William's Update

Remittances

William Clinton Co

September 30, 2025

Abstract

This document is a follow-up to the meeting on September 5th and addresses the items discussed during that meeting. It provides an update on current progress of the dataset completion and potential next steps, while providing an analysis of the current dataset at hand.

Table of contents

1 Introduction

1.1 Data Sources and Compilation

I received a reply from Mustafizur Rahman, who compiled the South Asian remittance dataset. I forwarded the reply to you. Overall, they confirm what we mostly already learned. Next, we will now be talking about current datasets the we have.

1.1.1 Potential Data Sources

- **Bank of Italy**: Rich dataset from 2005-present (quarterly frequency)
- **Italian Provincial Data**: Outgoing remittances by province (2011-2024), allowing for **provincial-level** analysis and regressions.
  - This is an interesting dataset because it contains outgoing remittances at the provincial level, providing additional depth compared to our country-level remittance data analysis.
  - Covering the years 2011 to 2024, this dataset reports annual outgoing remittance flows, detailing both the Italian province of origin and the destination country for each remittance.

1.1.2 Other Data Sources

This is the other central bank datasets we have acquired from other papers and academics.

- Bangladesh Central Bank dataset
- Bank of Italy dataset
- Philippine dataset
- South Asia dataset
- World Bank GDP data (inflation-adjusted, base period 2015)

2 Data Quality Issues and Corrections

With our current datasets (IMF Kpodar paper and Remitscope), we will assess their validity. Since the IMF/World Bank produce aggregate remittance data, we will aggregate our bilateral corridor-level remittance data and compare the results to these official aggregates.

2.1 Initial US Outflows Analysis

As discussed in our last meeting, we will now analyze US remittance outflows.

The initial analysis of US outflows revealed potential data quality issues:

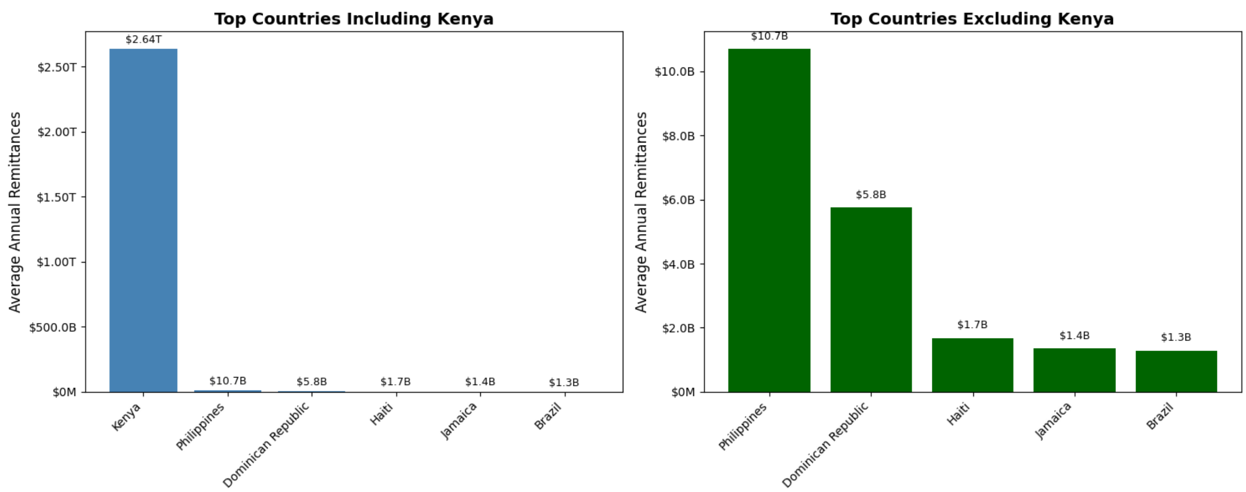

Figure 1: Initial US Outflows Analysis

As shown in Figure **??**, Kenya appeared suspiciously high in the remittance flows, prompting further investigation.

2.1.1 Kenya Data Correction

After removing Kenya from the initial analysis, the revised outflows map showed:

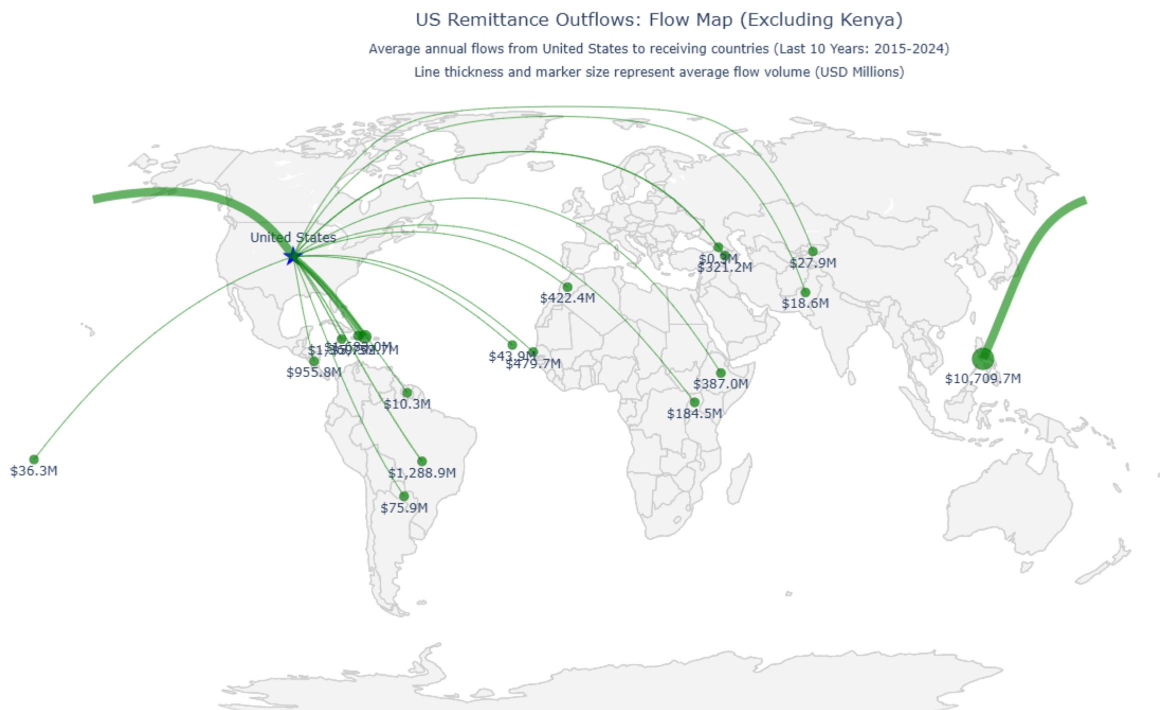

Figure 2: US Outflows Analysis (Kenya Removed)

The detailed interactive map can be accessed at: GitHub - Interactive Flow Map

2.1.2 Data Validation Process

Determining the accuracy of individual data points proved challenging. Cross-referencing with the Migration Data Portal revealed discrepancies of up to $1 billion USD in some cases. These inconsistencies likely stem from a combination of:

1. Known inaccuracies in the World Bank bilateral remittance matrix
2. Data collection and reporting errors

Kenya Data Verification:

I verified the Kenya figures against the Central Bank of Kenya's diaspora remittances data. The comparison revealed:

- Central Bank of Kenya estimate for North America: **$2.64 billion**
- RemitScope dataset figure: **$2.64 trillion** (entire order of magnitude error)

This was not a data scraping error but appeared to be a systematic issue in the Remitscope data. Evidence suggest this is not a single outlier as well, which will be seen in later analysis.

