William's Update

Remittances

William Clinton Co

October 1, 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. Potential next steps would be to gather more datasets from other country's central banks.

- 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:

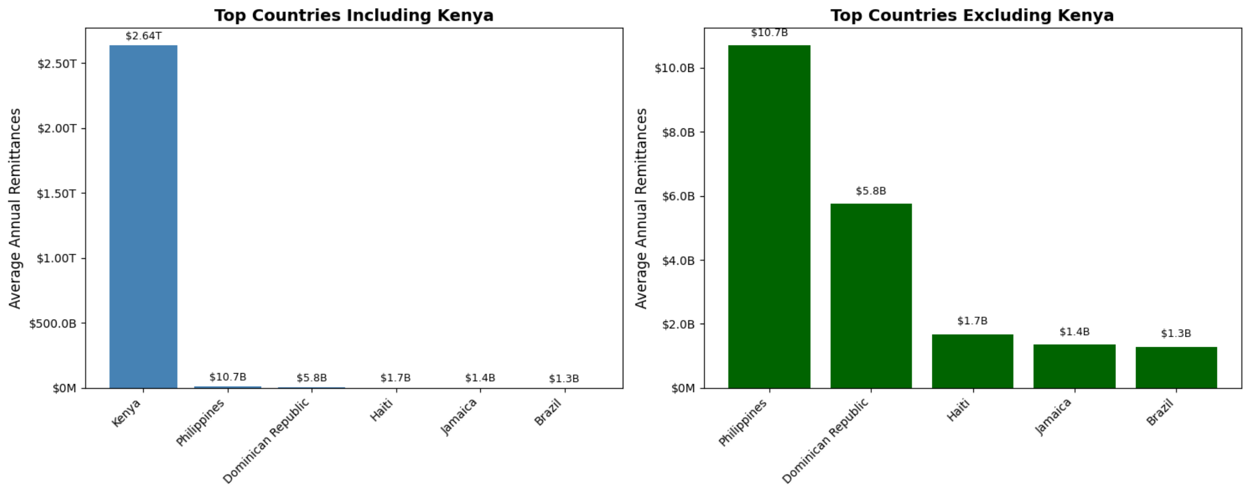

Figure 1: Initial US Outflows Analysis

As shown in Figure 1, 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:

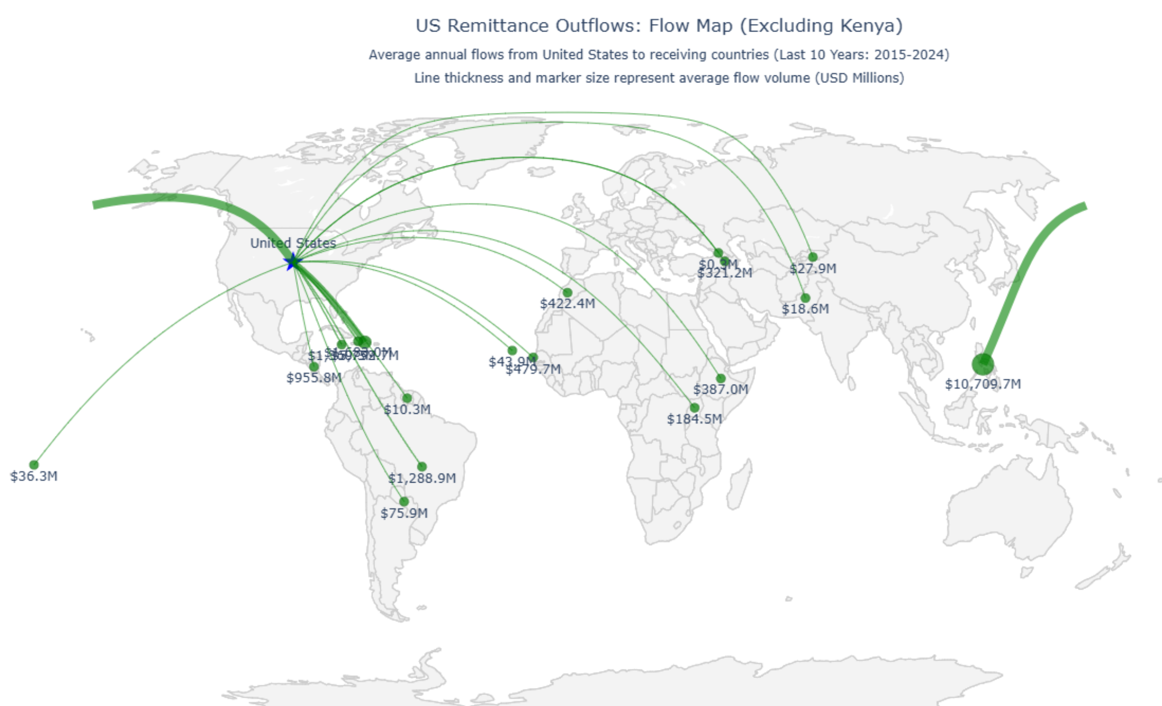

Figure 2: US Outflows Analysis (Kenya Removed)

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:

