Reall Household Income and Affordability Calculator

Methodology Paper: Côte d'Ivoire 2021/22

This document outlines the methodology used to process the 2021 Côte d'Ivoire Harmonized Survey on Household Living Conditions, 2021-2022 data for use in Reall's Household Income and Affordability Calculator. The calculator is available at www.reall.net/calculator.

Background and Data Source

This study uses the Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM-2) [1]2021/22 for Côte d'Ivoire, implemented by the Institut National de la Statistique (INS)[2] under the LSMS[3] harmonization program. The survey is nationally representative and was fielded in two waves to capture seasonality Nov 2021–Jan 2022 and Apr–Jul 2022 using CAPI. The final sample contains 13,693 households (≈ 5,675 urban and 8,018 rural), drawn via a two-stage, stratified design by region × urban/rural: primary sampling units were selected with probability proportional to size, and 12 households per cluster were interviewed (wave totals: 6,835 in wave 1; 6,858 in wave 2). Instruments comprise a household/individual questionnaire and a community questionnaire, covering demographics, education, health, employment and income, savings/credit, food and non-food consumption, enterprises, housing, assets, transfers, shocks/safety nets, agriculture/livestock/fishing/equipment, relative poverty, community services, and local prices.

Dataset Creation

Input files used:
The following data files and mappings were used:

- **ehcvm_welfare_civ2021.dta** – main welfare dataset with household expenditure (dtot), household size (hhsize), and survey weight (hhweight)
- **ehcvm_individu_civ2021.dta** – individual-level data for employment/activity (activ12m), urban/rural (milieu), region code (region), and agro-ecological zone code (zae)
- **region_mapping_cote.xlsx** – mapping of region → **RegionName**
- **zone_mapping_cote.xlsx** – mapping of zae → **ZoneName**

Variables extracted and transformations performed:

- **HH_exp**: Monthly total household expenditure = dtot / 12

[1] https://microdata.worldbank.org/index.php/catalog/6273/study-description
[2] https://www.anstat.ci/
[3] https://microdata.worldbank.org/index.php/collections/lsms

- **HH_size**: Household size = hhsize
- **no_of_earners**: Number of employed/active members in the past 12 months = count of activ12m ∈ {1, 3} per hhid where activ12m= 1(People employed), activ12m=2(People working in family businesses)
- **monthly_rent**: *Not available* in the public-use files for this wave (set to NA)
- **urban_rural**: "Urban" if milieu == 1, otherwise "Rural"
- **Percent_Income_Spent_on_Housing**: *Not constructed* (set to NA)
- **norm_weight**: Normalized survey weight = round(hhweight / 100)
- **Location**: **ZoneName** (agro-ecological zone) from zae via lookup; formatted in title case
- **City**: **RegionName** from region via lookup; formatted in title case
- Rows with missing HH_exp, HH_size, or norm_weight were dropped to ensure consistency in analysis.

Notes:

- Region and agro-ecological zone names were merged from the lookup tables and cleaned (title case) to ensure consistent naming for **Location** and **City**.
- Identifier/code fields were harmonised to match the lookup keys before joining (e.g., RegionCode, AgroZone).

```
# === Côte d'Ivoire 2021 Data Processing ===

# Step 1: Load welfare data

welfare <- read_dta("C:/Users/Pratyusha/Documents/Reall/Data/Côte
d'Ivoire/ehcvm_welfare_civ2021.dta") %>%

  select(hhid, hhsize, dtot, hhweight)

# Step 2: Load individual-level data for earners + location mapping

indiv <- read_dta("C:/Users/Pratyusha/Documents/Reall/Data/Côte
d'Ivoire/ehcvm_individu_civ2021.dta") %>%

  select(hhid, activ12m, milieu, region, zae)

# Step 3: Number of earners per household

earners <- indiv %>%

  filter(activ12m %in% c(1,2)) %>%

  group_by(hhid) %>%

  summarise(no_of_earners = n(), .groups = 'drop')

# Step 4: Extract urban/rural and region info (one row per household)

location_data <- indiv %>%

  group_by(hhid) %>%

  summarise(

    urban_rural = ifelse(first(milieu) == 1, "Urban", "Rural"),

    RegionCode = first(region),

    AgroZone = first(zae),

    .groups = "drop"

  )

# Step 5: Load region and agro zone mapping files

region_map <- read_excel("C:/Users/Pratyusha/Documents/Reall/Data/Côte
d'Ivoire/region_mapping_cote.xlsx") %>%

  rename(RegionCode = region, RegionName = region_name)

zone_map <- read_excel("C:/Users/Pratyusha/Documents/Reall/Data/Côte
d'Ivoire/zone_mapping_cote.xlsx") %>%

  rename(AgroZone = zae, ZoneName = zae_name)
```

```
# Step 6: Merge all datasets
merged_data <- welfare %>%
 left_join(earners, by = "hhid") %>%
 left_join(location_data, by = "hhid") %>%
 left_join(region_map, by = "RegionCode") %>%
 left_join(zone_map, by = "AgroZone") %>%
 mutate(
  HH_exp = dtot / 12,  # Convert annual to monthly expenditure
  norm_weight = round(hhweight / 100),
  Country = "Cote d Ivoire",
  Year = 2021,
  Location = str_to_title(ZoneName),
  City = str_to_title(RegionName),
  Percent_Income_Spent_on_Housing = NA_real_
 ) %>%
 drop_na(HH_exp, hhsize, norm_weight)
```