Data Correction Applied: I corrected the Kenya data by dividing by 1,000, resulting in the following US outflow dataset:

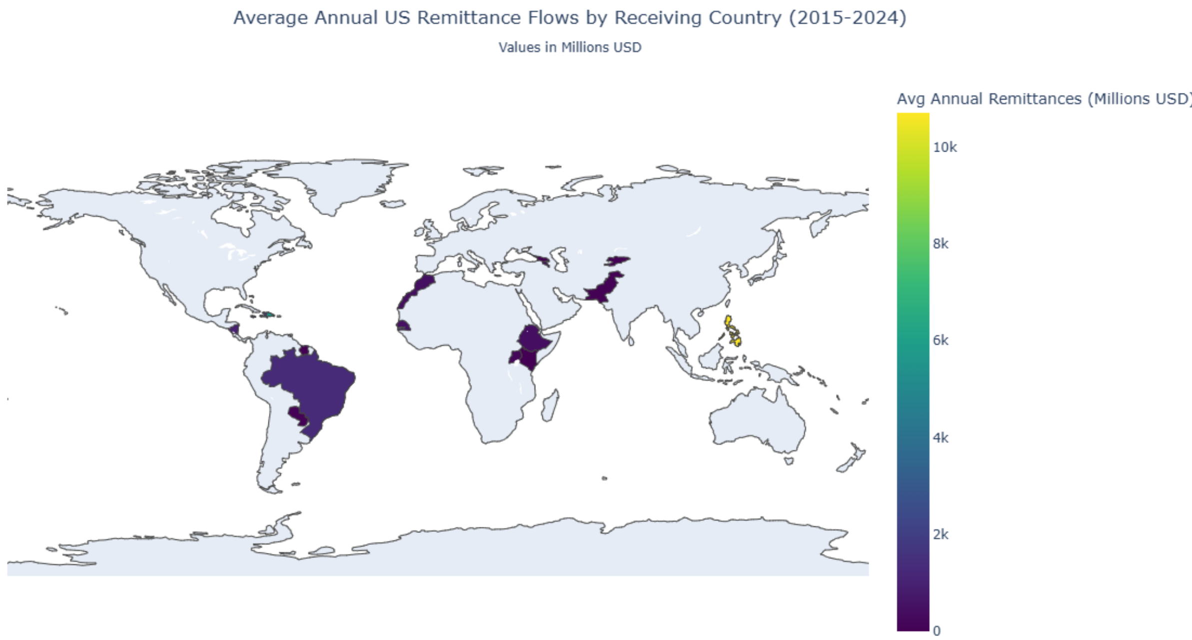

Figure 3: Corrected US Outflows Dataset

2.1.3 Top Receiving Countries

The analysis of corrected data reveals the ranking of top remittance-receiving countries from the United States:

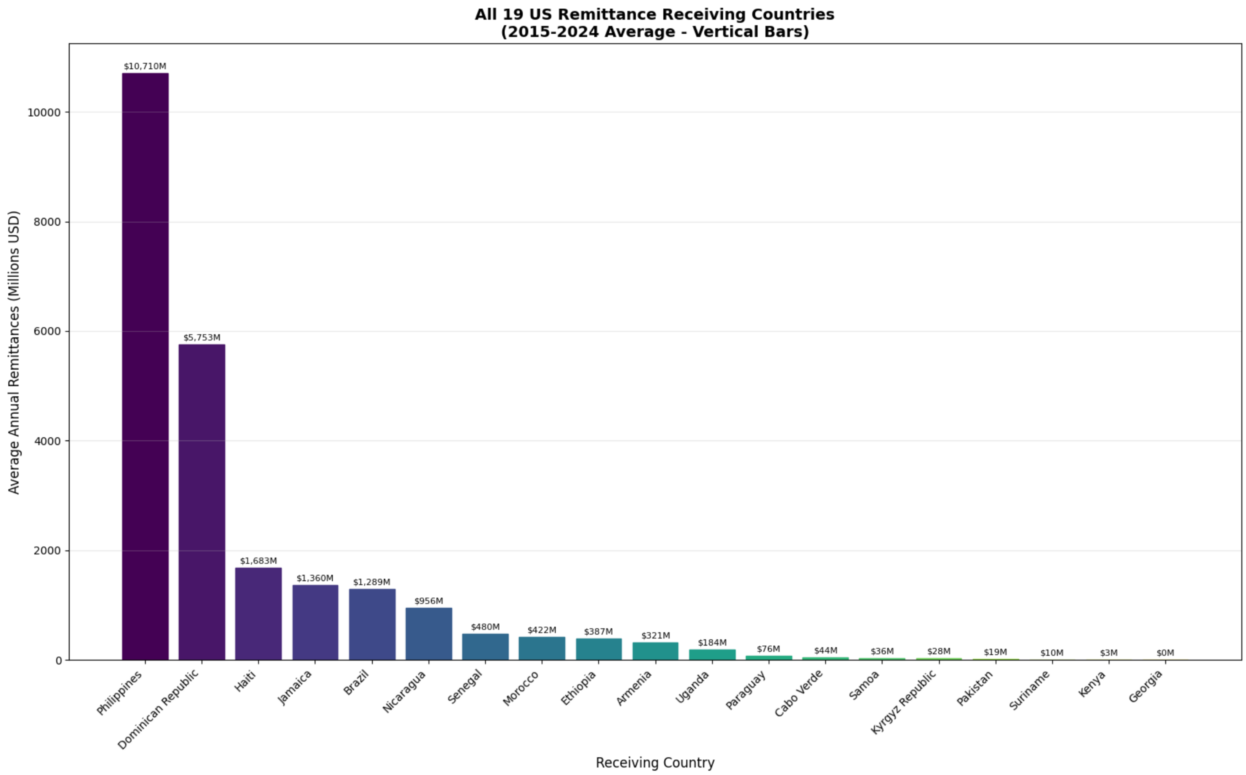

Figure 4: Top US Remittance Receiving Countries

Figure **??** illustrates the distribution of US remittance flows to major receiving countries. The observed results align with expectations based on established remittance patterns.

To further validate our dataset, we leverage the general accuracy of macro-level remittance statistics. Accordingly, we aggregate our bilateral corridor-level data and compare the results to official macro aggregates (IMF/WorldBank).