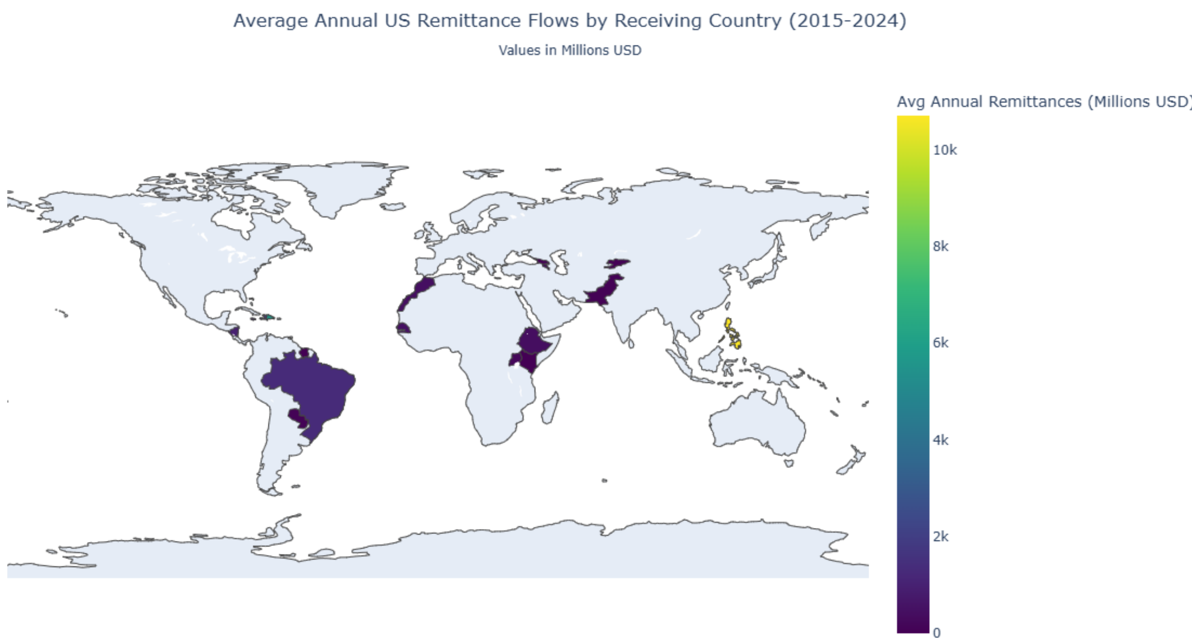

Figure 3: Corrected US Outflows Dataset

2.1.3 Top Receiving Countries

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

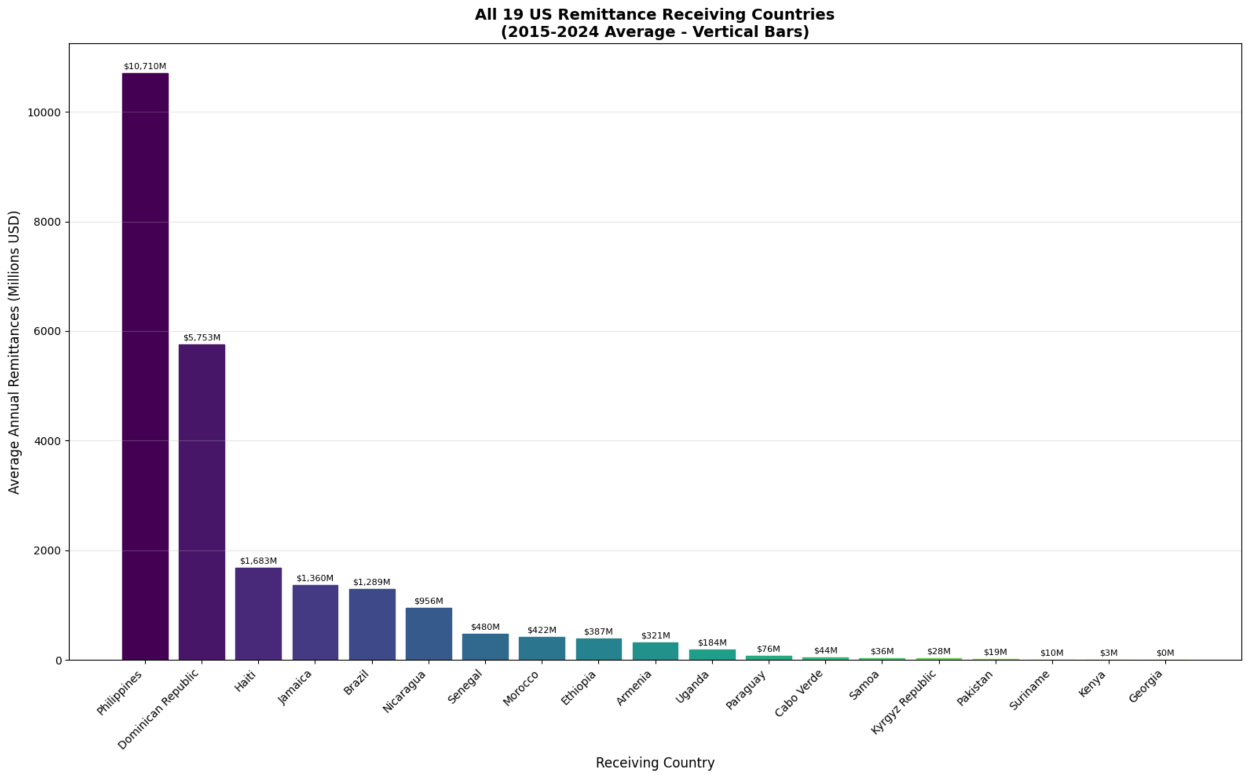

Figure 4: Top US Remittance Receiving Countries

Figure 4 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 5 .

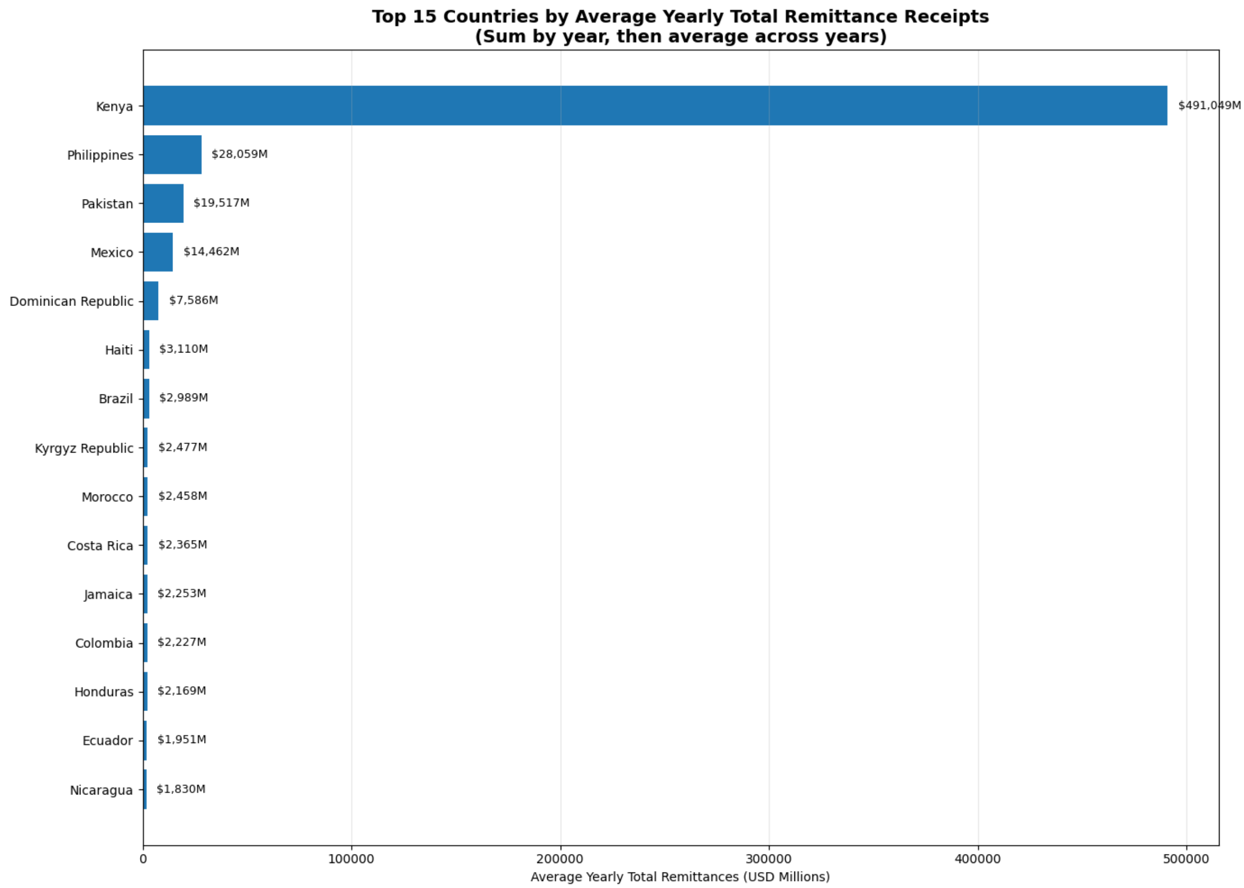

Figure 5: Kenya Systematic Issue

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 6.

