Economic Insights and Predictive Modeling of Personal Remittances in Saudi Arabia

Noor Alawlaqi
Effat University, Jeddah, Saudi Arabia
nolalwlaqi@effat.edu.sa

Maha Almashharawi
Effat University, Jeddah, Saudi Arabia
maaalmashharawi@effat.edu.sa

Mashael Alsalamah
Effat University, Jeddah, Saudi Arabia
Maoalsalamah@effat.edu.sa

Dr.Zain Balfagih
Effat University, Jeddah, Saudi Arabia
zbalfagih@effatuniversity.edu.sa

Abstract

In this project, we explored the relationship between economic indicators, including remittances, GDP growth, and migration, and their impact on global economic growth. Using data from key World Bank indicators from 1990 to 2020, we employ machine learning models—Linear Regression, Random Forest, and Gradient Boosting—to predict remittance flows based on these factors. We evaluate the models' performance using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). Our findings suggest that the Gradient Boosting model performs the best in predicting remittance flows, providing valuable insights for policymakers and researchers looking to understand the dynamic interaction between remittances and economic growth. The study also highlights the challenges of accounting for missing data and the complexity of the relationships between remittances and economic indicators.

1 Introduction

The global economy is influenced by a variety of factors—economic policies, financial flows, and demographic changes. Among these, remittances have emerged as a significant external financing source, particularly for developing countries. While remittances are known to contribute to economic growth, the extent and nature of this contribution remain debated. This paper examines the role of remittances in shaping economic outcomes, with a focus on understanding how they interact with other economic indicators like GDP growth, unemployment, and migration patterns.

We employ machine learning techniques to predict remittance flows, aiming to provide a clearer picture of how these financial transfers relate to broader economic trends. Our study uses data from key World Bank indicators, covering 1990 to 2020, and tests three machine learning models—Linear Regression, Random Forest, and Gradient Boosting. By analyzing the performance of these models, we aim to identify the most effective approach to forecasting remittance flows and better understand the complex relationships between remittances and economic indicators. Our approach offers a practical tool for policymakers and researchers who wish to leverage this data for economic forecasting.

2 Prior Literature

Economic growth is shaped by various factors, including financial flows, employment patterns, and environmental conditions. Remittances, as an important source of external financing, significantly influence economic performance, especially in developing countries. Studies have shown a positive relationship between remittances and growth, though the impact varies based on regional and economic contexts (Cazachevici et al., 2020). In countries like Saudi Arabia, remittance outflows are influenced by factors such as GDP and foreign employment, with policy measures like expatriate levies affecting these flows (Javid and Hasanov, 2023). Furthermore, remittances can have environmental implications, with their effects on carbon emissions differing across income levels, contributing to sustainable development in certain regions (Elbatanony et al., 2021). These factors collectively shape our understanding of economic growth and its drivers.

2.1 Remittances and Economic Growth

(Cazachevici et al., 2020) analyzed 95 studies to evaluate the effect of remittances on economic growth. The study found that the effect of remittances on economic growth varied widely across studies. About 40% of the studies showed a positive relationship, 20% showed a negative one, and the remaining 40% showed no significant effect. This variation is important for our research because we are analyzing how remittances influence economic growth alongside other factors like inflation and unemployment. They also emphasized the importance of controlling for other financial flows, such as FDI and foreign aid, and addressing endogeneity issues in research. This insight helped us in our project, as we needed to account for these complexities when building our model. For instance, we must ensure that we don't overestimate the effect of remittances by ignoring other influencing factors.

2.2 Remittances and Economic Activity

(Javid and Hasanov, 2023) examined factors influencing remittance outflows from Saudi Arabia. The study highlighted the role of GDP and foreign employment in determining remittance flows. Their study found that higher foreign employment and economic activity were associated with higher remittance outflows. This relationship is highly relevant to our research, as it suggests that an increase in migrant workers, which is common in countries like Saudi Arabia, would likely lead to higher remittances. Moreover, They discussed how remittances could act as a "leakage" in the economy, reducing the positive impacts of financial policies on growth.