Expanding Dataset by Weight

- The analysis sample was expanded to approximate the population distribution using the **normalized survey weight**: norm_weight = round(hhweight / 100).
- Records were replicated with uncount(weights = norm_weight) to create a weight-expanded micro-population for distributional statistics.
- **Note:** Sample sizes are computed **before** expansion (true interviewed counts); quantiles are computed on the **expanded** data.

```
# Step 7: Expand based on survey weights
expanded_data <- merged_data %>%
  uncount(weights = norm_weight)
```

Calculating Sample Sizes

We computed the number of valid survey households (i.e., sample size) for each **region-province-urban/rural** group. Two sets of sample sizes were calculated:

- **Urban/Rural disaggregated**
- **Combined Urban & Rural (All)**

```
# Step 8: Sample size summary
sample_size_ci <- merged_data %>%
  group_by(Location, City, urban_rural, Year) %>%
  summarise(sample_size = n(), .groups = 'drop') %>%
  mutate(Country = "Cote d Ivoire")


sample_size_ci_u_r <- merged_data %>%
  group_by(Location, City, urban_rural = "All", Year) %>%
  summarise(sample_size = n(), .groups = 'drop') %>%
  mutate(Country = "Cote d Ivoire")
```

Calculating Percentiles

Using a defined quantile probability range (1% to 99%), percentile values were computed for:

- HH_exp (household expenditure)
- HH_size (household size)
- no_of_earners

These were grouped by:

- Location (region)
- City (province)
- urban_rural (Urban / Rural / All)

```
# Step 10: Compute quantiles
ci_quantiles <- expanded_data %>%
  group_by(Location, City, urban_rural) %>%
  group_modify(~ calculate_quantiles(.x, quantile_probs)) %>%
  ungroup() %>%
  mutate(Country = "Cote d Ivoire", Year = 2021)

ci_quantiles_u_r <- expanded_data %>%
  group_by(Location, City, urban_rural = "All") %>%
  group_modify(~ calculate_quantiles(.x, quantile_probs)) %>%
  ungroup() %>%
  mutate(Country = "Cote d Ivoire", Year = 2021)
```