Upon aggregating the dataset, it became evident that the discrepancy associated with Kenya was systematic rather than an isolated , as illustrated in Figure **??** .

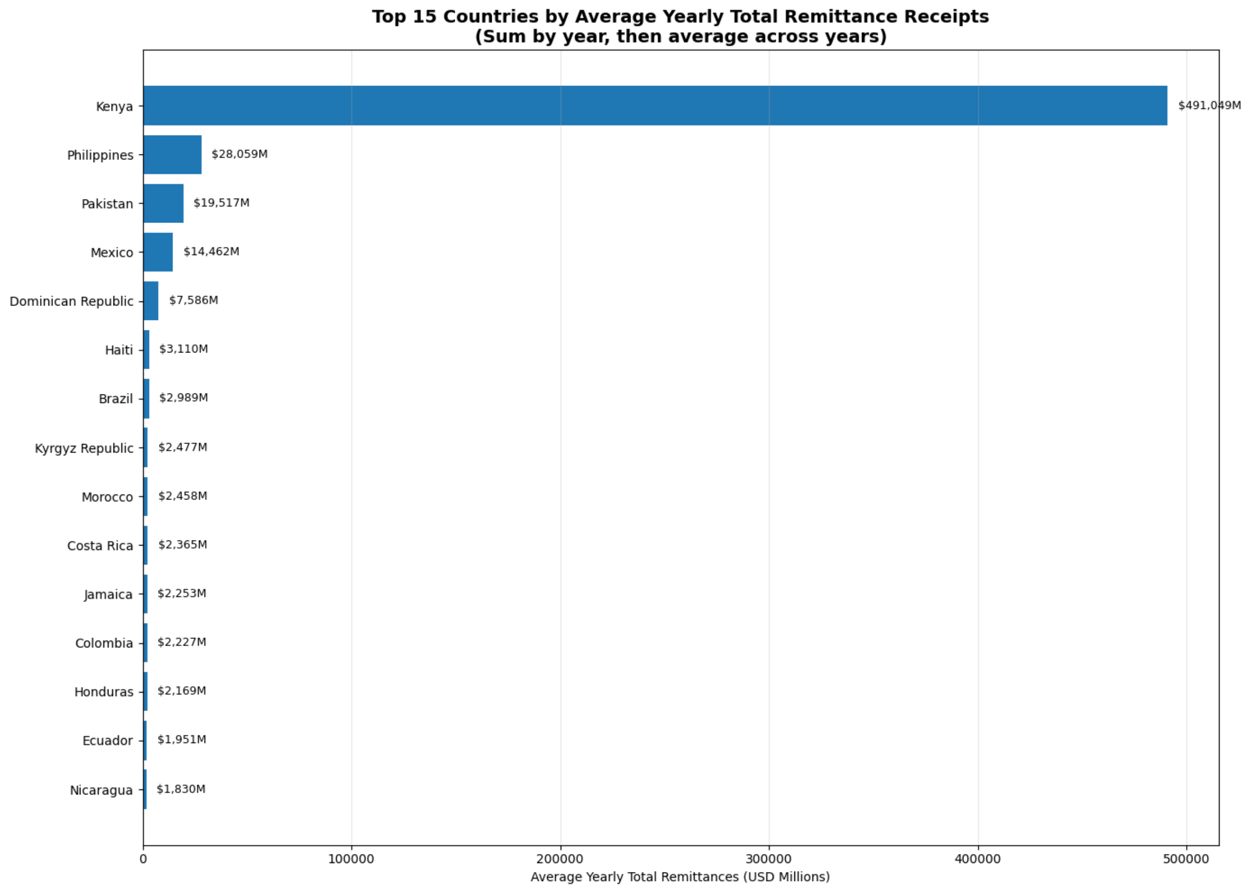

Figure 5: Kenya Systematic Issue

Specifically, the dataset reported remittance flows to Kenya at approximately $491 billion, whereas the World Bank macro remittance dataset indicates a figure closer to $4.3 billion USD. This substantial divergence highlights a persistent issue within the datasets.

To address this, we excluded Kenya from the analysis and employed the median as a robust measure to mitigate the impact of outliers. The resulting side-by-side comparison of the datasets is presented in Figure **??**.

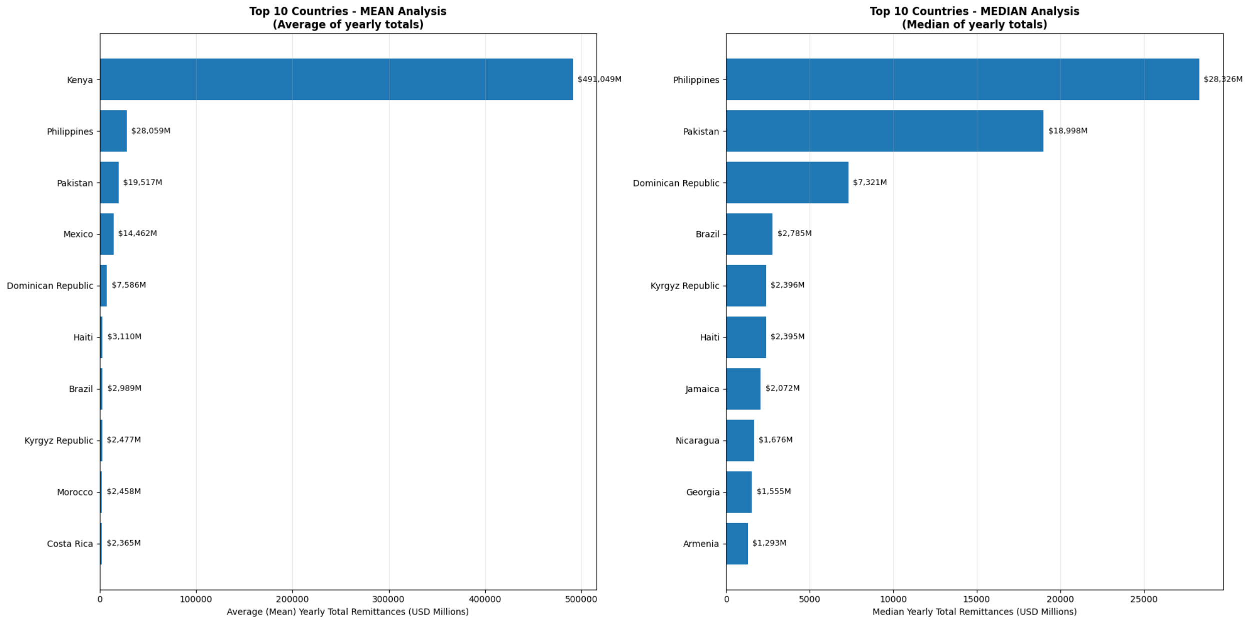

Figure 6: Dataset Side by Side

Further, Figure **??** displays the top 30 remittance-receiving countries, comparing both the median and mean after removing Kenya from the sample.