2.3 Environmental Impacts of Remittances

(Elbatanony et al., 2021) introduced the environmental implications of remittances, showing how remittances could play a role in promoting cleaner development in lower-income countries. This perspective will be useful in our project, as we aim to create a model that not only predicts economic growth but also considers sustainability factors. (Elbatanony et al., 2021), explored how remittance inflows affected environmental outcomes, particularly CO_2 emissions. The authors found that remittances reduced CO_2 emissions in low- and middle-income countries, especially where emissions were high. This could be important for our project as we consider how economic growth might be influenced not only by economic factors but also by environmental sustainability.

This review provided critical insights that directly influenced our research on predicting economic growth using remittances and other key economic indicators. By integrating these insights, we are better equipped to predict how remittances, alongside other economic indicators such as inflation, unemployment, and FDI, interact and affect economic growth.

3 Data

We used data from six key World Bank indicators: international migrant remittances received and sent (Bank, 2024d) (Bank, 2024c), total population (Bank, 2024e), annual GDP growth rate (Bank, 2024a), unemployment rate (Bank, 2024f), and net migration (Bank, 2024b). These indicators were chosen to analyze global economic and migration trends. Each dataset was loaded and processed using pandas. We melted the data, transforming year columns into rows, and filtered it to include only data from 1990 onward. We dropped missing values and ensured the 'Year' column was in numeric format for consistency.

After processing, we pivoted the data to create separate columns for each indicator. The 'Country Name', 'Country Code', and 'Year' served as the index, with each indicator as a separate column. Once all datasets were processed, we merged them into one. The final dataset, after pivoting, had a size of (9010, 9), with 9010 observations and 9 variables, including the country and year. This dataset was then used to analyze global migration and economic trends.

4 Models

In this study, we employed three machine learning models to predict Personal Remittances, Received based on various economic indicators. These models were selected for their ability to capture different types of relationships between the predictors and the target variable.

Before modeling, we applied KNN imputation to handle missing values in the dataset. This method uses the values from the nearest neighbors to impute missing data points. It's a model-based approach that assumes similar data points have similar values. We chose KNN imputation because it preserves the relationships between variables and works well with numerical data where patterns can be learned from neighboring values.

Linear regression was used as a baseline model to predict the target variable based on a linear relationship with the input features. It provides a simple way to assess the relationship between the variables and is easy to interpret.

Random Forest is an ensemble learning method that creates a "forest" of decision trees. Each tree in the forest is trained on a random subset of the data, and the final prediction is the average of all the trees' predictions. Random Forest is a powerful model for capturing complex, non-linear relationships between the features and the target. It is particularly effective when dealing with high-dimensional datasets with many features, as it can handle interactions between features without requiring explicit specification. In this study, Random Forest was chosen for its flexibility and ability to model complex patterns in the data.

Gradient Boosting is another ensemble method, but unlike Random Forest, it builds decision trees sequentially. Each tree corrects the errors made by the previous one, which helps improve predictive accuracy over time. This model is often more accurate than Random Forest, especially for datasets where relationships between features are non-linear and require iterative improvement. Gradient Boosting is known for its strong performance in regression tasks, and it was included in this study to test whether it could outperform Random Forest in predicting remittance flows.

5 Methods

The approach we followed for this study was focused on thorough data preprocessing, exploratory analysis, and predictive modeling, to predict Personal Remittances, Received for a certain country.

5.1 Data Preprocessing and Cleaning

We started by loading the datasets from multiple World Bank indicators. These datasets contained several economic and demographic variables for various countries. Each dataset was cleaned and transformed to ensure consistency across all indicators. First, we melted the data, which reshaped the datasets from a wide format into a long format, allowing us to handle each year's data as a separate observation. This transformation made it easier to analyze time-based trends for each country.

After melting, we filtered the data to include only records from the year 1990 onwards. This decision was driven by the need to focus on more recent data, ensuring that the trends we were analyzing were relevant and reflective of current economic conditions. The dataset was then cleaned by removing rows with missing values, which could introduce noise or bias into our analysis.

Next, we pivoted the data so that each indicator had its own column, while "Country Name," "Country Code," and "Year" were set as the index. This restructuring was essential to align the various economic indicators for each country across years. Once this was done, we merged the different datasets into a single, unified DataFrame, ensuring that all indicators were available for analysis and modeling.