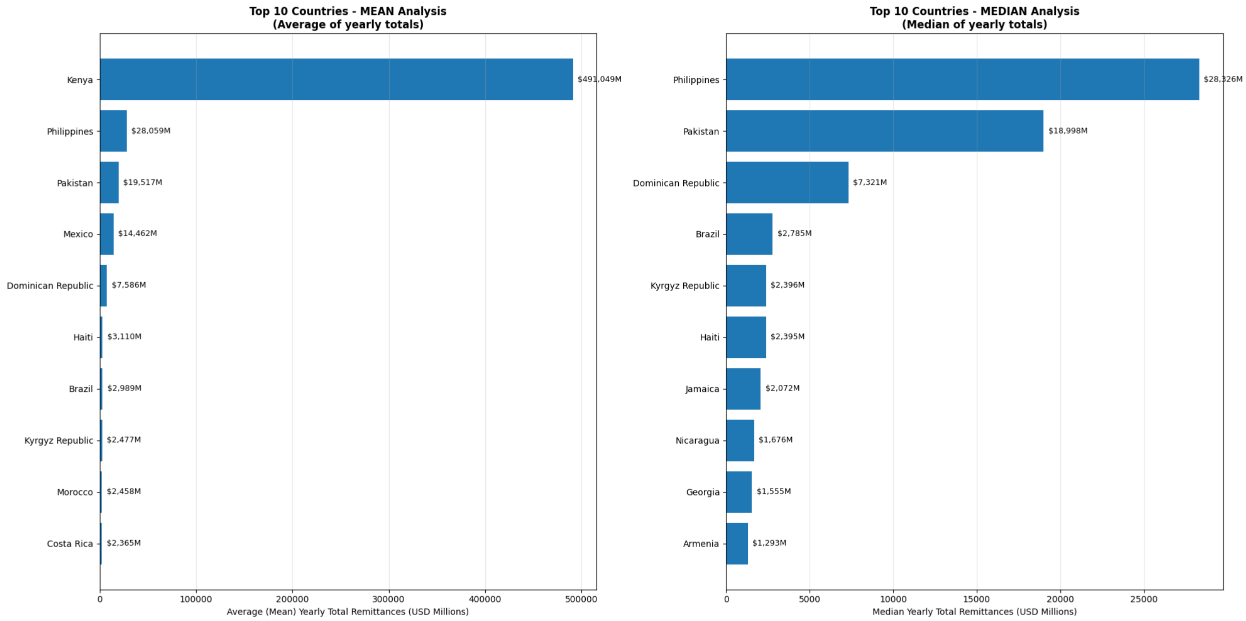

Figure 6: Dataset Side by Side

Further, Figure 7 displays the top 30 remittance-receiving countries, comparing both the median and mean after removing Kenya from the sample.

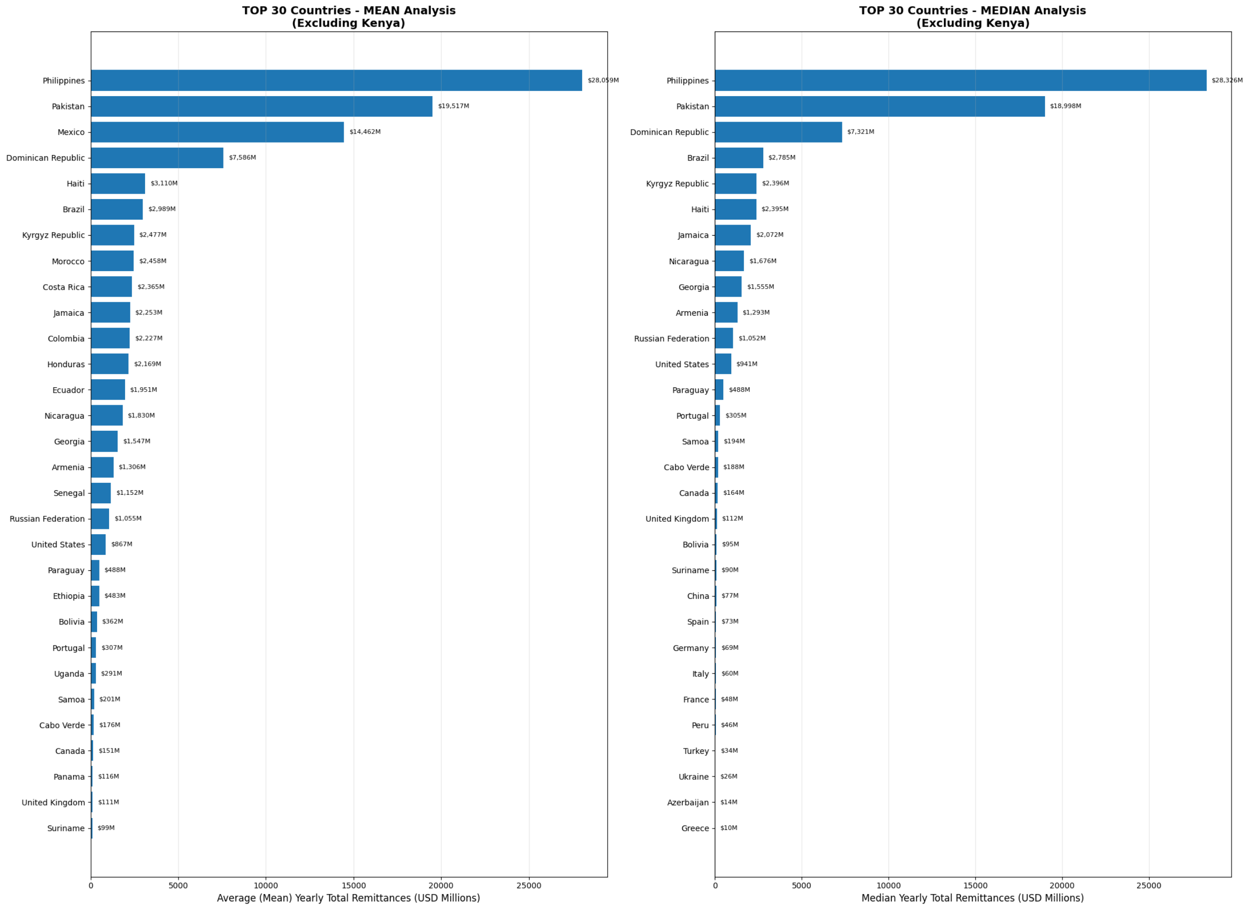

Figure 7: Top 30 Remittance Analysis

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 8.

