Faculty of Science
Department of Information and Computing Science

Determining the influence of natural disasters and public holidays on airtime top-up transfers

Written by

Gorantis Ilias
6743196

A thesis presented for the degree of
MSc Artificial Intelligence

First Supervisor: dr. A. A. Salah
Second Supervisor: dr. P. Yolum

In collaboration with:

Utrecht University
August 2021

Acknowledgements

There are many individuals without whom the thesis would not have been possible. First of all, obtaining my degree would not be possible with my first supervisor, Prof. dr. Albert Ali Salah. It has been more than one year of collaboration, a period that I learned a variety of different topics. He helped me with the machine learning topics, statistical approaches that I could consider for my analysis, and many more subjects. Apart from that, Albert gave me extensive feedback on how to write a proper scientific paper as well as writing tips which I found really useful. I would like to recognize the invaluable assistance and support he provided me, and I am grateful for having him as my supervisor. Moreover, I wish to show my gratitude to the second supervisor of this thesis, Prof. dr. Pinar Yolum. She provided me with useful comments and how I should structure my ideas properly. Finally, the nearly daily communication I had with Ph.D. Aydogdu Bilgecag could not be more important for both the research internship and the thesis. He encouraged me a lot with his ideas about the methodologies that I applied, the factors that we should consider, and several other topics.

During my research internship and the thesis, I collaborated with DT One company which provided me with the airtime top-ups data. I would like to thank the company for their support and the help during this period.

It is a pleasure to acknowledge my friends for their continuous support and help with their unique ways. They were my strength throughout these difficult times.

Finally, I could not close this section without expressing my deepest gratitude to my parents. I am forever indebted to them since they gave me the chance and the opportunities to achieve all my dreams and goals. I owe them a lot, and therefore I dedicate this degree to my wonderful parents.

Abstract

Remittance flows have a vast influence on the economies of many countries, primarily the developing ones. In this thesis, we utilize airtime top-up credits for 2020 received by fifteen countries. The data was obtained after collaborating with DTOne company. The aim is to determine the influence of natural disasters and public holidays on these top-up transfers. We developed a regression model to investigate the correlation among these factors. Our findings suggest that the most influenced natural hazards on the airtime top-ups are droughts, storms, floods, and epidemics. Among the results, we revealed that international migrants support their families back to their origin countries on important national or religious holidays. Moreover, ARIMAX and SARIMAX models were developed for Indonesia, Ethiopia, and the Philippines that forecast the next day's top-up credits. We observed that some models predict the peaks occurring due to holidays or natural hazards. However, there were some constraints when developing these models.

List of abbreviations

AI	Artificial Intelligence
ARIMAX	Autoregressive Integrated Moving Average with Exogenous Variables
CRED	Center of Research on the Epidemiology of Disasters
IOM	International Organization of Migration
ML	Machine Learning
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NN	Neural Network
OLS	Ordinary Least Squares
RMSE	Root Mean Square Error
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Variables

Table of Contents

1 Introduction

1.1 Problem Statement

1.1.1 International migration

Over many generations, migration is a significant topic that never got vanished and always increases. In the past years, due to the environmental change and the induced food shortages, many people are forced to leave their home countries and migrate to other countries. Apart from the climate changes, other factors are accounting for individuals migration. These are depending on the life cycle of one person, factors related to the origin and destination countries, as well as the economic conditions which represent a pivotal part in this decision [38]. According to the International Organization of Migration (IOM), 2020 report [43], only for 2019 there was an estimation of almost 272 million individuals that relocated from their home countries to others. Given that migrants consist of a vast volume of the world's population (3.5%), one could assume that they hold an exalted influence on the economy.

1.1.2 International remittances

Over the past years, many countries and companies facilitated international remittance flows. These represent the transactions applied mostly from migrants that left their country due to unemployment, human rights violations, poverty, etc. The remittances represent the money that migrants send back to their home countries to help their families [19].

According to [59], remittances have a key role in the growth and a notable impact on the economies, especially the ones of the developing countries. This finding aligns with the report of the International Organization of Migration (IOM) [43]. This increase is because migration rises rapidly (only for 2020, there was an estimation of almost 281 million individuals living outside of their home countries).

One limitation of studying the official data for remittance flows is the lack of detailed data. One could find a small portion of the actual size of remittances online, and if so, these would be aggregated mostly yearly or monthly. The official statistics do not provide information about money sent from informal transfers or digital transactions [1].

1.1.3 Airtime top-up transfers

Airtime top-up transfers are the transactions that individuals apply to send money back to their home countries via digital services. These top-ups are pre-paid mobile credits in the form of mobile data, SMS, calls, etc. To be used, there should be coordination between telecommunications in different countries, yet these are not available universally.

Previous studies stated that this kind of mobile transfer has a positive impact on the economic growth of the receiving countries [30]. Moreover, there are shreds of evidence that the international mobile transfers diminish the poverty of one nation [60] [48].

There are limited papers which research digital remittances data. Aydogdu et al. [10], explored the airtime top-up transfers in-depth. Among their findings, they stated that during Covid-19, the top-ups increased due to lockdowns. In a similar study, Shiqi [65] examined how mobile transfers were affected during Covid-19 in Kuwait and Italy.

For this thesis, we will utilize daily data of mobile credits. Our partner is a Business to Business (B2B) company called DT One, which provided airtime top-ups data for many country pairs. The company shared daily data for the top fifteen receiving countries for the year 2020, along with the sending countries of each one of them. However, the data should not be seen as a representative sample of the countries since the information we received corresponds to a limited number of the actual transactions that proceeded.

1.1.4 The effects of natural disasters on remittance flows

It is known that natural disasters harm the economic growth of the countries [36]. Previous research investigated their impact on international remittances. They ascertained that when a natural hazard occurs, the receiving remittances of the harmed countries increase [13][17]. Moreover, according to [9], the two most critical disasters are hurricanes and storms.