Final Dataset Preparation

- Quantile and sample size datasets were merged: The two Côte d'Ivoire quantile files (ci_quantiles, ci_quantiles_u_r) were merged with sample_size_ci and sample_size_ci_u_r, respectively, using left_join() on Country, Location, City, urban_rural, and Year. This ensured each quantile summary included the correct sample_size.
- Columns were aligned using a common schema: Variable names and structure were standardized with align_columns(ci_final_combined, common_columns) to ensure compatibility across country files.
- Combine relevant sample size datasets: All disaggregated and combined sample size datasets across countries (including sample_size_ci and sample_size_ci_u_r) were appended using bind_rows() into a unified dataset called combined_sample_size.
- Join combined sample size to summary dataset: The combined_sample_size dataset was merged with summary_dataset using left_join() on Country, urban_rural, Location, City, and Year.

```
# Step 11: Merge quantiles with sample size
ci_final <- ci_quantiles %>%
  left_join(sample_size_ci, by = c("Country", "Location", "City", "urban_rural", "Year")) %>%
  select(Country, Location, Year, Quantile, urban_rural, sample_size,
         HH_exp, no_of_earners, HH_size, Percent_Income_Spent_on_Housing, City)
ci_final_combined <- bind_rows(
  ci_final,
  ci_quantiles_u_r %>%
    left_join(sample_size_ci_u_r, by = c("Country", "Location", "City", "urban_rural", "Year")) %>%
    select(Country, Location, Year, Quantile, urban_rural, sample_size,
           HH_exp, no_of_earners, HH_size, Percent_Income_Spent_on_Housing, City))
# === Côte d'Ivoire 2021 Integration into Final Dataset ===
# Align columns
ci_final_combined_aligned <- align_columns(ci_final_combined, common_columns)
# Append to summary dataset and sample size table
summary_dataset <- bind_rows(summary_dataset, ci_final_combined_aligned)
# combined_sample_size <- bind_rows(combined_sample_size, sample_size_ci,
sample_size_ci_u_r)
```

```
combined_sample_size <- bind_rows(
  sample_size_india,  sample_size_india_u_r,  sample_size_uganda,  sample_size_uganda_u_r,
  sample_size_nigeria,  sample_size_nigeria_u_r,   sample_size_pakistan,
sample_size_pakistan_u_r,  sample_size_kenya,   sample_size_kenya_u_r,
  sample_size_kenya_2020,  sample_size_kenya_2020_u_r,  sample_size_india_2022,
  sample_size_india_u_r_2022,  sample_size_ph_2023,  sample_size_ph_u_r_2023,
sample_size_ph_2018,  sample_size_ph_u_r_2018,  sample_size_rw,  sample_size_rw_u_r,
sample_size_mw, sample_size_mw_u_r, sample_size_ci, sample_size_ci_u_r,sample_size_bf,
 sample_size_bf_u_r, sample_size_kenya_2021,  sample_size_all_kenya_2021)
summary_dataset <- summary_dataset %>%
  left_join(combined_sample_size, by = c("Country","urban_rural","Location","City","Year"))
```

Aggregation

In addition to the quantile-level summaries, we also generated **aggregated metrics at both state and national levels** to support broader comparisons within and across countries:

- **State-Level Aggregation:**
  The dataset was grouped by Country, Location (region), Year, Quantile, and urban_rural. Within each group, we calculated average household expenditure, household size, number of earners, and percentage of income spent on housing. The sample sizes were summed to reflect total representation. Kenya was excluded from this step as it is handled separately due to unique formatting.
- **National-Level Aggregation:**
  All regions within each country were collapsed into a national summary by grouping on Country, Year, Quantile, and urban_rural. This allowed for country-level benchmarking across time and income segments.
- **Combined Dataset Creation:**
  The original quantile-level data, state-level aggregates, and national-level aggregates were combined using bind_rows() into a single unified dataset.

```
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),
    Percent_Income_Spent_on_Housing = mean(Percent_Income_Spent_on_Housing, na.rm =
TRUE)
  ) %>%
 ungroup()
%>%
mutate(City = "All")
```

```
# ----------------------------------------
# National-Level Aggregation
# ----------------------------------------
# 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)
  #Percent_Income_Spent_on_Housing = mean(Percent_Income_Spent_on_Housing, na.rm = TRUE)
 ) %>%
 ungroup() %>%
 mutate(Location = "All",
      City = "All")
# Combine State and National Data
# Combine both the state-level and national-level data
final_dataset <- bind_rows(state_aggregated, national_aggregated,summary_dataset)
```

Calculating Income Growth

All household income data for Côte d'Ivoire was projected using **GDP per capita growth rates**.

For the historical period **2016–2024**, actual annual GDP per capita growth rates (sourced from the World Bank)[4] were applied directly to the dataset. For the forward-looking period **2025–2040**, a **median annual growth rate** of 2.426% was used.

[4] https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG

Year	GDP per capita growth rate (%)
2000	-2.98789
2001	-4.73707
2002	-5.21174
2003	-7.07102
2004	0.712926
2005	-1.38665
2006	0.42196
2007	-1.22416
2008	2.426
2009	1.326185
2010	4.519307
2011	-7.40425
2012	5.395836
2013	8.573224
2014	6.751698
2015	4.142329
2016	4.108584
2017	4.421612
2018	2.053349
2019	3.962447
2020	-1.81621
2021	4.447092
2022	3.756021
2023	3.818193
2024	3.400052
2025	2.426