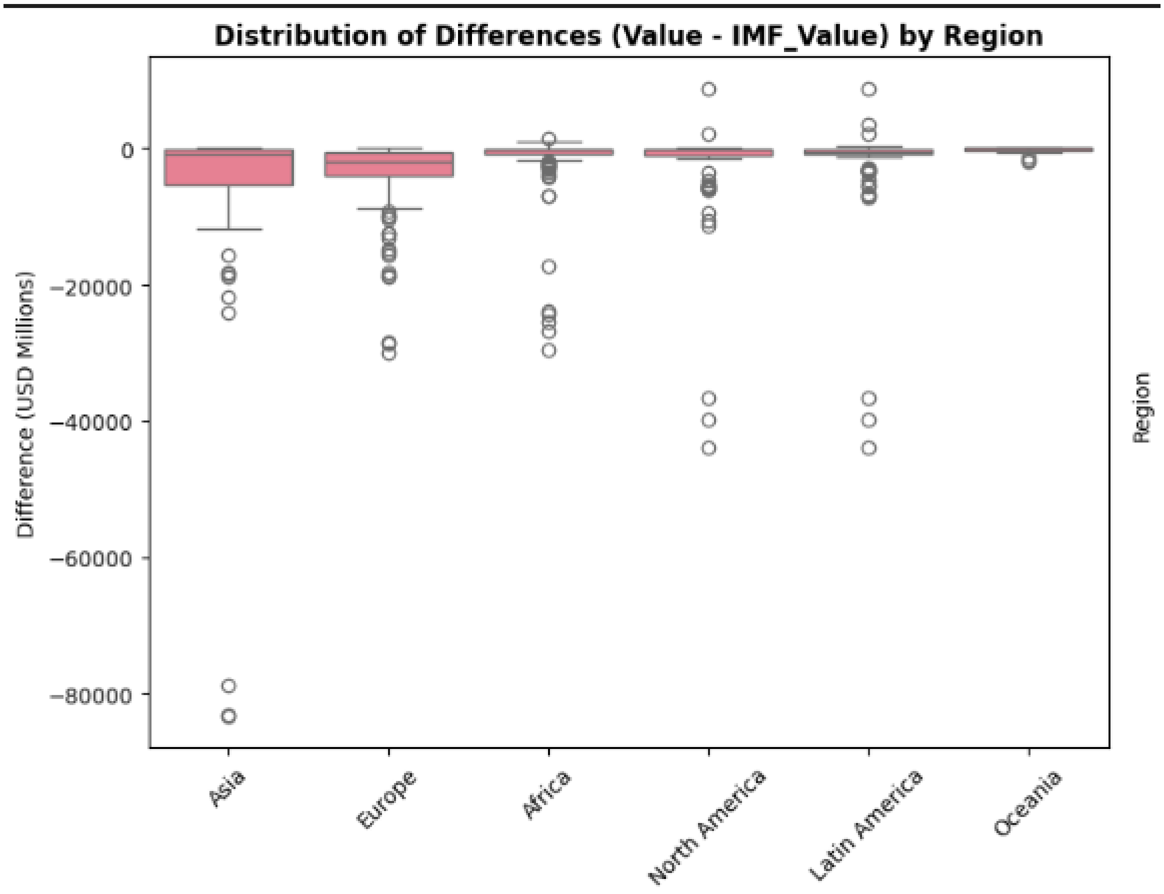

Figure 8: Regional Differences

- Temporal differences are presented in Figure 9

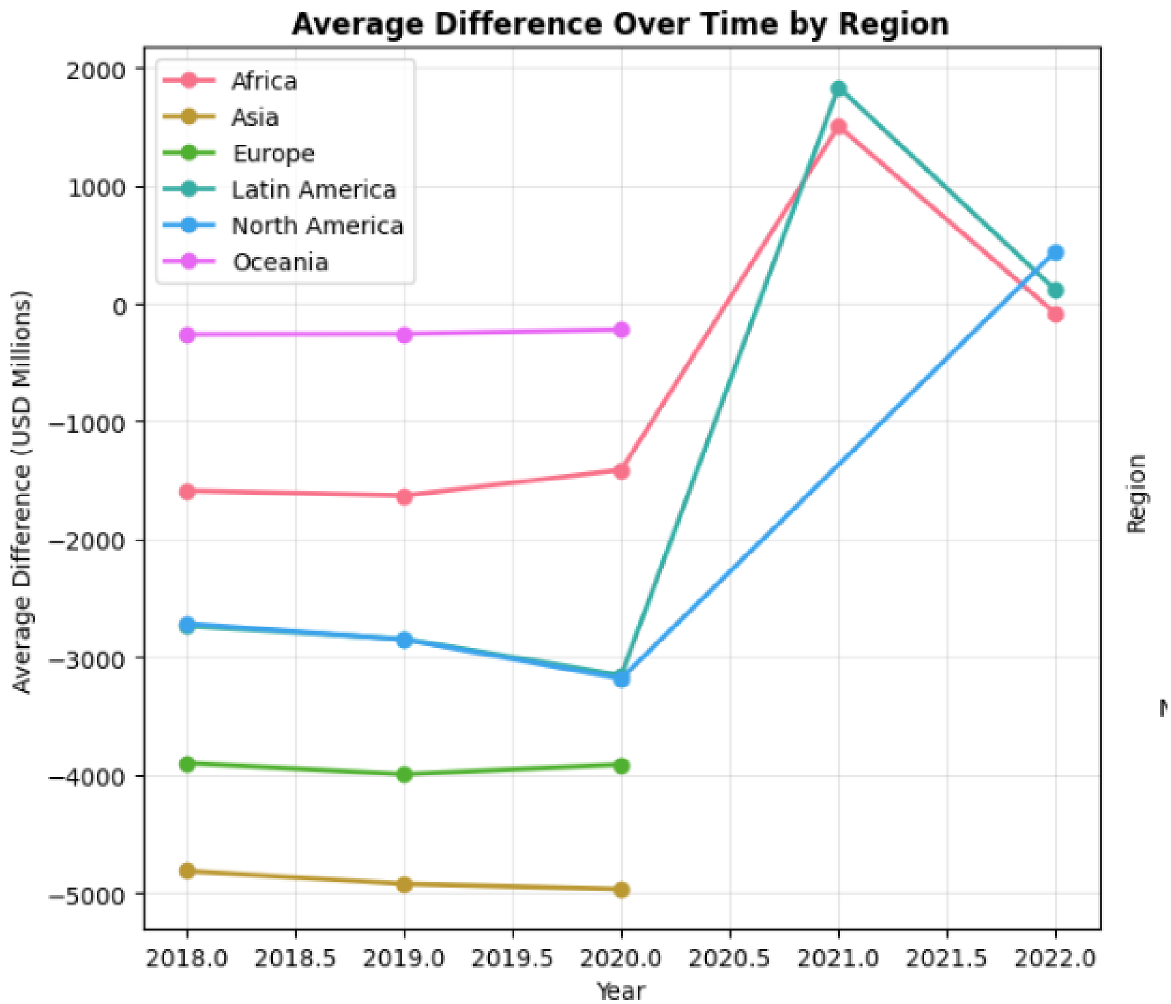

Figure 9: Time Series Differences

- Combined year and region analysis is shown in Figure 10

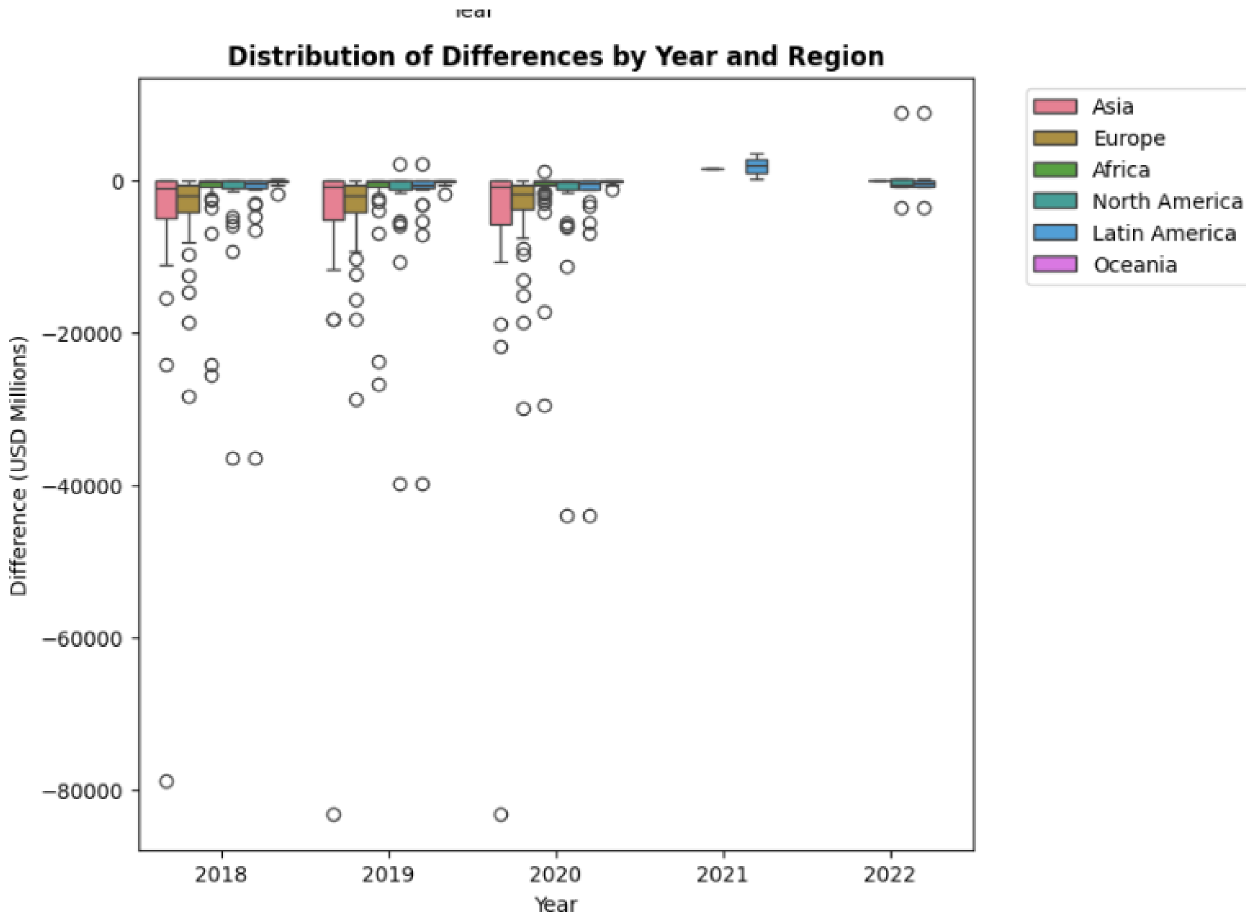

Figure 10: Year and Region Analysis