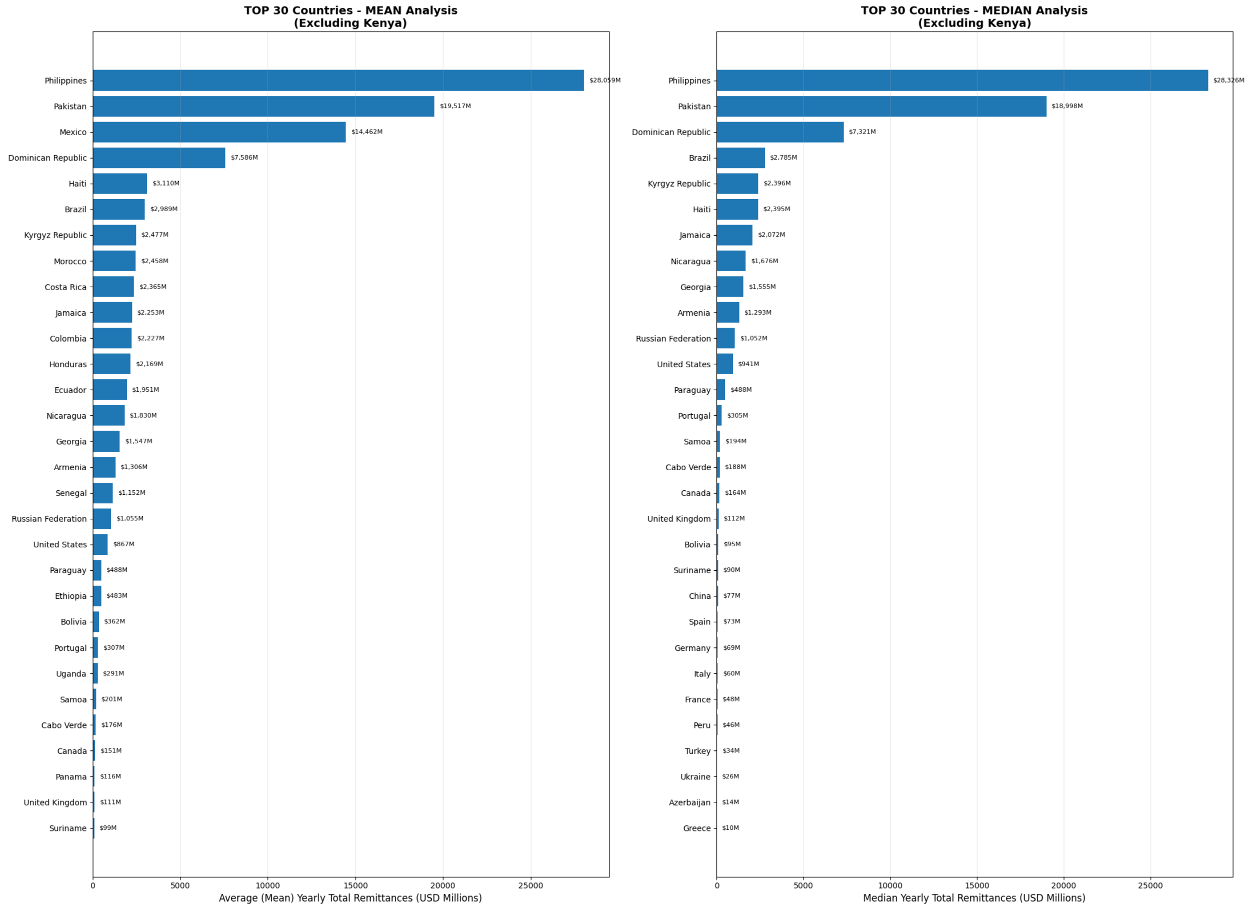

Figure 7: Top 30 Remittance Analysis

2.2 IMF Macro Remittance vs Dataset

To assess the consistency of our dataset with established benchmarks, we compare our remittance estimates to those reported in the IMF macro remittance dataset. This comparison is conducted across several dimensions:

- Regional differences are illustrated in Figure **??**.