One of the challenges we faced was dealing with missing values in the dataset. Rather than simply removing rows with missing data, which could lead to a loss of valuable information, we used KNN Imputation to fill in missing values. This method estimates the missing data points based on the similarity of neighboring data points. We selected 5 neighbors for imputation, ensuring that the missing values were estimated in a way that preserved the overall structure of the data.

In addition to imputing missing values, we engineered new features that we believed could provide more insight into the remittance trends. These included:

- Remittance Growth Rate, which tracked the year-over-year change in remittance amounts.
- Remittance per Capita, calculated by dividing total remittances by the population.
- Remittance Volatility, which measured the variation in remittance amounts over time.
- Remittance-to-GDP Ratio, a key indicator of the importance of remittances to the economy.

5.2 Exploratory Data Analysis

With the data in its final form, we moved on to the exploratory analysis phase. We aimed to understand the underlying patterns and relationships within the data before applying any machine learning models. First, we calculated summary statistics to get a sense of the distribution of key variables. This gave us insights into the central tendency, spread, and presence of any potential outliers or anomalies in the data.

We then created visualizations to better understand the trends over time. Line plots were used to observe the trends of key indicators like GDP Growth, Population, and Remittances Received. These visualizations helped us identify any clear patterns or seasonality in the data. To explore relationships between indicators, we generated correlation heatmaps, which revealed the strength of linear associations between different variables. Scatter plots were also used to further investigate how Personal Remittances related to other economic indicators like GDP Growth and Net Migration.

5.3 Model Implementation and Evaluation

Our goal was to predict Personal Remittances, Received based on a set of economic indicators. We implemented three different machine learning models to compare their performance and identify the best predictor.

The first model we applied was Linear Regression, a straightforward approach for capturing linear relationships between the target and the predictors. This model served as a baseline for our analysis. We used the default settings of fit_intercept=True and normalize=False for Linear Regression, as these settings are typically appropriate for datasets with no extreme scaling issues.

Next, we applied the Random Forest Regressor. This ensemble model was chosen for its ability to capture complex, non-linear relationships between the predictors and the target. The model was configured with 100 trees (n_estimators=100) and allowed to grow trees to their maximum depth (max_depth=None). We expected this model to handle feature interactions better than Linear Regression, particularly given the non-linear nature of economic data.

The third model tested was the Gradient Boosting Regressor. This model builds trees sequentially, with each tree learning from the mistakes of the previous one. We configured the model with 100 boosting stages (n_estimators=100), a learning rate of 0.1 (learning_rate=0.1), and a maximum depth of 3 (max_depth=3) to prevent overfitting. We chose Gradient Boosting for its ability to improve predictive performance iteratively by correcting errors.

To evaluate the models, we used a split of 80% training data and 20% testing data. The models were trained on the training set and evaluated on the testing set. For each model, we calculated the following performance metrics:

Mean Absolute Error (MAE) to measure the average magnitude of errors in the predictions. Root Mean Squared Error (RMSE) to give more weight to larger errors and assess overall prediction accuracy. R-Squared (R^2) to determine the proportion of variance in the target variable explained by the model.

6 Results and Discussion:

The heatmap shown in Figure 1 displays the correlation between various economic indicators, such as population, GDP growth, unemployment, and remittances. Strong positive correlations are seen between total population and remittances received (+0.60) and between remittances paid and remittances received (+0.51). This suggests that larger populations tend to receive more remittances, and countries paying remittances also often receive them. Weak to moderate correlations are observed between other indicators, highlighting how different economic factors are interrelated.

