Reall Household Income and Affordability Calculator

Methodology Paper: Uganda 2020

This paper lays out the methodology used to develop the 2020 Uganda data for Reall's Household Income and Affordability Calculator, which can be found at www.reall.net/calculator.

The Uganda National Panel Survey (UNPS) is produced by the Uganda Bureau of Statistics (UBS). UNPS is a nationally-representative household survey focused on presenting income and consumption expenditure data, with the current round covering 3,123 households. The full published report and tables provide a broad overview of the landscape, with the raw data freely available from the same page for more detailed analysis.[1]

Data is split into separate data files for different sections of the survey. Key sections of the UNPS survey used for this work are 'pov2019_20', which provides data on household sizes and monthly expenditure; 'gsec1', which provides district names; and 'gsec8', which provides employment information. Names and coding for Uganda's Subregions can be found on the Data Description tab under pov2019_20, [subreg].[2]

Methodology

The following actions were undertaken on these datasets:

Creating Dataset

- uganda_regions
  - Create new 'uganda_regions' dataset using subregion names from the survey webpage's Data Description tab (see above) – [subreg] and [subreg_value]
- gsec1
  - Create new [urban_rural] column, with a value of 'Rural' if [urban] is 0, or 'Urban' otherwise
  - Replace [region_value] codes with region names, where:
    - 1 is 'Central'
    - 2 is 'Eastern'
    - 3 is 'Northern'
    - 4 is 'Western'
- pov2019_20
  - Create as 'merged_dataset'
  - Merge [distict], [wgt] (weight), [urban_rural] and [region_value] columns from gsec1 by [hhid] (household ID)
  - Merge [subreg_value] (names of Uganda's subregions) from 'uganda_regions' by [subreg] (subregion codes)

[1] https://microdata.worldbank.org/index.php/catalog/3902/study-description
[2] https://microdata.worldbank.org/index.php/catalog/3902/study-description

    - Rename [nrrexp30] ('Monthly household expenditures in nominal prices') as [HH_exp]
    - Rename [hsize] ('Count of usual members present/absent' as [HH_size]
- gsec8
    - Create new [is_earner] column, with a value of 1 if [s8q14] ('Among the answers to q4,6,8,10 and 12, is there a YES?') is 1, or 0 otherwise
        - For reference:
            - Q4 – 'Did [NAME] work for wage, salary, commission or any payment in kind?'
            - Q6 – 'Did [NAME] run a business, of any size, for themself?'
            - Q8 – 'Did [NAME] help without being paid in any household business?'
            - Q10 – 'Was [NAME] an apprentice?'
            - Q12 – 'Did [NAME] work on this household's farm?'
        - Group dataset by [hhid]
        - Create new [no_of_earners] by summing [is_earner] for each household

```
# Data processing and merging
gsec1_dataset <- gsec1_dataset %>%
        mutate(urban_rural = ifelse(urban == 0, "Rural", "Urban"),
                region_value = case_when(
                region == 1 ~ "Central",
                region == 2 ~ "Eastern",
                region == 3 ~ "Northern",
                region == 4 ~ "Western",
                TRUE ~ "Unknown"
                ))

housing_expenses <- gsec15c_dataset %>%
        filter(CEC02 %in% c(301)) %>%
        group_by(hhid) %>%
        summarise(Housing_Expenditure = sum(CEC05, na.rm = TRUE))

merged_dataset <- pov19_20_dataset %>%
        left_join(gsec1_dataset[c('hhid', 'district', 'wgt', "urban_rural", "region_value")], by =
        "hhid") %>%
        left_join(sub_region_codes, by = "subreg") %>%
        left_join(housing_expenses, by = "hhid") %>%
        rename(HH_exp = nrrexp30, HH_size = hsize)

# Calculate the number of earners per household
gsec8_dataset <- gsec8_dataset %>%
        mutate(is_earner = ifelse(s8q14 == 1, 1, 0)) %>%
        group_by(hhid) %>%
        mutate(no_of_earners = sum(is_earner)) %>%
        ungroup()

        left_join(gsec8_dataset[c('hhid', 'no_of_earners')], by = "hhid") %>%
        distinct() %>%
```

Calculating Sample Sizes

- Group dataset by [subreg_value] and [region_value] to create sample sizes for each subregion
- Group dataset by [subreg_value], [region_value] and [urban_rural] to create separate urban and rural sample sizes for each subregion

```
sample_size_uganda <- merged_dataset %>%
        group_by(region_value, urban_rural, subreg_value, Year = 2020) %>%
        summarise(sample_size = n())

sample_size_uganda_u_r <- merged_dataset %>%
        group_by(region_value, subreg_value, urban_rural = "All", Year = 2020) %>%
        summarise(sample_size = n())
```