For this thesis, data from the Centre for Research on the Epidemiology of Disasters (CRED) will be used. They provide one of the most promising

[1] https://migrationdataportal.org/themes/remittances

datasets where there is detailed information about each crucial disaster that occurred in one country (total deaths, total homeless, etc.).

This thesis aims to analyze the impact of natural disasters on international mobile top-ups. Along with DT One and CRED data, the public holidays for 2020 will be used for each of the fifteen receiving countries that we have daily transactions.

For this research, three countries were selected as the main focus of the analysis, based on their economic situations. These are Indonesia, the Philippines, and Ethiopia. According to World Bank, they are labeled as higher than a middle-income country (MIC), MIC, and low-income country (LIC) respectively.

1.2 Research Approach

The nature of the data was challenging to interpret and discover any research questions. The volume of information received from the company was not adequate to find several insights. It was technically demanding to develop machine learning models to predict airtime top-ups, as they require lots of data to achieve worthwhile results. Despite the several obstacles, we defined three research questions. Apart from the prediction part of the thesis, we examine the correlation between the airtime top-ups and the natural disasters - public holidays.

1.2.1 Research Questions

The main research questions are:

RQ1: What is the relationship between natural disasters and/or public holidays to airtime top-ups?

RQ2: What insights can we gain about regular or irregular migration between corridors, based on airtime top-ups?

RQ3: Can we find factors useful for predicting airtime top-up transfers?

RQ1: As mentioned earlier, previous studies are showing the effects of natural disasters on remittances. However, since the availability of the airtime top-ups data cannot be obtained easily, there is a limited literature showing the influence of natural hazards on them. Consequently, it is a great opportunity to investigate this correlation utilizing this set of data.

For this research question, a regression model was used to analyze the effects of natural disasters and public holidays on airtime top-ups.

RQ2: Many corridors could prove valuable candidates and provide insights to explain the behavior of migrants. Rivaling the remittances observed after the disaster with the mobile transfers before it, we could infer valuable conclusions about individuals' behavior on how they manage and support their relationships (friends-families) after a natural hazard. That is, this research question refers to the extent that migrants apply digital remittances in the wake of a disaster and/or the occurrence of a public, national, or religious day.

RQ3: For each of the three countries mentioned previously, one prediction model was developed. ARIMAX or SARIMAX model is used in all cases, where we predict the receiving top-ups of the next day, by adding the components of natural disasters and the holidays of each origin country. The aim is to develop a reliable prediction model and determine if these factors affect the prognosticated results.

1.3 Thesis Outline

The thesis is organized as follows: in Section 2 we outline the basic definitions of migration, remittances, and airtime top-ups that are valuable for this thesis. Moreover, we review previous studies about the prediction of those transactions and the correlation with natural disasters. In Section 3, we mention and analyze all the datasets that we use for the thesis, along with the limitations and the pre-processing steps that were implemented. The next section presents the methodologies that were developed for the analysis. Section 5 corresponds to the results that elucidate the three research questions of the thesis. In the Discussion section, the limitations and the future work are discussed. Finally, in Section 7, a summary of the thesis can be found.

2 Literature Review

2.1 Migration

2.1.1 Definition, types and leading factors

As the migration topic is a pre-condition of the remittance flows and therefore profoundly correlated, it should be sensible to state the definition of migration. It represents the flows of people from one place to another, who, due to a variety of reasons had to leave from their habitual residence. The general definition of a migrant is well defined by the IOM; a migrant could be characterized as an individual that has to leave or already left his/her origin home and moved across an international border or within a state of his/her country [2].

According to IOM, there are several types of human migration such as internal, external, emigration, immigration, etc. The type of migration changes depending on the push and pull factors. Briefly, the first corresponds to socioeconomic instabilities, climate changes and natural disasters, civil wars, human rights violations, etc. On the other hand, some pull factors are; higher wages to the destination country, better living conditions, better education, and health services, etc.

Many individuals migrate as a choice, to enhance their quality of life and have more favorable economic opportunities (even if we discuss internal or international migration). Besides, the underdevelopment of a country is a crucial factor that boosts the migration to other nations [33]. However, many people move involuntarily from their homes due to unforeseen circumstances such as wars, disasters, or violations. According to United Nations, women migrants consist of almost the 50% of the international migrants [51].

For this research, we have data for international remittance flows, between different countries, and these are correlated with the international migration flows. These groups of people moving across their national borders to other countries, affect both home and host regions [15]. For the latter, according to [27], migrants pay taxes to the host country, therefore they have a crucial influence on its social security systems.

International Monetary Fund stated that the stock of migrants from Emerging Market and Developing Countries (EMDCs) towards developed

[2] https://www.iom.int/glossary-migration-2019

countries (DCs), had dramatically increased by 5% from 1990 (where it was 4%) to 9% in 2019 [23]. Moreover, the migration between EMDCs increased by 2%, and the migrants from a DC to another remained stable.

2.1.2 International migration during 2020

At the beginning of 2020, a new pandemic started spreading worldwide. The Covid-19 disease affected the individuals' lives by governmental restraints and the economy of each country. In 2020 various measures were exerted by the governments to constrain the spread of the virus. The pandemic influenced multiple factors, one of them was human migration.

As mentioned beforehand, the international migration flows were rising till 2019. Yet, for 2020 there was a drop of individuals leaving their origin countries because of Covid-19 [11]. World Bank states that many migrants working in low-skilled jobs were displaced and lost their incomes. The loss of jobs left several migrants unexposed to unemployment, compared to individuals living in their home countries. According to International Migration Outlook [55], for the first six months in 2020, there was a 46% decline of new visa/permits in many Organisation for Economic Co-operation and Development (OECD) countries. This drop in human migration to the developed countries could be explained by the measures against Covid-19 that were implemented by many countries.