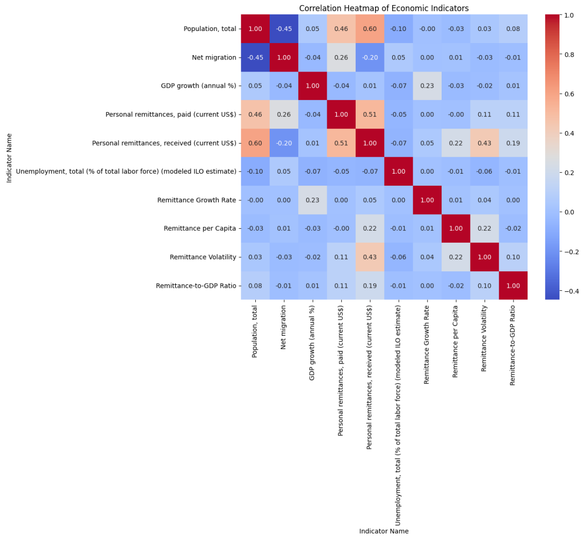

Figure 1: This heatmap shows the correlation between various economic indicators, including population, GDP growth, unemployment, and remittances. Positive correlations are observed between population and remittances received, and between remittances paid and remittances received.

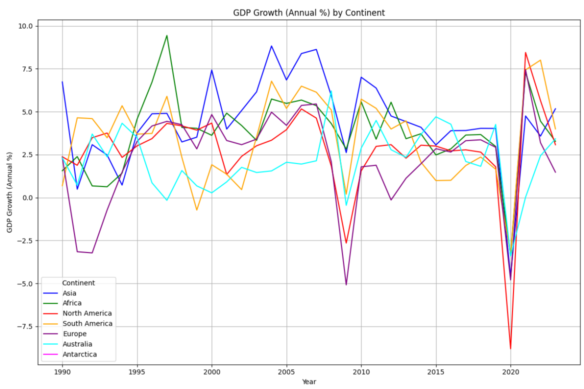

Figure 2: The plot shows the trend of GDP growth (annual %) across different continents over time. Each line represents a continent.

The line plot in Figure 2 shows the annual GDP growth (%) for each continent over the years. Asia exhibits the highest and most consistent growth, with dips during the Asian financial crisis (1997-

1998) and the global financial crisis (2008), followed by a sharp decline in 2020 due to the COVID-19 pandemic. Africa and South America show more volatility, with fluctuations likely tied to economic instability and external factors. North America and Europe display steady but lower growth, with noticeable declines in 2008 and 2020. Australia maintains stable, moderate growth, while Antarctica shows no significant economic activity.

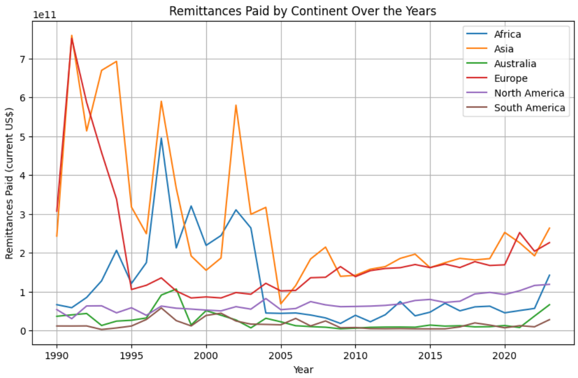

Figure 3: This plot tracks the total remittances paid by each continent from 1990 to 2020. Asia saw a gradual decline after the 1990s, while Europe showed a steady increase, surpassing Asia in the mid-2000s. North America's growth was slower but steady. Africa, South America, and Australia had minimal contributions, with stable trends over time.

The plot seen in Figure 3 shows the total remittances paid by continent from 1990 to 2020. Asia had the highest remittances paid in the 1990s, but from around 2005 onward, there was a gradual decline, with Asia's total remittances remaining consistently lower than in the 1990s. Europe, on the other hand, shows a steady and continuous increase, overtaking Asia's remittance outflows in the mid-2000s. North America also shows gradual growth, though at a slower rate compared to Europe. Africa, South America, and Australia have low, stable levels of remittances, with no significant changes or growth. The growth rate of remittances from all continents appears to slow slightly toward the end of the period (around 2020), but there is no sharp drop or noticeable dip.