Calculating Percentiles

- Multiply each [hhid] record by [wgt] to create full weighted dataset
- Group data by [subreg_value] and [region_value], and extracting data for records at 1% increments, creating figures for each percentile of every district
- Repeat the step above but also grouping by [urban_rural], creating separate urban and rural percentile figures for each district

```
# Function to calculate quantiles and related statistics
calculate_quantiles <- function(data, quantiles) {
        do.call(rbind, lapply(quantiles, function(q) {
                data.frame(
                        Quantile = q * 100,  # Convert quantile probability to percentage
                        HH_exp = quantile(data$HH_exp, probs = q, na.rm = TRUE)
                        )
                }))
        }

all_data_uganda <- merged_dataset %>%
        uncount(weights = as.integer(wgt/100))

# Calculate quantiles for each location
uganda_location_quantiles <- all_data_uganda %>%
        group_by(Country = "Uganda", urban_rural, Location = region_value,City =
        subreg_value, Year = 2020) %>%
        group_modify(~ calculate_quantiles(.x, quantile_probs)) %>%
        ungroup()

# Additional mixed urban/rural quantiles for each location
quantile_probs <- seq(0.01, 0.99, by = 0.01)
uganda_location_quantiles_u_r <- all_data_uganda %>%
```

```
    group_by(Country = "Uganda", urban_rural = "All", Location = region_value, City =
    subreg_value, Year = 2020) %>%
    group_modify(~ calculate_quantiles(.x, quantile_probs)) %>%
    ungroup()
```

Creating Final Dataset

- Define and align common columns to enable merging of quantiles datasets
- Combine relevant 'quantiles' and 'quantiles_u_r' datasets into a single 'summary_dataset'
- Define and align common columns to enable merging of sample size datasets
- Combine relevant 'sample_size' datasets into a single 'combined_sample_size'
- Join 'combined_sample_size' dataset to 'summary_dataset'
- Create a 'state_aggregated' version of 'summary_dataset' by grouping all household data by state
- Create a 'national_aggregated' version of 'summary_dataset' by grouping all household data by country
- Join 'state_aggregated' and 'national_aggregated' datasets to 'summary_dataset'

```
# Define common columns for final summary
common_columns <- c("Country", "urban_rural", "Location", "City", "Year", "Quantile",
"HH_exp", "no_of_earners", "HH_size", "Percent_Income_Spent_on_Housing")

# Function to align columns across datasets
align_columns <- function(df, common_cols) {
    df %>%
    mutate(across(setdiff(common_cols, colnames(df)), ~ NA)) %>%
    select(all_of(common_cols))
    }

# Align columns and combine all datasets
Uganda_location_quantiles <- align_columns(Uganda_location_quantiles,
common_columns)
Uganda_location_quantiles_u_r <- align_columns(Uganda_location_quantiles_u_r,
common_columns)

# Combine all country datasets into a single summary dataset
summary_dataset <- bind_rows(Uganda_location_quantiles, Uganda_location_quantiles_u_r)

sample_size_Uganda <- sample_size_Uganda %>%
    rename(Location = state, sample_size = sample_size) %>%
    mutate(Country = "Uganda")

sample_size_Uganda_u_r <- sample_size_Uganda_u_r %>%
    rename(Location = state, sample_size = sample_size) %>%
    mutate(Country = "Uganda")

# Combine the sample size tables into one
```

```
combined_sample_size <- bind_rows(
        sample_size_Uganda,
        sample_size_Uganda_u_r,

summary_dataset <- summary_dataset %>%
        left_join(combined_sample_size, by = c ("Country", "urban_rural", "Location", "City",
        "Year"))

# Aggregate data for state level
state_aggregated <- summary_dataset %>%
        group_by(Country, Location, Year, Quantile,urban_rural) %>%
        summarise(
                sample_size = sum(sample_size),
                HH_exp = mean(HH_exp, na.rm = TRUE),
                no_of_earners = mean(no_of_earners, na.rm = TRUE),
                HH_size = mean(HH_size, na.rm = TRUE),
                ) %>%
        ungroup()%>%
        mutate(City = "All")

# Aggregate data for the national level by combining all states within each country
national_aggregated <- summary_dataset %>%
        group_by(Country, Year, Quantile,urban_rural) %>%
        summarise(
                sample_size = sum(sample_size),
                HH_exp = mean(HH_exp, na.rm = TRUE),
                no_of_earners = mean(no_of_earners, na.rm = TRUE),
                HH_size = mean(HH_size, na.rm = TRUE)
                ) %>%
        ungroup() %>%
        mutate(Location = "All", City = "All")

# Combine both the state-level and national-level data
final_dataset <- bind_rows(state_aggregated, national_aggregated,summary_dataset)
```

Calculating Inflation

All data is inflated using median annual inflation rates from 2010-23. Median rates were used rather than actual figures to help compensate for large-scale inflation across many economies in 2022 and 2023.

Consumer Price Index inflation rates were taken from the World Bank[3] and consisted of the following figures, creating a final median rate of 5.13%.

CPI	Uganda
2000	3.392022
2001	1.865125

[3] https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG

2002	-0.28751
2003	8.680477
2004	3.721287
2005	8.448726
2006	7.310676
2007	6.138511
2008	12.05086
2009	13.01726
2010	3.976553
2011	15.12515
2012	12.67874
2013	4.902714
2014	3.074678
2015	5.410004
2016	5.445759
2017	5.640932
2018	2.623975
2019	2.869094
2020	3.794423
2021	2.204572
2022	7.195789
2023	5.350948