2.2 Remittances and migration

Remittance flows represent the incomes that mostly migrants transfer back to the domestic households in their home countries [72], contributing to the incomes of many countries [59] [43], particularly the developing ones [26]. Stark et al. [68] noted that the family relationship is the key to explain remittance behaviors. This conclusion aligns with an earlier study [34], where Johnson and Whitelaw stated that the reason for transmitting remittances is pure altruism, and they are highly correlated with the family ties that a migrant has back to his/her origin country.

Remittance flows could be in form of cash or goods to support the families. An individual could remit via formal channels, such as banks or credit unions, or informal channels. However, when remittances being applied with the latter, it contributes as a big challenge for analyzing remittances since the availability of that data is hard to be found.

Many papers address the factors that influence remittances. At the microeconomic level, the most promising indicators are the income level of the household, education, and health conditions of the family [5] [68] [41]. On the other hand, there are macroeconomic variables that influence these flows. Economic conditions, fluctuations, political instabilities, exchange rates, inflation rates, or GDP values of both origin and host country could be some indicators [1] [61] [70].

Many other micro and macro indicators could demonstrate how the individuals remit, yet this research area is not related to our analysis. Since for the thesis daily top-ups data will be used, that makes it rather difficult to include these indicators in this research. Microeconomic and macroeconomic values are not easy to be found, and it is extremely unlikely to find daily information for most of them.

2.2.1 International remittance flows during 2020

Throughout 2020, a variety of national constraints and lockdowns took place due to Covid-19, which affected the financial sector globally [54]. During that time, researches showed that restricting the individuals has several immediate consequences on the economy [40][29]. As long as there will be such restraints, economic improvements will persist to deteriorate. Moreover, according to estimations [29][56], the world's GDP growth decreased in 2020 due to the pandemic.

In 2020, World Bank estimated the consequences of the pandemic on these flows. According to their report on October 2020 [11], remittances that were going to be sent to low-income countries (LICs) and middle-income countries (MICs), would be dropped dramatically by 7.2% for the year 2020. It is mentioned that the increased unemployment rates for migrants to the host countries, the exchange rates against the US dollar, the oil prices to home countries, are some crucial features that would contribute to these drops inflows. Nevertheless, World Bank published the latest report on May 2021, where the actual impact of Covid-19 on remittances is analyzed.

In May's report [12] it is stated that the remittances slightly dwindled by 1.6% since 2019, which indicates that from 548$ billion transmitted in 2019, they decreased to 540$ billion in 2019. Even though the predictions revealed that remittances would have a notable fall due to the unemployment of migrants and the restrictions of each country, the latest report showed that individuals supported their origin countries during the pandemic. The remittances received by many regions appeared to have grown compared to 2019; the Middle East and North Africa (2.3%), South Asia (5.2%), East Asia and the Pacific (7.9%), and Sub-Saharan Africa (12.5%). Yet, for Central Asia, there was a shrinkage of 9.7% for 2020 as opposed to 2019. More precisely, the remittances started increasing after June 2020, where many countries loosened the governmental restrictions of Covid-19.

In the World's Bank report for 2020, it is stated that migrants located in the United States, United Arab Emirates, Saudi Arabia, and Russia seemed to have the most crucial influence on remittance flows, as they are the top sending countries during the pandemic. As for the top receiving countries in US dollar terms, these remained the same as the previous years; India (83$ billion), China (60$ billion), Mexico (43$ billion), the Philippines (35$ billion), Egypt (30$ billion), Pakistan (26$ billion), Bangladesh (22$ billion), Nigeria (17$ billion), Vietnam (17$ billion), and Ukraine (15$ billion). Even if the pandemic persisted after 2020, the World Bank predicted that the remittances will rise by 2.2% for 2021. However, this is highly influenced by the Covid regulations of each country, and the results could be different in the end.

2.2.2 Digital remittances

Since the remittances have a crucial influence on the world's economy and notably the economies of developing countries [44], there was an enhanced interest by the companies to adopt an alternative system for mobile transactions in the past years. It was essential to ascertain something distinctive from traditional services and something that would be serviceable and uncomplicated for the customers. The new way of transferring money that was established was the airtime top-up transfer. These are money transactions that an individual could purchase and represent a Business-to-Business (B2B) model. There should be connections across the telecommunications of the countries to be applied, yet these are not available to all regions. The airtime top-ups are mostly used for SMS, calls, or mobile data and consider to be fast, low cost, secure and simple. Implementing that kind of mobile service to individuals' routines is presumably one step further to solve several daily difficulties that were occurring (waiting in the queue at the banks to transmit money, etc.).

According to Ngugi et al. [53], a study conducted in Kenya, they stated that adopting this new mobile technology, will not have an economic consequence on traditional services like banks. In [21] one of the conclusions was that there should be an association between banks and services providing airtime top-up transactions to achieve, among other things, economic growth. Having accomplished that, individuals (particularly the poor ones), would prefer to use that kind of service if they are served at low-cost prices. Not only are there avails from the customer's perspective, but the providers will gain more perks since they will be capable to perform more enhanced risk evaluations and improve their customer services [35].

Merritt [42] stated that in the last years, mobile transfers have become more advanced, implying that they switch from traditional to wireless providers. Digital services are more affordable for the customers since these providers yield more negligible cost services and not only. Besides, he mentioned that wireless transfers come to be the future of transactions and that back in 2010 where this study was conducted, mobile transactions were an ongoing challenge. Additionally, for the developed countries where mobile infrastructures are more reliable than the developing countries, Merritt states that fresh and stabler services of airtime top-up transfers could be established and be discovered. As a result, they could be adopted by other emerging nations.