Figure 4 illustrates the year-over-year incremental changes in remittances received globally from 1990 to 2020. The plot highlights periods of significant growth, such as the early 1990s and the mid-2000s, as well as sharp declines, particularly around 2005. In contrast, the period between 2010 and 2020 shows smaller fluctuations, suggesting

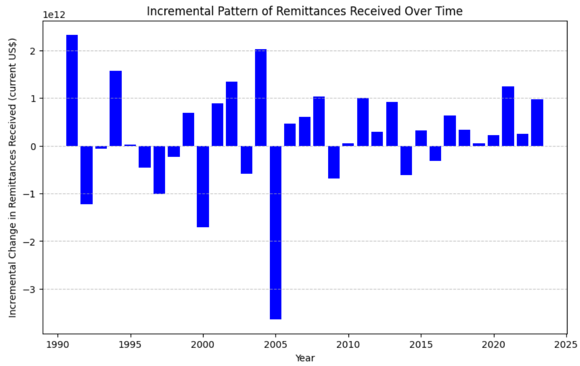

Figure 4: This bar plot tracks year-over-year changes in remittances received globally, highlighting periods of growth, such as the 1990s and early 2020s, and sharp declines, such as in 2005.

a more stable trend in remittance inflows. This visualization provides insights into the variability and key turning points in global remittance patterns over time.

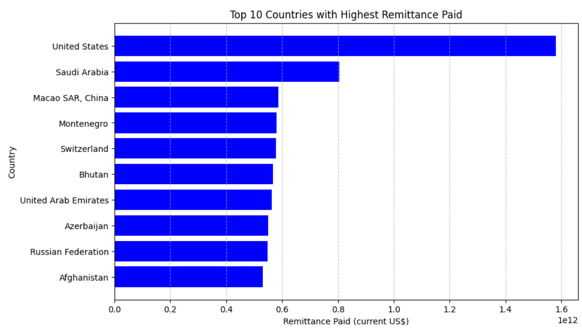

Figure 5: This horizontal bar plot shows the top 10 countries by total remittances paid, with the largest contributors placed at the top.

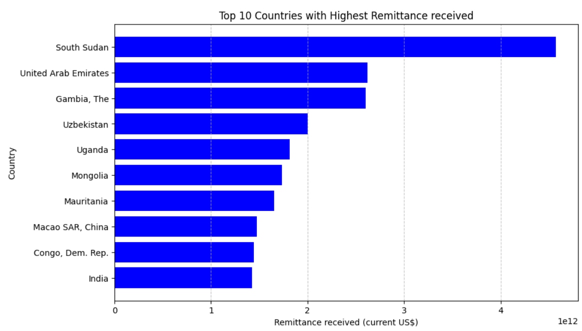

Figure 6: This horizontal bar plot displays the top 10 countries by total remittances received.

Figures 5 and 6 compare the top 10 countries by remittances paid and received, respectively. Figure 5 highlights the countries with the highest remittance outflows, while Figure 6 focuses on those with the highest remittance inflows.

Saudi Arabia appears in the top 10 countries with the highest remittance paid, reflecting its significant role as a host for migrant workers. However, it is absent from the top 10 countries receiving remittances, indicating that Saudi Arabia primarily acts as a sender rather than a recipient in the global remittance cycle.

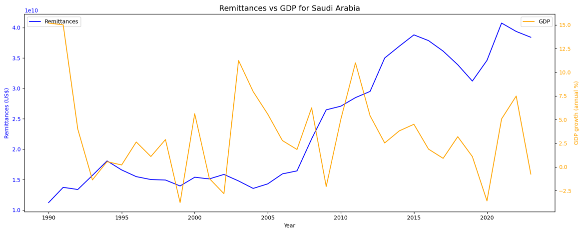

Figure 7: Trend of Remittances Paid and GDP Growth in Saudi Arabia (1990-2020).

The plot in Figure 7 reveals an interesting relationship between remittances paid and GDP growth in Saudi Arabia. While remittances show a steady increase over time, reflecting the growing number of migrant workers and their remittance outflows, GDP growth fluctuates significantly, with sharp declines in 1997, 2008, and 2020. These declines in GDP growth do not directly correspond to the trend in remittances, which continues to rise despite economic downturns. This suggests that while GDP growth is affected by external factors like oil prices or global crises, remittance outflows are driven more by the presence of migrant workers and less by short-term economic fluctuations. The two variables do not show a clear, consistent correlation, indicating that the factors influencing remittances and GDP growth in Saudi Arabia may be largely independent.