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 11 :

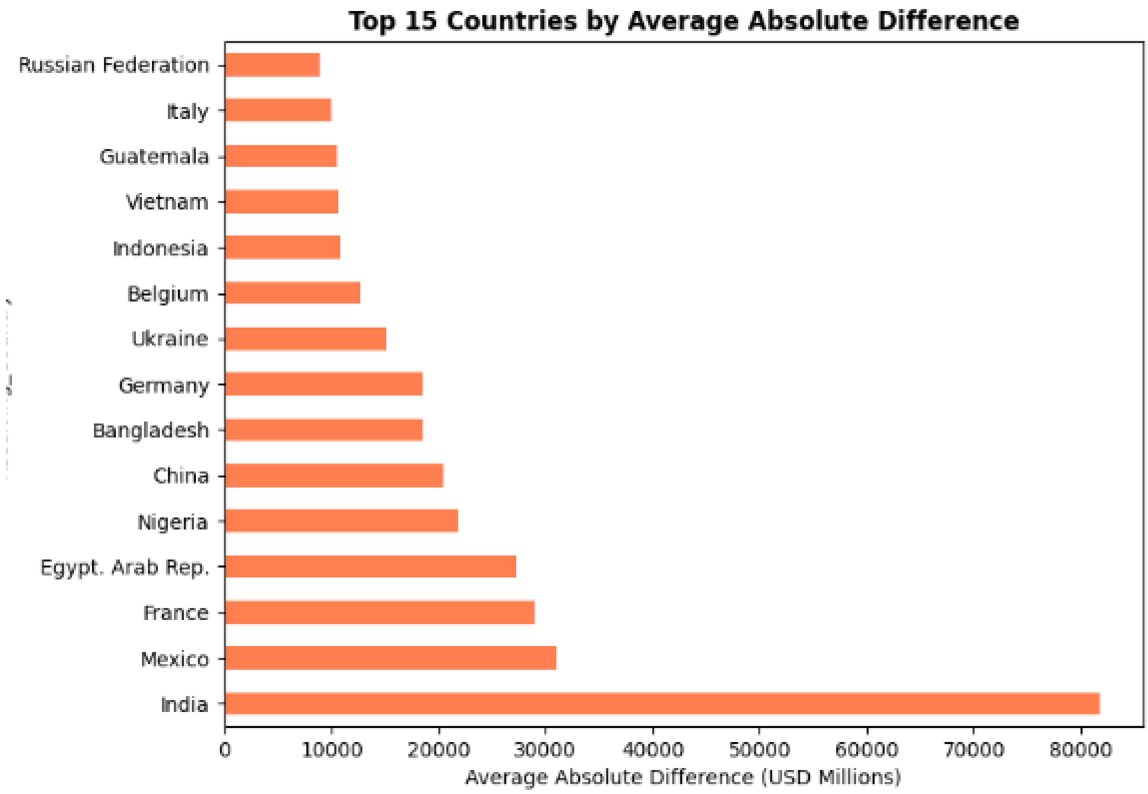

Figure 11: Country-Level Discrepancies

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 12.

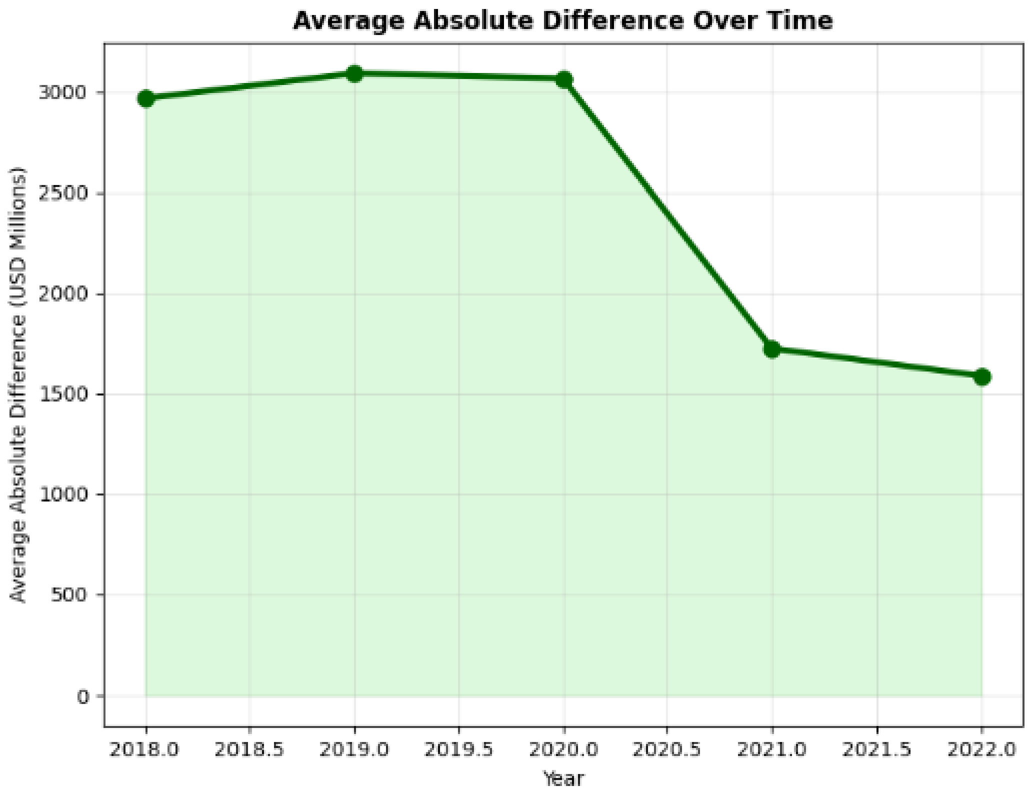

Figure 12: Average Country Discrepancies

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 13 and Figure 14.