Erickson [48] stated that given the fact that mobile transfers are related to the reduction of the poverty of one developing country, governments should invest in mobile infrastructures to improve and expand these services. In a same study, Adams et al. [3] discovered that international receiving remittances for 71 developing countries, could reduce the poverty of one nation. Precisely, they state that on average a 10% increase of these flows could reduce by nearly 3.5% the number of people living in poor conditions.

The airtime top-ups have a crucial impact on the economy of each country. Mawejje and Lakuma [39], examined the influences of mobile transfers as a macroeconomic effect in Uganda. They revealed that mobile money is positively correlated with a bunch of other macroeconomic variables, e.g. private sector credits, etc. One of the outcomes implied that mobile transactions are deeply associated with the economic development of this country. Daily data for macroeconomic variables such as exchange rates are not easily obtained, and therefore cannot be used for this thesis.

According to World Bank, many migrants chose to transmit money to their home countries via digital services. These were increased from almost 8$ billion (2019) to 12$ billion in 2020. The official remittance flows records, are not substantial portions of the actual flows being applied from migrants. There are unofficial channels to transfer money to the origin countries, for instance, hand carry. World Bank stated that during the pandemic, many unrecorded transactions, transmitted via these techniques (digital ways of sending money to home countries). That was the result of the restrictions and lockdowns to many countries since there was no alternative way for migrants to send money to their families.

2.2.3 The impact of natural disasters on remittances

Since this paper aims to analyze data of airtime top-up transfers, there are some researches where the correlation of natural disasters (earthquakes, floods, hurricanes, tornadoes, etc.) with the remittance flows is examined.

Bettin and Zazzaro [13] utilized data for 98 countries from 1990 till 2010. They revealed that when a natural disaster occurs, the affected regions gain more remittances. Mohapatra et al. [45] concluded similar outcomes when applying an analysis for Burkina Faso, Ghana, Bangladesh, and Ethiopia. They proved that earthquakes, droughts, and floods had a positive impact on receiving remittances.

In another study, Attzs [9] indicated equivalent results concentrating on the Caribbean and Central America. In this paper, they examined the effect of remittances in these areas. The findings showed that these flows rise after the strike of hurricanes and storms. Yang [71] using data from 1970 to 2002, observed that the hurricanes have an influence on the receiving remittances mostly on the developing countries.

Blumenstock et al. [14] applied an experiment following an earthquake shock on Lake Kivu during 2008. They utilize a four-year dataset with mobile phone activities and the airtime transactions that occurred through this time. The findings suggest that there was an uptick in the airtime top-up transfers when the individuals commenced practicing that kind of mobile service. One more finding is that people transmit money to the affected people more frequently when an earthquake befalls. However, the bulk of receivers were prosperous people than poor ones. This finding aligns with [37].

Since many studies conducted and analyzed the effects of natural disasters on remittances, in this particular thesis, we examine how these disasters are correlated with digital remittances.

2.3 Time series forecasting

Time-series data is the category where the features are gathered sequentially through time [18] and the data points could be daily, weekly, monthly, etc. [16]. Moreover, when we present forecasting techniques, we refer to models that perform predictions for the future. The period of forecasting depends on the information that is used and the horizon that a company, researcher, etc. that favors having some insights about the future trends of their data.

There are two different forecasting categories, the qualitative and the quantitative techniques. The former is when no prior data could be used to construct the model and the information cannot be measured. The latter is when one base the results on historical data (used mostly for short-term period forecasting purposes) [47].

Many statistical models are widely used for time-series analysis. The autoregressive moving average model (ARMA) is one example [31]. Other forms of ARMA can be used, which the best model depends on the nature of the data. One of them is the Autoregressive Integrated Moving Average (ARIMA) which is widely used with time-series data [73]. An extended form is the ARIMAX which handles independent variables along with the dependent one.

Various studies conducted using these models, and more precisely for the forecast of macroeconomic variables. Durka and Pastorekova [57], compared ARIMA and ARIMAX models by utilizing quarterly data of GDP per capita and unemployment rates from 2000 till 2011. They discovered that the validation methods Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE) of the ARIMA, are slightly more trustworthy than the ARIMAX's model. They indicated that even if we manage a more complex model, the results will not always be more accurate. Since the ARIMAX model is highly recommended when one uses time-series data, this approach will be used for the thesis. Given the fact that there are limited airtime top-ups data, we assume that an ARIMAX model will be the best option for this analysis.

Apart from the classical approaches of time series analysis, studies develop complex machine learning (ML) models for that kind of research. In [7], they explicated that the Artificial Neural Network (ANN) models perform better for time-series data as opposed to the common ones mentioned earlier.

Many studies could be discussed, contributing to the literature, where they used ANNs models utilizing time-series data. For instance, Nakamura [49] explored the potentiality of managing Neural Network (NN) models for forecasting economic variables such as inflations. By this research, NNs models present a valuable contribution when being used for macroeconomic studies, comparable, for instance, to auto-regressive models.

Other studies developed a Support Vector Machine (SVM) for time-series analysis. In [64] they used daily stock markets data from January 2000 till October 2012 for predicting the next day. Apart from the stock markets data, Shen et al. utilized other features with daily information, that account as interventions of the stock values. However, for this thesis, a NN or SVM could not be used since there were limited data. These papers are mentioned to present that there are different types of time series methods that could be developed to accomplish a good prediction model.

One widely technique that is used for time-series analysis is the long short-term memory (LSTM) approach [32] [62]. There are studies where that kind of neural network was compared to other simple prediction models, such as ARIMA [66] where it outperformed the latter. Even if the LSTM is supposed to be a good model for time-series analysis, it is not presented in this thesis because it failed. Since a neural network as the LSTM requires many data as input points, it was not feasible to demonstrate this approach for the prediction of airtime top-ups. Nevertheless, many LSTM models with different parameters were created for this purpose, but the length of the data was not enough to develop a good and accurate model.