Model	MAE	R-squared	RMSE
Linear Regression	22,581,836.08	0.9934	23,171,064.01
Random Forest	25,539,693.05	0.9810	39,367,267.88
Gradient Boosting	16,557,204.53	0.9944	21,337,922.47

Table 1: Model Performance Results

The models were evaluated based on three performance metrics: Mean Absolute Error (MAE), R-squared (R^2), and Root Mean Squared Error (RMSE). Each of these metrics provides a different perspective on the model's predictive accuracy, with MAE reflecting the average prediction error, R^2 indicating the proportion of variance explained by the model, and RMSE penalizing larger errors more heavily. Below are the results for each model:

Linear Regression The Linear Regression model achieved an R-squared value of 0.9934, suggesting that it explained approximately 99.34% of the variance in the target variable, Personal Remittances, Received. This high R-squared value indicates that the model did a good job of capturing the overall trend in the data. However, the model also had a relatively high MAE of 22,581,836.08 and RMSE of 23,171,064.01, which suggests that while it performed well on average, its predictions were not always very accurate, particularly in cases where the actual values deviated significantly from the predicted values.

Random Forest Regressor The Random Forest model produced an R-squared of 0.9810, which, while lower than Linear Regression, still indicates that the model was able to explain 98.1% of the variance in the target variable. The MAE for this model was 25,539,693.05, and the RMSE was much higher at 39,367,267.88. This suggests that although the Random Forest model captured the general trends in the data, it struggled more with individual predictions, particularly with larger errors. The relatively high RMSE indicates that the model might have overfitted to the data, especially if it was influenced by noisy or outlier data points.

Gradient Boosting Regressor The Gradient Boosting model outperformed both the Linear Regression and Random Forest models in terms of accuracy. It achieved an R-squared value of 0.9944, meaning that it explained 99.44% of the variance in remittances, which was the highest among the three models. This high R^2 indicates that Gradient Boosting was the most accurate in capturing the relationship between the predictors and the target variable. The MAE for Gradient Boosting was 16,557,204.53, which was the lowest of all the models, indicating smaller average prediction errors. The RMSE was also the lowest at 21,337,922.47, further supporting the model's overall superior performance in terms of minimizing large prediction errors.

7 Conclusion

This study confirms the importance of remittances in shaping economic growth, with a particular emphasis on how these financial flows interact with other key economic indicators such as GDP growth, unemployment, and migration. By applying three machine learning models—Linear Regression, Random Forest, and Gradient Boosting—we demonstrate the varying degrees of predictive accuracy each model offers. The Gradient Boosting model outperforms the others, providing the most reliable predictions for remittance flows.

Our results highlight the complex and multi-faceted relationship between remittances and economic growth. While remittances can support growth, their effects can be moderated by factors like inflation and foreign direct investment, suggesting that a holistic view of the economy is essential for accurate forecasting. For future work, refining our models and incorporating additional factors, such as policy changes or environmental impacts, could further improve prediction accuracy and provide a more comprehensive understanding of global economic dynamics.

References

World Bank. 2024a. Gdp growth (annual %). Accessed: 2024-12-01.

World Bank. 2024b. Net migration. Accessed: 2024-12-01.

World Bank. 2024c. Personal remittances, paid (current us$). Accessed: 2024-12-01.

World Bank. 2024d. Personal remittances, received (current us$). Accessed: 2024-12-01.

World Bank. 2024e. Population, total. Accessed: 2024-12-01.

World Bank. 2024f. Unemployment, total (% of total labor force). Accessed: 2024-12-01.

Alina Cazachevici, Tomas Havranek, and Roman Horvath. 2020. Remittances and economic growth: A meta-analysis. *World Development*, 134:105021.

Mahmoud Elbatanony, Imed Attiaoui, Ibrahim Mohamed Ali Ali, Nahla Nasser, and Monaem Tarchoun. 2021. The environmental impact of remittance inflows in developing countries: evidence from method of moments quantile regression. *Environmental Science and Pollution Research*, 28(35):48222–48235.

Muhammad Javid and Fakhri J Hasanov. 2023. Determinants of remittance outflows: The case of saudi arabia. *OPEC Energy Review*, 47(4):320–335.