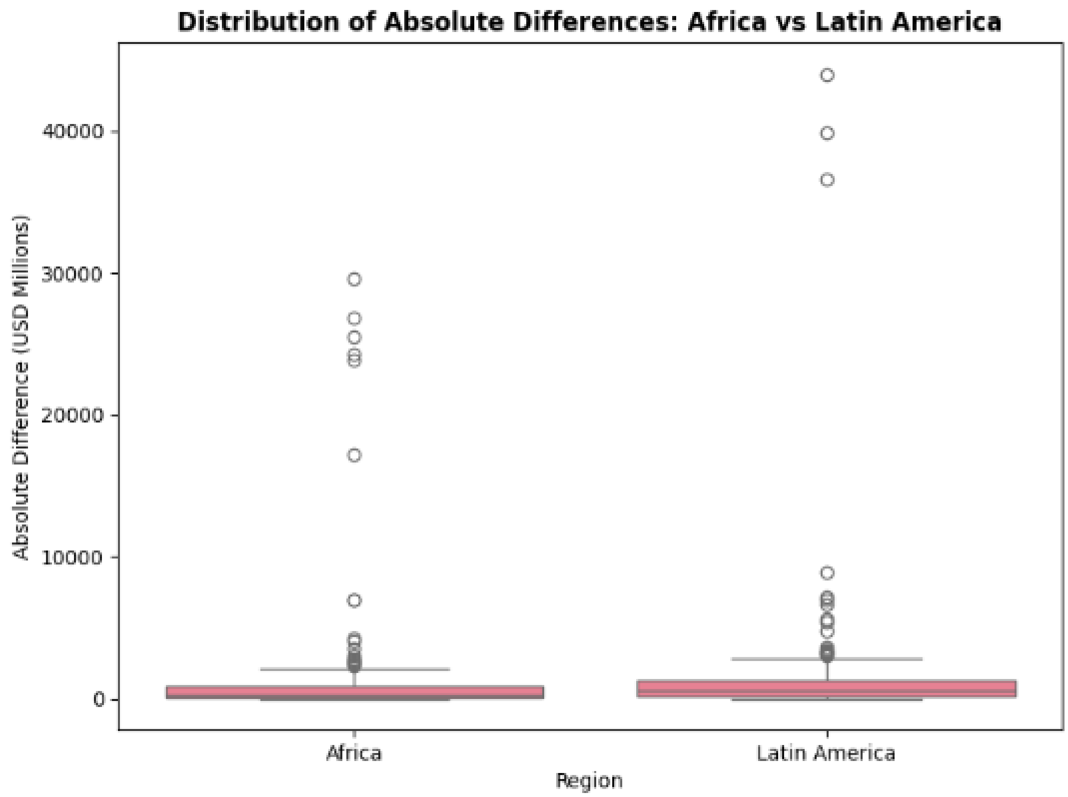

Figure 13: Latin America Focus 1

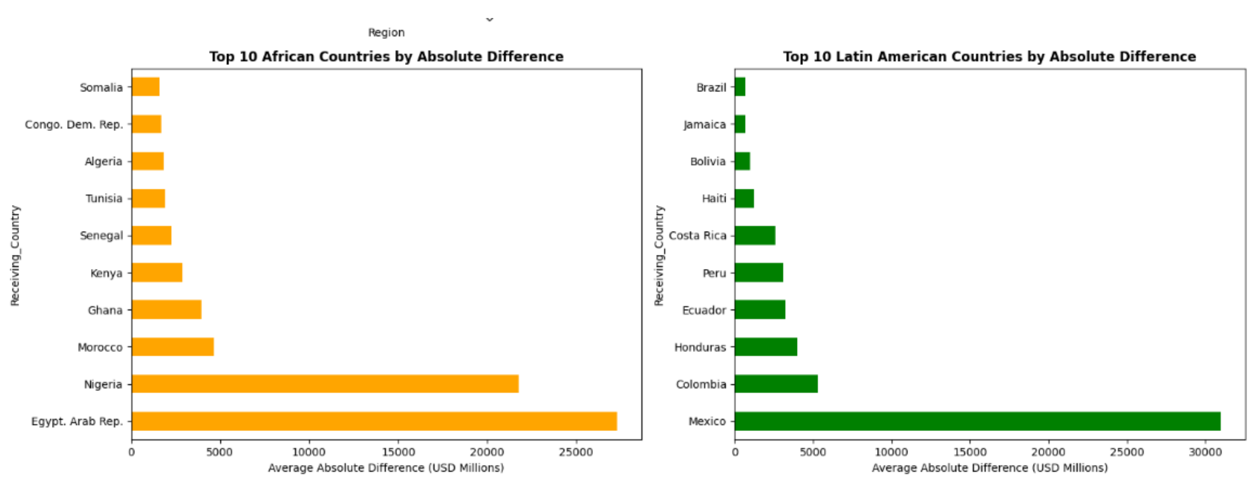

Figure 14: Latin America Focus 2

To further refine the analysis, we remove prominent outliers such as Egypt, Nigeria, and Mexico. The impact of this adjustment is shown in Figure 15 and Figure 16.