2.4 Forecasting remittances

There is limited literature for forecasting remittances, particularly for daily predictions. The availability of these flows is hard to be found as daily data. Therefore, the research on this field has potentialities to be examined in greater depth.

One of the first studies forecasting remittances was performed by Mohapatra and Ratha [46]. They handled remittance data for 200 countries throughout the global financial crisis that appeared in 2008. They intended to predict what would occur in a country-level approach in the next years. They developed two approaches, the remittance matrix-based approach, and the elasticity-based approach. The assumption of the former is that the remittances will depend on migrant incomes of the origin country, and the latter assumes that these flows will increase faster than the incomes of individuals in the host country. The flows from a sending country using the remittance matrix-based approach, depending on the outflows ratio of the sending country to receiving country, as well as on the GDP predictions for the next time step (these are calculated by the World Bank and they are available online). For the second approach, the authors calculated the sending remittances concerning the elasticity of remittances, the migrant incomes, and the predictions of the GDP. For both approaches, they added the share remittances of the home country in the outflows of the host country. The results revealed that the income remittances of the developing countries would decrease after the crisis. However, Mohapatra and Ratha stated that these models have multiple limitations. These refer to other factors such as exchange rates, return migration, or immigration controls that they did not use for the models.

Adedokun [4] developed an ARIMA model to predict the received remittances to Nigeria till 2019. In this paper, the data resembles the percentage of remittances to GDP from 1977 to 2009. This corresponds to a simple model since it utilizes only the variable that they want to predict which corresponds to the remittances. Nevertheless, the findings suggested that these flows as a percentage of GDP in Nigeria, would be rising for the next 10 years (2010-2019).

Garcia-Alonso et al. [24] developed a model where they predict the remittance flows while using other datasets that the influence to the former is known. The model is based on the Monte-Carlo approach and prognosticates the remittances, where three principal variables are taken into account; the average salaries that a migrant receives, the migrant stocks of the host country gained by one country, and lastly one variable that includes many sub-variables like the skills of the migrant, or educational background, etc. The data used for the analysis corresponds to the USA and El Salvador, which represent the host and home countries respectively. One of their findings is that the model can be used when natural disasters or economic shocks occur, to predict the financial consequences of the receiving remittances.

The previous studies utilized remittances with a bunch of other variables that influence them. For the thesis, these factors (GDP values or exchange rates which were presented earlier) cannot be used since daily data is in place for the analysis.

2.5 Other approaches analyzing remittances

The following research study [69] focuses on the international remittances in Nigeria and the effect on the growth of the economy and the development of the country. The data are remittances from 1970 - 2008, along with other datasets (e.g. the GDP growth of the country). Utilizing the Autoregressive Distributed Lag (ARDL) bounds testing approach, the findings implied that educational reforms necessitate being established, as well as, improvements to technologies and other factors, to boost the economy of Nigeria. Moreover, the skills, knowledge, and experiences of an individual could enhance the receiving remittances of this country. This finding aligns with [2], where they state that poorly educated migrants are more likely to practice transfers back to their home countries than highly educated individuals.

Solomon managed remittances data for Ethiopia from 1980-2017 [67]. The outcomes from the bound testing approach revealed that the remittances have a long-run association with the economic growth of the country. In another study [6], the ARDL approach was tested again. Moreover, the relationship between the remittances and the misery index (a single statistic that sums the unemployment rate with the inflation rate) was studied. It embodies one useful topic to comprehend how remittances affect the economy of one country. GDP per capita growth, inflation and unemployment rates, inflow remittances in Turkey from 1975 to 2011 were managed for this analysis. The conclusions showed that the remittances and the misery index have a long-turn relationship. Another finding was that if there is an uptick in unemployment and inflation rates, there will be an increase of the inflow remittances.

Serino and Kim [63], analyzed panel data for 66 developing countries by developing Ordinary Least Squared (OLS) models. The remittances corresponded to data from 1980 till 2005. They examined the influence of these flows on the poverty of these countries. More variables were used on the regression model, which are the GDP of the home countries, income inequalities, and foreign direct investments. The findings suggested that remittances reduce the poverty of a country, especially the low-income ones. This form of the regression model will be used in this thesis. The aim is to analyze the daily data of airtime top-ups and understand the relationship between natural disasters and public holidays with them.

3 Data

3.1 Airtime top-up data

For this thesis, the data corresponds to bilateral airtime top-up transactions data. DT One is an innovative company that provides digital connectivity solutions which shared that kind of data with us. It is a Business-to-Business (B2B) network, and its purpose is to design a network where users have the opportunity to transfer mobile top-ups around the world. The company has 1.000 global partners approximately in more than 160 countries. DT One's customers are mostly located in countries where there are emerging economies. The company's target is to yield solutions and improvements to universal digital communications.

DT One shared two different datasets for previous analysis and the thesis. Both of them are explained in the next subsections.

3.1.1 First dataset: Top 14 sending countries

The first data that was shared, corresponds to daily mobile transactions that occurred from January 2019 till July 2020. It consists of the top fourteen sending countries for this timespan, presented in table 1, and most of them are associated with twenty receiving countries. There are almost 300 corridors in the dataset, excluding the flows within the same country. A corridor represents the country pair, meaning the set of two countries (e.g., USA-HKG). The former resembles the sending country, and the latter the receiving country. Moreover, the shared information existing in the dataset is aggregated for each corridor, and there are no sensitive data about any of the DT One customers.

There are some corridors where the top-ups are not transferring each day but merely particular days of the week. One reason that could explain this situation, is the substantial low transaction volumes that happened from individuals to these countries. Hence, for some receiving countries, the data points for some days do not exist in this dataset.