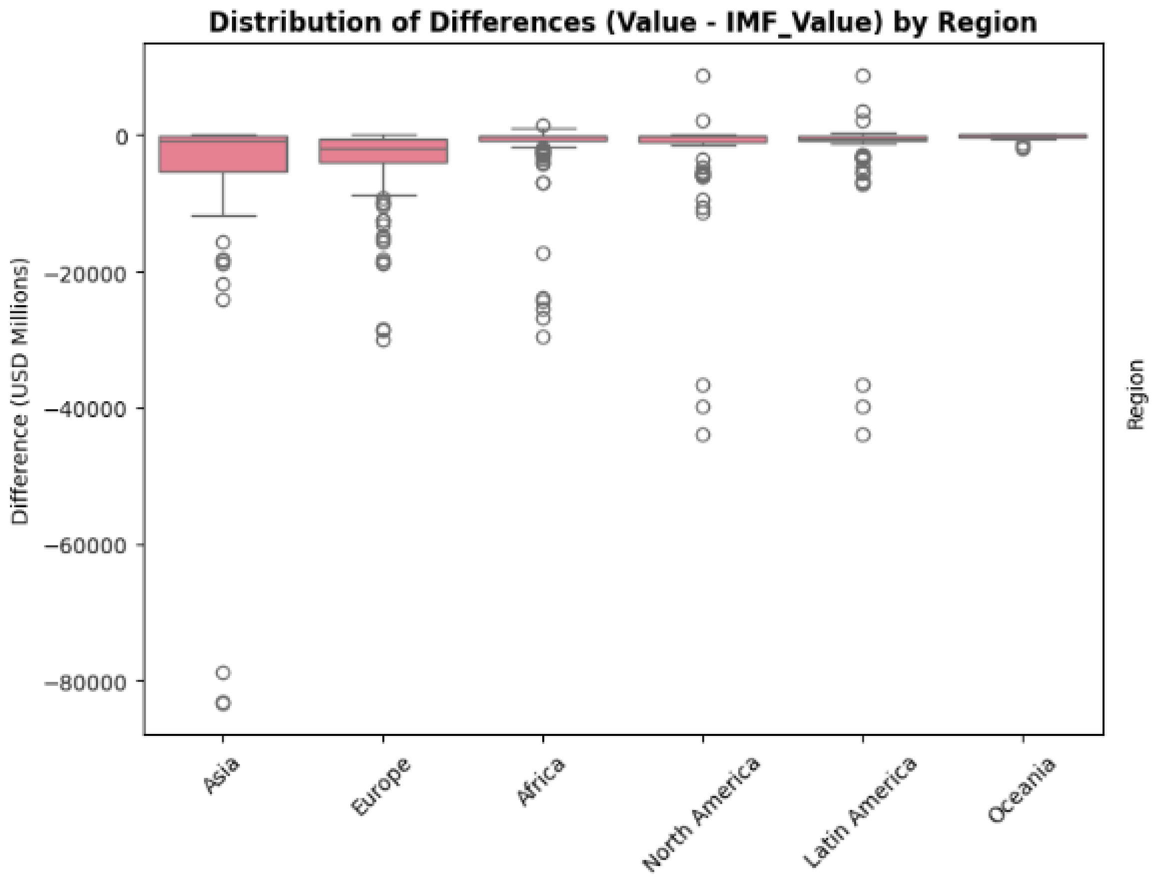

Figure 8: Regional Differences

- Temporal differences are presented in Figure **??**

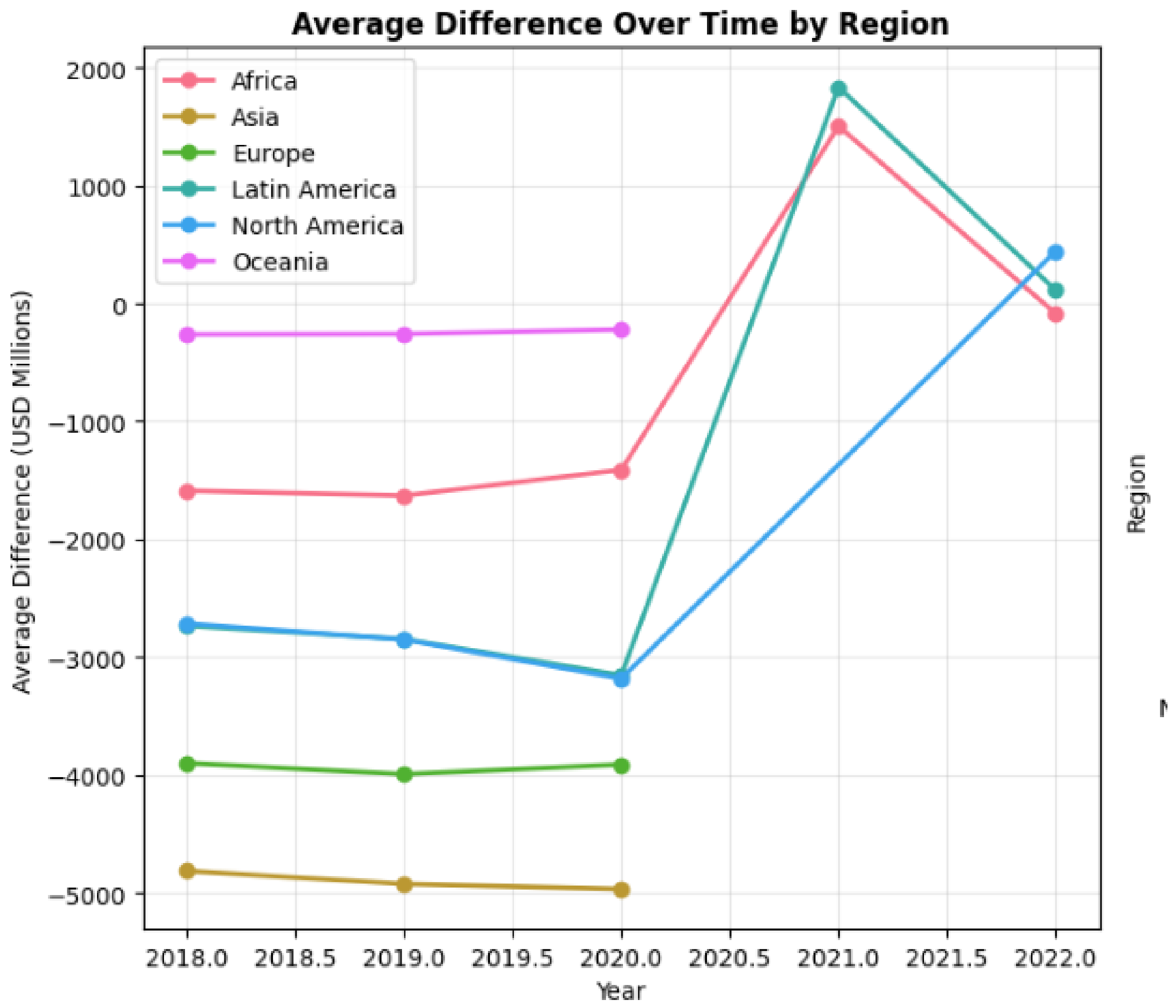

Figure 9: Time Series Differences

- Combined year and region analysis is shown in Figure **??**

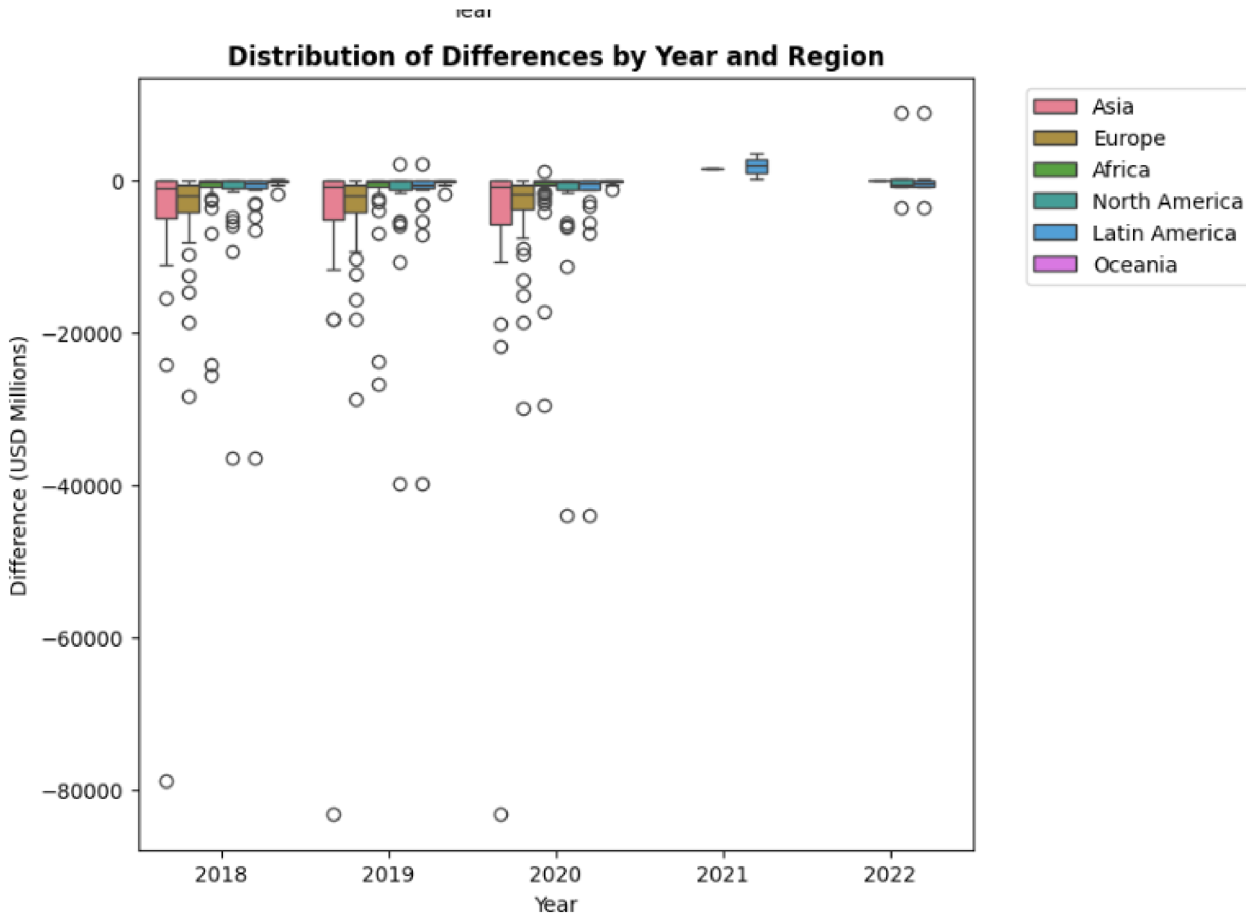

Figure 10: Year and Region Analysis

Next, we sort the data by country to highlight the magnitude of observed discrepancies. Notably, several countries exhibit substantial differences, with discrepancies ranging from \$10 billion to \$80 billion USD between our dataset and official macro-level statistics Figure **??** :

