The inequality measure used in this research is the WIID Companion dataset from the United Nations (2023). This dataset provides information on annual per capita income distributions at both the country and global levels. The key variables of interest are the standardized Gini Index and the income share of the top and bottom 20 percent. The dataset partially covers 201 countries from 1890 to 2022. These two variables are chosen because the Gini Index offers a comprehensive measure of overall income inequality, while the income shares of the top and bottom 20 percent highlight disparities in wealth distribution.

Our explanatory variable, which is the reliance of a country’s economy on remittances, is represented by the personal remittances received as a percentage of GDP (World Bank, 2024). These values are based on World Bank staff estimates, using IMF balance of payments data as well as World Bank and OECD GDP estimates. IMF data on worker remittances is based on official banking reports and may underestimate actual flows, as informal transfers are not accounted for. Consequently, the estimated coefficients likely represent a lower bound of remittances' true impact on income inequality (Koechlin and Leon, 2006). The dataset partially covers 266 countries from 1890 to 2023.

First, international remittances paid and received are included, sourced from the same data as the explanatory variable described in the previous paragraph, with the data coverage also remaining consistent. Second, following Murodova (2018), education level is considered an important factor when examining the impact of international remittances on inequality. To account for this, we use school life expectancy at the tertiary education level, drawing from two datasets that separately cover expectancy for ISCED levels 1–3 and ISCED levels 5–8 (UNdata, 2016). ISCED levels 1–3 correspond to primary school through high school, while ISCED levels 5–8 are equivalent to short-cycle tertiary education through a doctorate or equivalent. The former reflects the overall accessibility of education in a country, while the latter indicates the availability and development of higher education. Both datasets provide only partial coverage for the period 1975–2016, with the dataset for ISCED levels 1–3 including 216 countries and the dataset for ISCED levels 5–8 covering 193 countries.

Additionally, we aim to include data that reflect a country's overall economic conditions. For this, we use the World Economic Outlook database (IMF, 2024). This dataset provides incomplete coverage of 44 different economic indicators across 196 countries from 1980 to 2023. To enable cross-country comparisons over time, only data presented in the form of indexes, percentages, or international dollars are used. As a result, a total of 12 indicators are included in the analysis.

It is important to note that all the datasets we use are incomplete. Therefore, our final research dataset is constructed as the intersection of all the aforementioned datasets, presented in a panel data format. In addition to data related to the economy, remittances, and inequality, we also retain information on each country's geographic region and income group to account for geographic and income-level factors, both sourced from the WIID Companion dataset from the United Nations (2023). After processing and synthesis, the final dataset consists of 386 observations, covering 52 different countries over the period from 1980 to 2015. The definitions of all datasets are provided in the Variable Descriptions Table.

Most numbers in the descriptive statistics table make sense in terms of value, and there are no dummy variables. There is no variable with low variation, but many variables exhibit very high standard deviations. This can be explained by the significant differences in economic status across countries and time periods, reflecting the diverse levels of development, income distribution, and macroeconomic conditions in the dataset. To some extent, this also reflects the broad coverage of our dataset. One limitation of our dataset is there are more observations in recent years compared to earlier times due to the incompleteness of our data in earlier periods. One limitation of our dataset is the uneven distribution of observations over time, with more data available in recent years due to incomplete coverage in earlier periods. This may limit the representativeness of our results for earlier years, making it particularly challenging to account for yearly fixed effects.