3.1.2 Second dataset: Top 15 receiving countries

Apart from the aforementioned dataset, the company shared data that corresponds to the top fifteen receiving countries for the year 2020. There are

First Dataset: Top 14 Sending Countries				
United States	Indonesia	Indonesia	France	Thailand
United Arab Emirates	Philippines	Germany	Sweden	South Africa
United Kingdom	Saudi Arabia		Spain	Hong Kong
Second Dataset: Top 15 Receiving Countries				
Philippines	Myanmar	Mexico	Mali	India
Indonesia	Malaysia	Cuba	Ethiopia	Nigeria
Argentina	Bangladesh	Nepal	Senegal	Pakistan

Table 1: The sending and receiving countries shared from DT One.

254 corridors in this dataset, excluding the domestic transactions within one country. The top fifteen operating countries according to the data of DT One, presented in table 1.

Combining these datasets would produce bias results since the data are different. On the one hand, there are the top 14 sending countries for 2019 till July 2020, and on the other hand only data for 2020 for the top 15 receiving countries. If we were to supplement the top sending countries to this second dataset, the results would not be reliable.

One more reason to work with the second dataset is that we study the effects of natural disasters and holidays, on the receiving countries. Hence, if we were to investigate the first dataset, the outcomes would not be accurate since the information resembles the top sending countries of DT One data.

3.1.3 Preprocessing steps and limitations

Before proceeding to the analysis of the top-ups, some preprocessing steps were implemented to clean the data. One feature that was not taken into account for the models, were the corridors that had the same sending and receiving country (the domestic flows in a country). The primary goal is the analysis of the impact of natural disasters on the receiving top-ups to host countries, as well as how much other countries have applied transactions throughout this time. However, in Section 5, there is a comparison of the transactions received from one country, with its domestic flows for 2020.

One drawback that we had to deal, is that there were repeated values for one day for a small portion of the countries in the dataset. The United States of America appeared with three different sending names, which are; the USA, the United States, and the United States of America. There were days that those variables, had different values for the same day. For instance, on 15^{th} of January 2020, the number of top-ups that were sent to Cuba was 25, 26.862, 2.798 respectively. There was no information if this was a technical issue in the company's servers, therefore we kept the values only for the sending country that appeared as "United States". Given that the United States was among the top sending country to many of the fifteen receiving countries, the idea was to keep only the highest amount of daily transactions from these three variables that discussed above. Furthermore, the sum of these values could be another solution, yet the results would not have an immense variation from the "United States" variable that we chose to work with. The duplicated values problem appeared with the United Arab Emirates, which had two values as sending country. The same process as the United States was followed.

The company mentioned that Singapore should be completely excluded as a sending country because the values were not correct. Furthermore, multiple transactions had an instance called "Global" as a host country. Since there was no information about which sending country it was representing, there was no value including it in the final dataset and it was erased from the analysis.

The data obtained from the company should not be seen as a representative sample of the whole picture of remittances. Not all migrants use that kind of digital service and the volume from one country to another changes accordingly. The selected corridors existing in the shared data were selected by the company. We are not aware if the data reflect the exact size of transactions, therefore should not be accounted for like that. For instance, some countries could be excluded from the dataset due to the sensitivity of the information or the selection that happened for business purposes.

3.1.4 Normalized data

For the investigation of the airtime top-ups, the data was normalized applying the z-normalization. Practicing this method, we assure that the output data has a zero mean and a standard deviation close to 1. Moreover, with z-normalization, we ensure to not lose any data valuable for the analysis (e.g. peaks for some days). Normalizing the data was essential to compare

Features of the CRED data		
Total Deaths	Total Affected	Total Injured
Total Homeless	Total Damages	Reconstruction Costs
Disaster Group	Disaster Subgroup	Disaster Type
Associated Dis 1	Associated Dis 2	Declaration
Event Name	Country, Region, Continent, Location	Disaster Subtype
Distance Mag Value - Scale	Geo Locations	Latitude Longitude

Table 2: The complete features of the CRED natural disasters dataset.

several time series data that had different ranges.

3.2 Natural disasters data

The second dataset concerns the natural disasters for 2020. The EM-DAT dataset was formulated by the Centre for Research on the Epidemiology of Disasters (CRED), and it is termed as Emergency Events Database (EM-DAT)[3] [28]. It is one of the few databases that consist of geocoded data for natural disasters, as well as technological disasters (transport, industrial, and miscellaneous accidents). The whole information one could find starts from 1900 till the present. The EM-DAT is one of the most promising, useful, and prominent databases that could be employed for research purposes. Moreover, the technological category is not applicable for the thesis and was not selected.

EM-DAT provides a variety of information such as how many people were affected by the damages, the number of homeless individuals, total deaths, etc. Table 2 displays all the variables for the natural disasters of each country that exist in EM-DAT. According to CRED, the sources of the EM-DAT dataset, are UN agencies, non-governmental organizations, research institutes, etc. The criteria of a disaster to be included in the dataset, are the following;

- The total deaths of a disaster should be higher than 10.

[3] `https://www.emdat.be/database`

Natural Disasters of the CRED dataset	
Meteorological	
Extreme Temperature	Storm
Climatological	
Wildfire	Drought
Hydrological	
Landslide	Flood
Geophysical	
Volcanic Activity	Earthquake
Biological	
Insect Infestation	Epidemic

Table 3: The natural disasters that exist in CRED dataset.

- There are more than 100 affected, injured, or homeless people.
- The country that was stroked by one disaster declared a state of emergency or asked for international help.

Some of the disasters that one could observe in the EM-DAT dataset, are earthquakes, volcanoes, storms, floods, etc. There are five groups where the disasters are categorized into. An extended list of all the natural disasters that happened in the countries that we have mobile transactions data from DT One, can be found in table 3.

The EM-DAT dataset contains multiple variables that illustrate the severity of a natural catastrophe. However, there are many missing values for various features, such as the total deaths, injures, etc. The records for these variables are limited to only a small portion of the disasters in the dataset. Moreover, for some of the natural disasters, there were missing values for the dates. These represent the data that did not have a starting or ending month and were excluded. Another shortcoming of the dataset is that there were missing values for the starting or ending days, which were replaced by the mean value of the whole dataset.