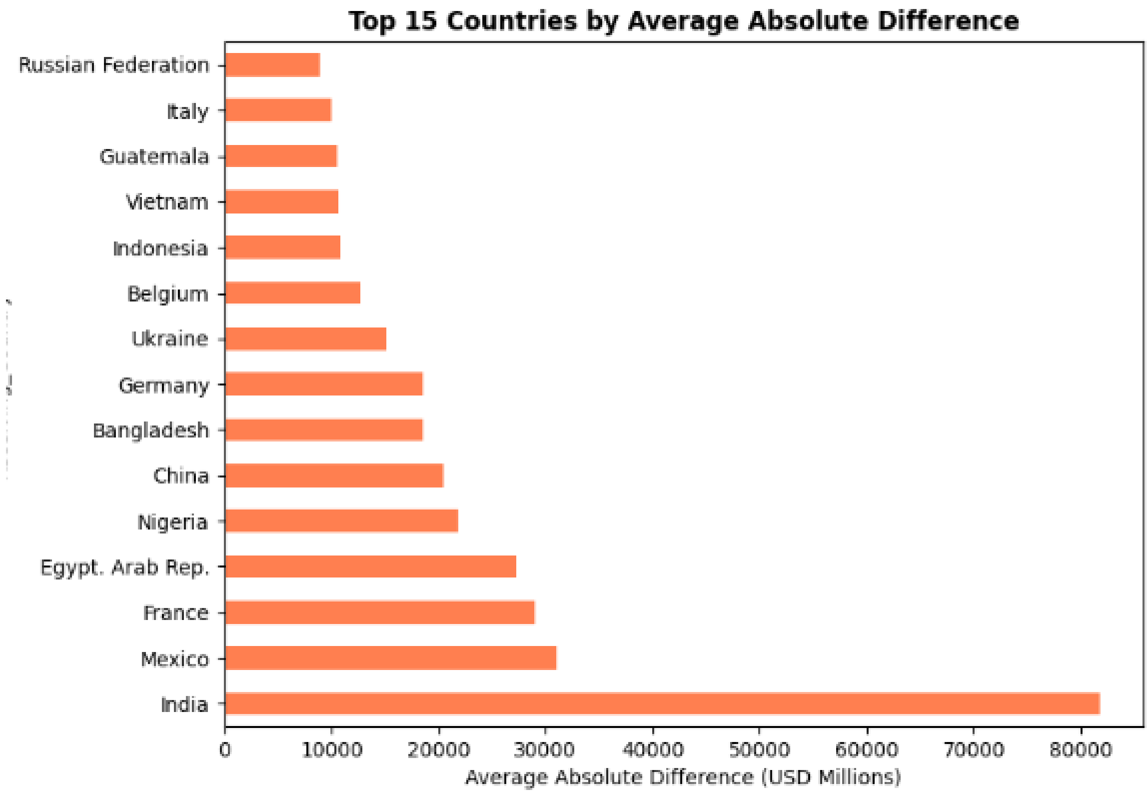

Figure 11: Country-Level Discrepancies

To further contextualize these discrepancies, we compute the average difference at the country level. While this approach yields more moderate values, the discrepancies remain substantial, ranging from approximately \$1.5 billion to \$3 billion USD. The results are presented in Figure **??**.

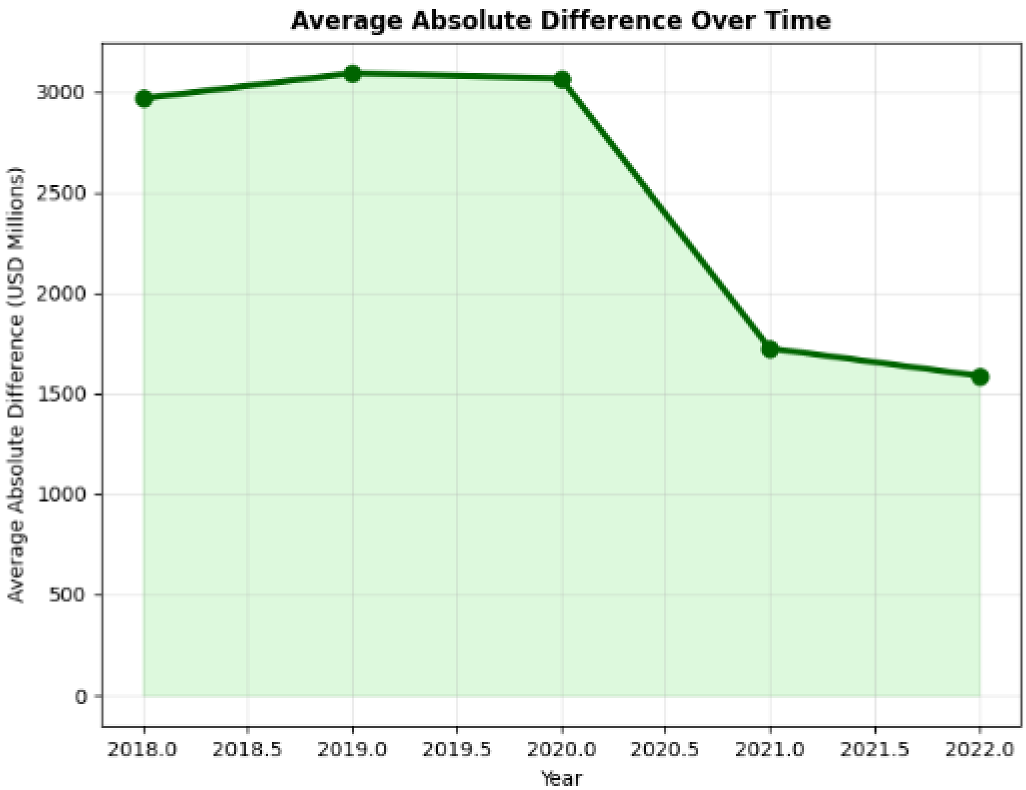

Figure 12: Average Country Discrepancies

Given RemitScope's emphasis on Africa and Latin America, we subset our dataset to focus specifically on these regions. The results of this regional analysis are presented in Figure **??** and Figure **??**.