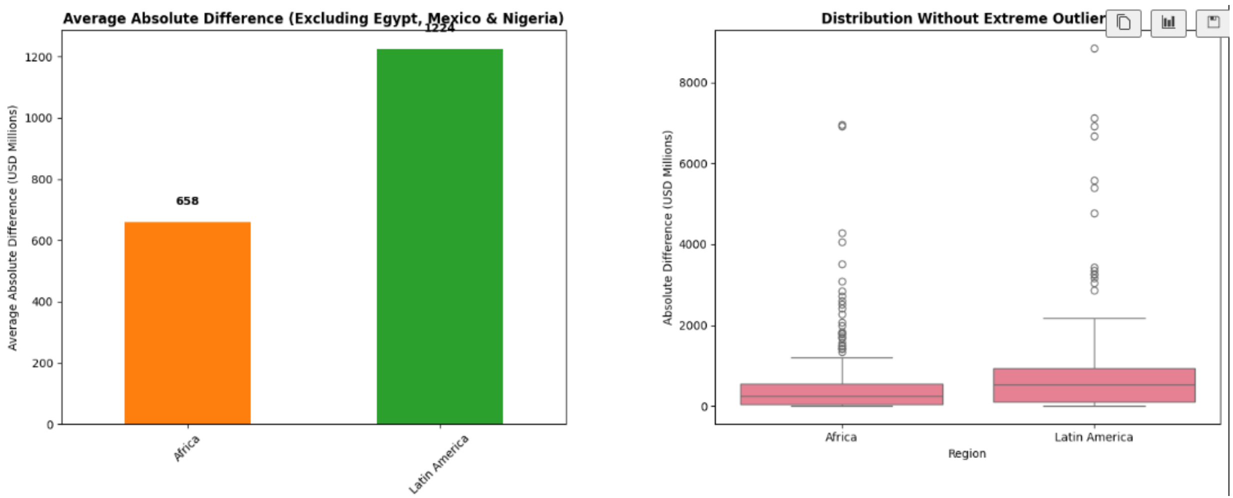

Figure 15: Outliers Removed 1

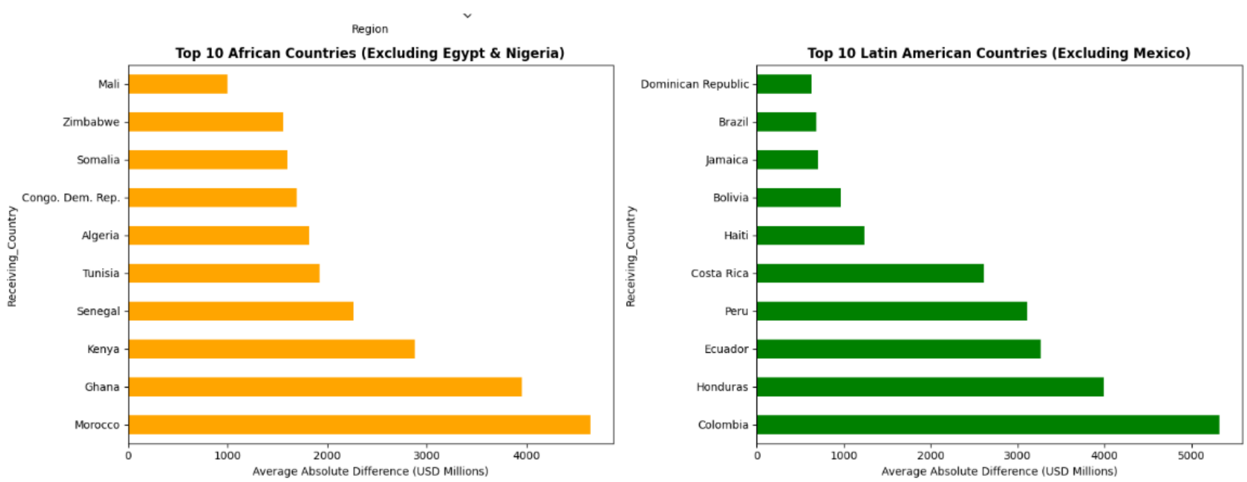

Figure 16: Outliers Removed 2

Despite these adjustments, the discrepancies remain substantial, often exceeding $1 billion USD. A country-level isolation of the data is provided in Figure 17.

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.

3 Empirical Analysis: Drivers of Remittances

With the corrected dataset, I proceed to analyze the determinants of remittance flows through regression analysis. Real GDP is used to account for inflation, while remittance values are expressed in nominal terms. Outliers were identified and removed using the interquartile range (IQR) method, specifically excluding observations below the 25th percentile minus 1.5 times the IQR or above the 75th percentile plus 1.5 times the IQR.

3.0.1 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 or thousands USD

The regression results are presented in Table 1 below:

Table 1: All Three Specifications

	Remit. (thousands)		**Log(Remit. millions)**		**Remit. (millions)**	
	W/ Outliers Spec 1 (1)	W/o Outliers Spec 1 (2)	W/ Outliers Spec 2 (3)	W/o Outliers Spec 2 (4)	W/ Outliers Spec 4 (5)	W/o Outliers Spec 4 (6)
—	—	—	—	—	—	—
Sending GDP (millions)	0.164**	0.138***				
	(0.071)	(0.040)				
Receiving GDP (millions)	-0.061	-0.034				
	(0.112)	(0.063)				
Log(Sending GDP)			0.959***	0.955***	420.9***	269.5***
			(0.026)	(0.026)	(94.8)	(53.5)
Log(Receiving GDP)			0.561***	0.561***	67.8	54.1
			(0.033)	(0.033)	(120.3)	(67.9)
Constant	691,496***	434,212***	-17.49***	-17.45***	-4,583***	-3,032***
	(247,445)	(139,546)	(0.474)	(0.473)	(1,740)	(982)
—	—	—	—	—	—	—
Obser.	2,837	2,835	2,837	2,835	2,837	2,835
R^2	0.002	0.004	0.370	0.370	0.007	0.009
Adjusted R^2	0.001	0.004	0.370	0.369	0.006	0.008
Residual Std. Error	12,261,774	6,913,052	3.334	3.324	12,231	6,896
	(df = 2834)	(df = 2832)	(df = 2834)	(df = 2832)	(df = 2834)	(df = 2832)
F Statistic	2.855*	6.088***	832.3***	830.2***	10.01***	13.00***
	(df = 2; 2834)	(df = 2; 2832)	(df = 2; 2834)	(df = 2; 2832)	(df = 2; 2834)	(df = 2; 2832)

Notes: Spec 1: Linear (thousands), Spec 2: Log-Log, Spec 4: Linear-Log. *p<0.1;* ***p<0.05;*** p<0.01. Standard errors in parentheses.

3.1 Lagged Model

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

Table 2: Remittances and Lagged GDP (t-1)

	Remit. (thousands)		**Log(Remit. millions)**		**Remit. (millions)**	
	W/ Outliers Spec 1 (1)	W/o Outliers Spec 1 (2)	W/ Outliers Spec 2 (3)	W/o Outliers Spec 2 (4)	W/ Outliers Spec 4 (5)	W/o Outliers Spec 4 (6)
—	—	—	—	—	—	—
Sending GDP (t-1, millions)	0.072***	0.072***				
	(0.005)	(0.005)				
Receiving GDP (t-1, millions)	0.002	0.002				
	(0.007)	(0.007)				
Log(Sending GDP (t-1))			0.930***	0.930***	63.746***	63.746***
			(0.034)	(0.034)	(6.451)	(6.451)
Log(Receiving GDP (t-1))			0.694***	0.694***	27.370***	27.370***
			(0.041)	(0.041)	(7.755)	(7.755)
Constant	42,784***	42,784***	-18.504***	-18.504***	-897.360***	-897.360***
	(16,033)	(16,033)	(0.602)	(0.602)	(113.856)	(113.856)
—	—	—	—	—	—	—
Obser.	1,467	1,467	1,467	1,467	1,467	1,467
R^2	0.121	0.121	0.410	0.410	0.069	0.069
Adjusted R^2	0.120	0.120	0.409	0.409	0.068	0.068
Residual Std. Error	573,098	573,098	3.118	3.118	589.692	589.692
	(df = 1464)	(df = 1464)	(df = 1464)	(df = 1464)	(df = 1464)	(df = 1464)
F Statistic	100.756***	100.756***	508.309***	508.309***	54.548***	54.548***
	(df = 2; 1464)	(df = 2; 1464)	(df = 2; 1464)	(df = 2; 1464)	(df = 2; 1464)	(df = 2; 1464)

Notes: Spec 1: Linear (thousands), Spec 2: Log-Log, Spec 4: Linear-Log. GDP variables lagged by 1 period (t-1). *$p<0.1$;* ***$p<0.05$;*** $p<0.01$. Standard errors in parentheses.

3.2 Lagged GDP Per Capita Model

To reduce concerns about reverse causality and address the interpretational challenges with absolute GDP levels, we re-estimate our models using lagged GDP per capita variables: We also redefine our utlier variable to remove 5th and 95th percentile.

Table 3: Remittances and Lagged GDP Per Capita (t-1)

	Remit. (thousands)		**Log(Remit. millions)**		**Remit. (millions)**	
	W/ Outliers Spec 1 (1)	W/o Outliers Spec 1 (2)	W/ Outliers Spec 2 (3)	W/o Outliers Spec 2 (4)	W/ Outliers Spec 4 (5)	W/o Outliers Spec 4 (6)
—	—	—	—	—	—	—
Sending GDP Per Capita (t-1, USD)	4.734***	4.734***				
	(0.699)	(0.699)				
Receiving GDP Per Capita (t-1, USD)	-0.845	-0.845				
	(1.185)	(1.185)				
Log(Sending GDP Per Capita (t-1))			1.529***	1.529***	85.960***	85.960***
			(0.073)	(0.073)	(12.243)	(12.243)
Log(Receiving GDP Per Capita (t-1))			0.144	0.144	-38.180**	-38.180**
			(0.104)	(0.104)	(17.581)	(17.581)
Constant	27,486.240	27,486.240	-15.615***	-15.615***	-347.266*	-347.266*
	(23,165.860)	(23,165.860)	(1.173)	(1.173)	(197.355)	(197.355)
—	—	—	—	—	—	—
Obser.	1,472	1,472	1,472	1,472	1,472	1,472
R^2	0.032	0.032	0.232	0.232	0.038	0.038
Adjusted R^2	0.031	0.031	0.231	0.231	0.037	0.037
Residual Std. Error	600,276.400	600,276.400	3.556	3.556	598.410	598.410

	Remit. (thousands)		Log(Remit. millions)		Remit. (millions)	
	(df = 1469)	(df = 1469)	(df = 1469)	(df = 1469)	(df = 1469)	(df = 1469)
F Statistic	24.632***	24.632***	221.589***	221.589***	29.376***	29.376***
	(df = 2; 1469)	(df = 2; 1469)	(df = 2; 1469)	(df = 2; 1469)	(df = 2; 1469)	(df = 2; 1469)

Notes: Spec 1: Linear (thousands), Spec 2: Log-Log, Spec 4: Linear-Log. GDP Per Capita variables lagged by 1 period (t-1). *p<0.1;* ***p<0.05;*** p<0.01. Standard errors in parentheses.