3.3 Public holidays data

For the thesis, the public holidays for 2020 of the operating countries will be used. There is not a comprehensive dataset with all countries. Most

Dataset	Chosen features for the analysis	
DT One	ISO3 codes for sending countries	ISO3 codes for operating countries
	Corridor codes	The daily sending top-ups from sending to receiving country
Office Holidays	Public holidays for each country, including religion ones	
CRED	ISO3 code of the country that a natural disaster happened	The daily natural disasters happened to each country, if any (flood, etc.)
	The daily amount of the natural disasters happened to each country, if any (1,2, etc.)	Binary Variable; 1 - if a disaster happened, 0 - if not

Table 4: The chosen features for the analysis and the sources.

importantly, in the existing datasets, there were missing values for the corridors that we have airtime top-ups data.

One of the aforementioned incomplete data sources is the Azure dataset [4] which includes universal information for public holidays from 1970-2099 for 38 countries. One can obtain regional holidays in the dataset, but not for all of them. Moreover, the Azure dataset consists of data obtained by the PyPI holidays library [5], and Wikipedia. Nevertheless, even if the Azure dataset contains countries from the PyPI holidays library, we noticed countries that exist in the latter and not in the former. Possibly this happened due to upgrades where more countries were added to this python library. Yet, both datasets include information for a small portion of the home countries and there are missing data for countries that DT One distributed with us.

Since the Azure and PyPI holiday datasets are not suitable for the analysis, we decided to develop a script and generate a new dataset with all the countries with their public, religious, and national holidays. The "Office Holidays"[6] provides access to data where one can locate all the informa-

[4]`https://docs.microsoft.com/en-us/azure/open-datasets/dataset-public-holidays`

[5]`https://pypi.org/project/holidays/`

[6]`https://www.officeholidays.com/countries`

tion for the holidays of each country. Mostly for all the countries, this site delivers information for each country from 2015-2022. Since the company granted us airtime top-up flows data for the years 2020, then this site was the ideal source to retrieve the holidays' information. For this purpose, we developed a script to download the data for all the countries presented on this site. After this step, the public holidays for each receiving country were combined with the dataset of the mobile top-up transfers. Therefore, working with the "Office Holidays" led to a complete dataset with no missing values for the countries that we had transactions data.

Overall, in table 4 one could find the data sources and the variables that were selected from each source. Some of the features were not in the sources but were created manually and they are based on some variables of each dataset. They represent mostly numeric values that are more flexible to be employed in the research. For instance; there is a binary variable if a natural hazard occurred one particular day, and it is saved as 1 if one happened, 0 if not.

4 Methodology

Analyzing the remittance flows, one should describe the problem mathematically and give a formalization of it. For this paper, the formula for the digital remittances that is going to be used is as follows;

$$R_{ijt} = f(A_{ijt}, y_{jt}, x_{jt}) \tag{1}$$

where R_{ijt} represents the remittances from the sending country (i) to receiving country (j) on day t, A_{ijt} are the transactions from country i to country j on day t, x_{jt} is the set of the natural disasters at the j country that happened a specific day t, and finally, the y_{jt} corresponds to the public, national, or religious holiday of the j country on day t.

As mentioned in Section 3, the data about natural disasters could be expressed as follows;

1. $x_{jt} = 1$: if a natural disaster happened in the country j on day t.
2. $x_{jt} = 0$: if not.

and respectively for the holidays, $y_{jt} = 1$ if on the day t it is holiday in the receiving country, and $y_{jt} = 0$ if not.

The values of natural disasters and public holidays are considered as independent variables for the following models. These variables could also be called exogenous variables.

4.1 OLS Regression model

Ordinary Least Squares (OLS) is one of the most popular linear regression models in statistics [58]. It is widely used to estimate the coefficients of the independent factors that influence the dependent variable. The basic form of an OLS regression model can be formulated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \epsilon \tag{2}$$

where Y is the dependent variable, β_0 is the unknown coefficient of this variable, β_1 and β_2 are the unknown coefficients of each independent variable, and ϵ is the random error of the model.

The primary goal of an OLS regression model is to determine the effects that the independent variables cause on the dependent variable. For instance, if one independent variable increases by one unit with the rest of the variables being constant, one can determine the effect's magnitude of this independent variable to the Y [20].

The goodness of fit can be found from the R-squared value. The R^2 is a statistical measure representing the proportion of the variance of a dependent variable that is explained by the independent variables. The range of this metric is $(0, 1)$ and the more it increases, the better the model fits the data. A high R^2 indicates that the target variable can be explained by the shifts of the independent factors which were used on the regression model. More precisely, the relationship between Y and X's variables can be derived by the regression model.

Adjusted R-squared is a moderated version of R-squared, which estimates the proportion of the variance described by the independent variables that explain the target variable Y. The adjusted R-squared could have a lower value than R-squared if an independent variable is not needed for the model. Hence, the difference between R-squared and adjusted R-squared is the degree of freedom that the latter provides.

The difference between these two metrics is that as long as an independent variable is added to the model, the R-squared increases, whereas the adjusted R-squared rises only if the independent variable influences the target variable Y.

4.2 ARIMAX and SARIMAX models

One of the most promising models for time-series forecasting processes is the Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box and Jenkins [16]. It employs past data to predict the future cycles of the determinant that an individual is interested in. ARIMA models not only have been applied for predicting macroeconomic determinants but for remittances also, as shown before in the literature [4].

One model could utilize more determinants that count as an intervention to the predictions. Hence, the updated model of ARIMA, which includes these changes, is denominated as the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model. Mostly in economics, these external interventions (or independent variables) of the target variable Y, are called exogenous variables. The ARIMAX model could be seen as a regression model, and it is comprised of three terms [52]: Autoregressive model (AR), Order of Integration (I), Moving Average (MA), and the exogenous variables (X).