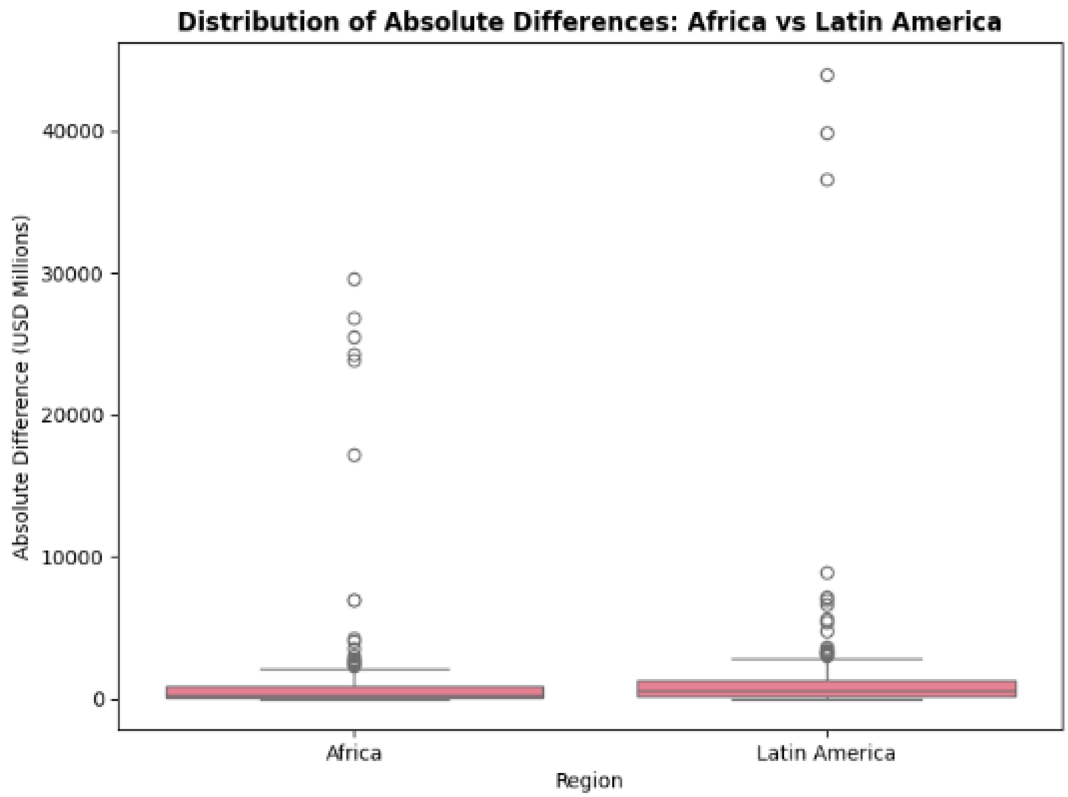

Figure 13: Latin America Focus 1

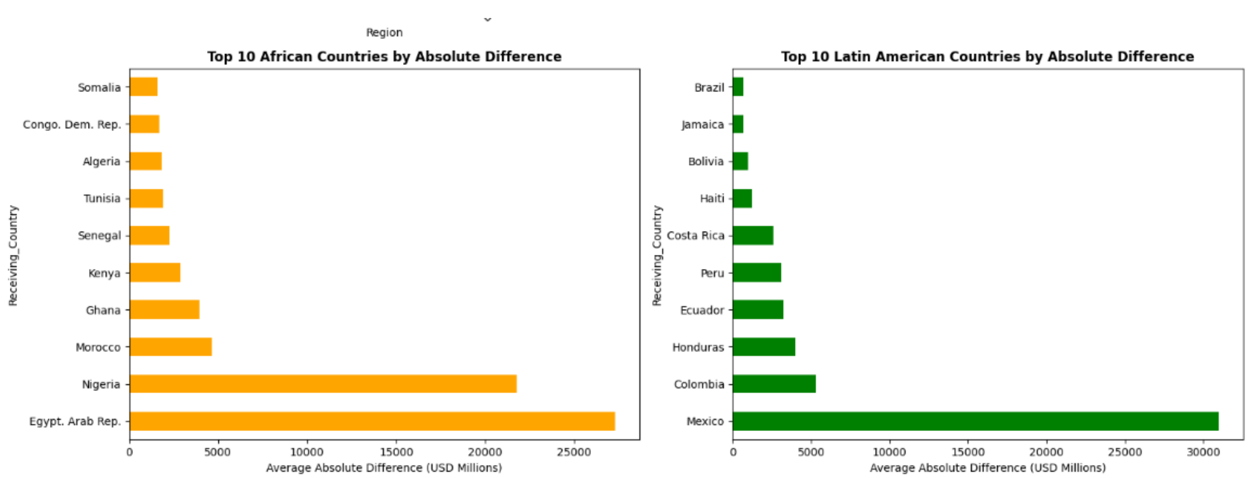

Figure 14: Latin America Focus 2

To further refine the analysis, we remove prominent outliers such as Egypt, Nigeria, and Mexico. The impact of this adjustment is shown in Figure **??** and Figure **??**.

