Life Expectancy Analysis: Data Processing and Fixed Effects Estimation

Luis Alejandro Rubiano

Data Processing

We analyze the WHO Life Expectancy dataset. The data comprises 2,938 country-year observations across 22 variables. Preprocessing followed a two-tiered imputation strategy:

1. **Type 2 missingness (structural):** Countries missing an entire variable across all years were removed, preserving statistical coherence.

2. **Type 1 missingness (sporadic):** Time-series interpolation was applied per country and variable, followed by forward and backward fill to ensure temporal continuity.

We eliminated variables with multicollinearity risks or derivative redundancy:

- `Total expenditure = GDP × percentage expenditure`

- `Adult Mortality` ≈ inverse of `Life expectancy`

- `infant deaths` highly collinear with `under-five deaths`

We log-transformed GDP to reduce skewness and discarded invariant categorical features (`is_developed`). Final dataset: 2,128 observations across 133 countries.

Statistical Analysis

We estimate a fixed-effects panel regression using `Life expectancy` as the outcome, with 15 time-varying predictors. The model controls for country-specific unobservables (e.g., institutions, geography) and clusters standard errors at the country level.

Model Specification

$$y_{it} = \alpha_i + \beta X_{it} + u_{it}$$

- y_{it}: Life expectancy in country i at year t

- X_{it}: vector of health, demographic, and economic indicators

- α_i: country fixed effect

- u_{it}: idiosyncratic error term

Model Results

- **Within $R^2 = 0.455$:** covariates explain 45.5% of intra-country variation.

- **Key coefficients (selected):**

  - HIV/AIDS: -0.40 ($p < 0.001$)
  - Income Composition: $+3.30$ ($p < 0.001$)
  - Schooling: $+0.62$ ($p < 0.001$)
  - GDP (log): $+0.53$ ($p = 0.04$)

- **Poolability test:** $p < 0.001$ confirms necessity of fixed effects.

Key Insights & Implications

Insight 1: HIV prevalence reduces life expectancy by 0.4 years per unit: A stark and statistically robust relationship suggesting that targeted interventions in HIV/AIDS control can yield significant longevity benefits within countries.

Insight 2: Human capital drives longevity: A one-point increase in education (schooling) is associated with a 0.62-year increase in life expectancy, underscoring the critical role of education in health outcomes.

Insight 3: Income composition outperforms raw GDP: While GDP is significant, the income composition index has a larger and more stable impact, suggesting that inclusive economic structures enhance life expectancy more than absolute economic scale.

These insights advocate for health and education-focused policies over purely growth-driven strategies in development agendas.