These results indicate that, although the coefficient for receiving country GDP is not strictly negative, it is consistently lower than that of the sending country. This finding suggests that countries receiving remittances tend to have lower GDP overall, consistent with the notion that remittance flows are directed toward less affluent economies. In the following section, I address potential issues of non-stationarity and upward trends in both GDP and remittance series by considering first-difference specifications.

3.3 First-Difference GDP Models

To address potential trending behavior in both GDP and remittance series, we estimate first-difference models that examine changes in remittances as a function of changes in GDP. This approach helps remove persistent upward trend.

3.3.1 Level Specification Models

We begin with level specifications where changes in remittances (in thousands USD) are regressed on changes in GDP per capita:

Table 4: Change in Remittances Models - Level Specification

	ΔRemit (000s USD)			
	Cont.+Out (1)	**Cont.-Out (2)**	**Lag+Out (3)**	**Lag-Out (4)**
—	—	—	—	—
ΔGDP/cap Send	0.1205	2.6539*		
	(0.9827)	(1.3837)		
ΔGDP/cap Recv	2.6243	−0.7886		
	(2.7470)	(3.0939)		
L.ΔGDP/cap Send			3.7535	2.3964
			(3.5454)	(5.5185)
L.ΔGDP/cap Recv			−20.3995	0.3755
			(12.5156)	(12.0929)
Constant	5,003.8290**	423.9368	5,757.2720	−1,999.8670
	(2,391.0170)	(866.5253)	(4,179.6280)	(2,797.0050)

	ΔRemit (000s USD)			
—	—	—	—	—
Observations	1,467	1,067	750	550
R^2	0.0007	0.0040	0.0040	0.0003
Adjusted R^2	−0.0007	0.0021	0.0013	−0.0033

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.*

3.3.2 Log-Log Specification Models

Table 5: Change in Remittances Models - Log-Log Specification

	Δlog(Remit)			
	Cont.+Out (1)	**Cont.-Out (2)**	**Lag+Out (3)**	**Lag-Out (4)**
—	—	—	—	—
Δlog(GDP/cap Send)	0.4191	0.1655		
	(0.3428)	(0.7112)		
Δlog(GDP/cap Recv)	1.1349***	0.7387		
	(0.3939)	(0.5833)		
L.Δlog(GDP/cap Send)			−0.6308	0.9403
			(0.7759)	(1.4655)
L.Δlog(GDP/cap Recv)			−2.0513**	3.1176*
			(0.8764)	(1.6508)
Constant	−0.0178	0.0128	−0.0587	−0.2065***
	(0.0258)	(0.0287)	(0.0400)	(0.0763)
—	—	—	—	—
Observations	1,467	1,120	750	557
R^2	0.0112	0.0029	0.0109	0.0068
Adjusted R^2	0.0099	0.0011	0.0083	0.0032

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.*

3.3.3 Alternative GDP Specifications

3.3.3.1 Total GDP Models

Table 6: Change in Remittances Models - Level Specification (GDP Total)

	ΔRemit (000s USD)			
	Cont.+Out (1)	**Cont.-Out (2)**	**Lag+Out (3)**	**Lag-Out (4)**
—	—	—	—	—
ΔGDP Send	0.0373	0.3407***		
	(0.0265)	(0.0780)		

	ΔRemit (000s USD)			
ΔGDP Recv	0.0371	0.0075		
	(0.0305)	(0.0464)		
L.ΔGDP Send			0.1320***	−0.3050
			(0.0431)	(0.1860)
L.ΔGDP Recv			−0.0607	−0.0630
			(0.0454)	(0.1493)
Constant	4,541.9500*	27.7142	2,733.6210	81.5042
	(2,336.8530)	(796.6574)	(3,907.2870)	(2,146.8610)
—	—	—	—	—
Observations	1,467	1,120	750	554
R^2	0.0025	0.0180	0.0146	0.0051
Adjusted R^2	0.0012	0.0163	0.0120	0.0015

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.*

Table 7: Change in Remittances Models - Log-Log Specification (GDP Total)

	Δlog(Remit)			
	Cont.+Out (1)	**Cont.-Out (2)**	**Lag+Out (3)**	**Lag-Out (4)**
—	—	—	—	—
Δlog(GDP Send)	0.3836	0.0210		
	(0.3446)	(0.7112)		
Δlog(GDP Recv)	1.1890***	0.9777*		
	(0.3955)	(0.5925)		
L.Δlog(GDP Send)			−0.5063	0.3709
			(0.8275)	(1.6071)
L.Δlog(GDP Recv)			−2.2794**	2.4886
			(0.9356)	(1.7842)
Constant	−0.0298	0.0021	−0.0337	−0.1911**
	(0.0251)	(0.0279)	(0.0444)	(0.0921)
—	—	—	—	—
Observations	1,467	1,103	750	550
R^2	0.0117	0.0039	0.0119	0.0037
Adjusted R^2	0.0103	0.0021	0.0092	0.0001

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.*

3.3.3.2 Linear-Log Specifications

The linear-log models examine how percentage changes in GDP affect absolute changes in remittances:

Table 8: Change in Remittances Models - Linear-Log Specification

	ΔRemit (1000s)	
	Cont.+Out (1)	**Cont.-Out (2)**
—	—	—
Δlog(GDP/cap) Send	−5,724.3720	24,112.9400
	(32,287.6700)	(21,794.2100)
Δlog(GDP/cap) Recv	52,138.2200	6,233.4400
	(37,098.2000)	(17,876.3200)
Constant	5,362.9530**	131.5970
	(2,425.7380)	(878.2345)
—	—	—
Observations	1,467	1,120
R^2	0.0015	0.0026
Adjusted R^2	0.0002	0.0008

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.*

Table 9: Change in Remittances Models - Linear-Log Specification (GDP Total)

	ΔRemit (1000s)	
	Cont.+Out (1)	**Cont.-Out (2)**
—	—	—
Δlog(GDP) Send	−5,577.1460	11,784.8800
	(32,455.5100)	(21,933.1700)
Δlog(GDP) Recv	58,509.2600	14,008.5700
	(37,253.1600)	(18,273.8800)
Constant	5,049.8510**	−237.2238
	(2,366.4090)	(861.3333)
—	—	—
Observations	1,467	1,103
R^2	0.0020	0.0019
Adjusted R^2	0.0006	0.0001

Note: p<0.1; **p<0.05;** *p<0.01. Standard errors in parentheses.**

3.4 Extensions

Additional control variables may be added.

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