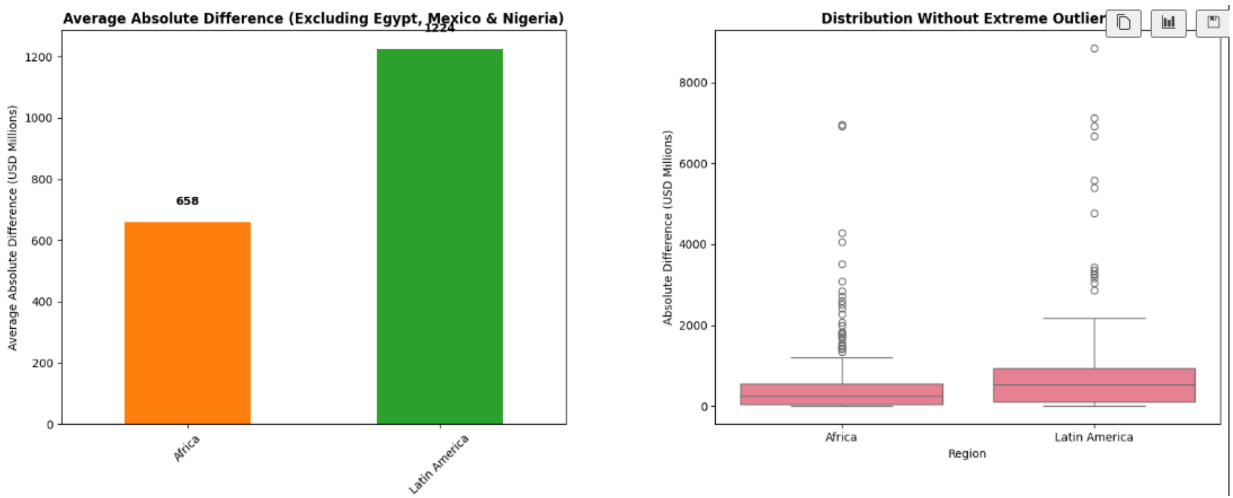

Figure 15: Outliers Removed 1

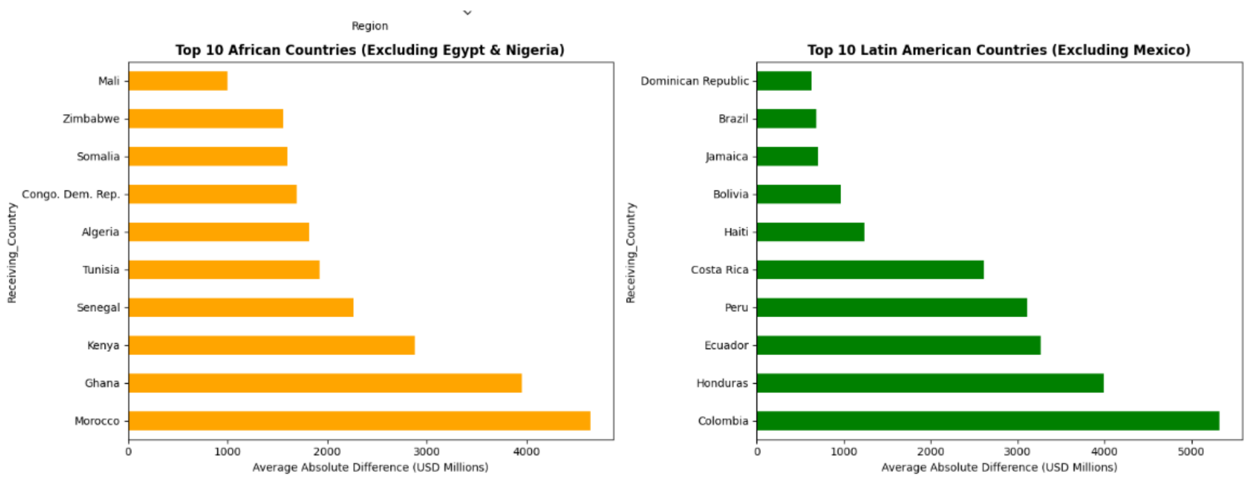

Figure 16: Outliers Removed 2

Despite these adjustments, the discrepancies remain substantial, often exceeding $1 billion USD. A country-level isolation of the data is provided in Figure **??**.

Figure 17: Country Isolation

These errors can be partially attributed to several factors: 1. Discrepancies in macro-level data sources may arise due to differences in reporting standards and aggregation methods. 2. Data encoding errors within RemitScope can introduce inconsistencies (as shown with Kenya). 3. Central bank datasets often prioritize major remittance corridors, frequently excluding countries with minimal remittance flows. As a result, published data may only reflect the top 20 remitting countries or similar subsets.

In summary, while the dataset exhibits notable discrepancies, these issues can be managed through careful methodological adjustments.

2.3 Empirical Analysis: Drivers of Remittances

With the corrected dataset, I proceed to analyze the determinants of remittance flows through regression analysis.

2.4 Econometric Specifications

The GDP mapping process was challenging due to unstructured data, but I successfully completed it. ### Linear Model (Levels)

For comparison purposes, I estimate a linear specification in levels:

$$\text{Remit}_{ij,t} = \alpha_0 + \alpha_1 \text{GDP}_{\text{sender},i,t} + \alpha_2 \text{GDP}_{\text{receiver},j,t} + u_{ij,t} \tag{1}$$

Where:

- Dependent variable: **Remittances (level)** in millions USD
- α_1 and α_2 measure the **marginal change** in remittances for a one-unit change in GDP (e.g., $1 billion)
- $u_{ij,t}$ is the error term

2.4.1 Log-Log (Gravity) Model

Following the gravity model literature, I estimate a log-linear specification:

$$\ln(\text{Remit}_{ij,t}) = \beta_0 + \beta_1 \ln(\text{GDP}_{\text{sender},i,t}) + \beta_2 \ln(\text{GDP}_{\text{receiver},j,t}) + \varepsilon_{ij,t} \tag{2}$$

Where:

- Dependent variable: ln(Remittances)
- Key independent variables: $\ln(\text{GDP}_{\text{sender}})$, $\ln(\text{GDP}_{\text{receiver}})$
- $\varepsilon_{ij,t}$ is the error term

Interpretation:

- β_1 = elasticity of remittances with respect to **sender GDP**
- β_2 = elasticity of remittances with respect to **receiver GDP**

The gravity model specification allows for percentage interpretations: a 1% increase in sender GDP leads to a β_1% change in remittances, holding receiver GDP constant.

2.5 Unlagged Model

A baseline specification can be written as:

$$Remittances_{i,t} = \alpha + \beta_1\, GDP_{i,t}^{sending} + \beta_2\, GDP_{i,t}^{receiving} + \epsilon_{i,t}$$

where:

- $Remittances_{i,t}$ = remittance inflows from country i at time t
- $GDP_{i,t}^{sending}$ = GDP of the migrant-hosting (sending) country
- $GDP_{i,t}^{receiving}$ = GDP of the migrant-origin (receiving) country
- $\epsilon_{i,t}$ = error term

2.6 Lagged Model

To reduce concerns about reverse causality (i.e., remittances also influencing GDP), we can use lagged GDP values:

$$Remittances_{i,t} = \alpha + \beta_1\, GDP_{i,t-1}^{sending} + \beta_2\, GDP_{i,t-1}^{receiving} + \epsilon_{i,t}$$

In this model, current remittances are explained by *past* GDP levels in both the sending and receiving countries. This assumes that today's remittances cannot affect yesterday's GDP.

2.7 Extensions

Additional control variables may be added to improve robustness, for example:

- Exchange rate: $ExRate_{i,t}$
- Migrant stock: $MigrantStock_{i,t}$
- Unemployment in sending country: $Unemp_{i,t}^{sending}$

A more general model could look like:

$$Remittances_{i,t} = \alpha + \beta_1 \, GDP_{i,t-1}^{sending} + \beta_2 \, GDP_{i,t-1}^{receiving} + \beta_3 \, ExRate_{i,t} + \beta_4 \, MigrantStock_{i,t} + \epsilon_{i,t}$$

2.8 Interpretation

- $\beta_1 > 0$: Higher GDP in the sending country increases remittance flows.
- $\beta_2 < 0$: Higher GDP in the receiving country reduces the need for remittances.
- β_3 and β_4: Context-dependent but expected to be positive in most cases.