overall sample into more than two sub-groups, or if a particular sub-group does not represent a large portion of the overall sample.

FINMARK TRUST

COVID-19 tracking survey | Methodology

Sanofi House, Second Floor,
44 on Grand Central Office Park,
2 Bond Street, Grand Central
Ext 1, Midrand

Tel: +27 11 315 9197
Fax: +27 86 518 3579
info@finmark.org.za
www.finmark.org.za