Autoregressive model or AR(p): It represents a regression model where the current time series data depends on previous steps with one or more lags. The p presents the most significant lag of the time series, and it can be determined manually by the Partial Autocorrelation Function (PACF). The AR of order p can be written as follows:

$$y_t = c + \vartheta_1 y_{t_1} + ... + \vartheta_p y_{t-p} + \varepsilon_t \tag{3}$$

where ε represents the noise and the ϑ corresponds to the coefficients of the previous steps. Lastly, the AR is a multiple regression model with the previous lag(s) of the current value.

Order of integration or I(d): If the data is not stationary, then this variable is used to make it stationary. That is, if the data is replaced with the difference between the current points and the previous values, then d should be used in the model. The most common value of d is 1 or 2. For instance, if we use first differences, then the formula of the integration would be;

$$y'_t = y_t - y_{t-1} \tag{4}$$

Moving Average model or MA(q): This term calculates the moving average error for a number of periods. It can be interpreted as the prior errors (or noises) that affect the current value. Therefore, the term q is used to reduce the noise out of the model and smooth it out.The MA of order q can be written as follows:

$$y_t = c + \varepsilon_t + \varphi_1 \varepsilon_{t_1} + ... + \varphi_q \varepsilon_{t-q} \tag{5}$$

where the φ are the coefficients of the errors of the previous steps. One can find the q value by looking at the Autocorrelation Function (ACF) plot.

In many cases, the data present trends, or seasonality. In the last case, the ARIMAX model is called Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), where more parameters should be imported. This model consists of the ARIMAX model order (p,d,q) and the seasonal order (P, D, Q, s). The latter order corresponds to the seasonal Autoregressive model (P), seasonal order of integration (I), and Q which is the seasonal moving average term.

Seasonality(s): If the data shows seasonality, then the latter is represented as s. Seasonality indicates that some fluctuations or patterns repeat over a period (yearly, weekly, etc.).

The formulas for the SARIMAX components (P,D,Q) are the same as the equations 3, 4, and 5 respectively. The only change is that for the seasonal order we use a time-step of s.

The equation that handles the exogenous variables in both ARIMAX and SARIMAX models is as follows [8]:

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_n x_{nt} + \varepsilon_t \tag{6}$$

where x is the set of the n independent variables at time t, and the β represents the regression coefficients of the x.

The final formula for SARIMAX model is [22]:

$$\begin{aligned}(1 - \varphi_1 G - \varphi_2 G^2 - ... - \varphi_p G^p) \times (1 - \Phi_1 G^s - \Phi_2 G^{2s} ... - \Phi_P G^{Ps}) \\ \times (1 - G)^d \times (1 - G^s)^D \times (y_t - \beta_0 - \beta_1 x_{1t} - ... - \beta_n x_{nt}) \\ = (1 + \vartheta_1 G + \vartheta_2 G^2 + ... + \vartheta_q G^q) \\ \times (1 + \Theta_1 G^s + \Theta_2 G^{2s} + ... + \Theta_Q G^{Qs}) \times \varepsilon_t\end{aligned} \tag{7}$$

where the first two parts correspond to AR and seasonal AR respectively. Then, the integration orders are presented along with the formula for the independent variables.

The parameters of each model can be estimated using maximum likelihood functions [22] [25]. Two measures should be checked before proceeding with the best model, and understand which one fits the data the best. Akaike Information Criterion (AIC) is a measure that can be used for that purpose. Among a collection of models, the AIC estimates the quality of each model. It penalizes the ones that are too complex and rewards the others which have better goodness of fit. The formula is as follows:

$$AIC = 2K - 2\ln(L) \tag{8}$$

where K describes the number of independent parameters, and L is the maximized likelihood estimate.

The second option to evaluate a model is the Bayesian Information Criterion (BIC). It is related to AIC but in the case of BIC, the penalty term is higher. For ARIMAX and SARIMAX models, the AIC is used mostly to select the best parameters of the model. Nevertheless, the formula for BIC can be written as:

$$BIC = K \ln(n) - 2\ln(L) \tag{9}$$

where n denotes the observations, and the rest of the parameters are the same as equation 8.

There are some pre-processing steps before proceeding with the development of the (S)ARIMAX models. The first step is to split the data into training and test data (80% and 20% respectively) and then normalize the data. This process was mentioned earlier in Section 3.

The second step is to recognize if the data is stationary or not. When the data is stationary, it means that there is no trend, there are no fluctuations over time (in most cases it means that there is no seasonality), and there is a constant variance over time [50]. One can use the Augmented Dickey-Fuller (ADF) test to determine if there is stationarity. If the test rejects the null hypothesis (that is, the process has a unit root), then the data is stationary.

The next steps are to identify the parameters (p,d,q) from the ACF and PACF plots. If there is seasonality in the data, then one should check and the rest of the parameters (P, D, Q, s). However, determining these values is not such a precise method. "Auto Arima" [7] is a library provided in both Python and R programming languages that help with this process. It performs many iterations to find the best parameters that minimize the AIC and BIC measures. There are many parameters that one can choose to get the best results from this library. For instance, a component of the library finds the optimal parameters by not fitting all the possible hyper-parameter combinations (step-wise). According to the documentation of this library, implementing this approach in the model will slightly prevent it from overfitting. This feature was used for all the models in the case studies.

After fitting the model, one should investigate the diagnostic tests, such as QQ plot and histogram. Both plots verify if the data are normally distributed. In the QQ plot, the x-axis is the normal distribution with a zero mean and standard deviation which equals one. Nevertheless, if the parameters that were chosen pass the tests, then this model is evaluated in the test data.

[7] `https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html`

The aim of developing the ARIMAX and SARIMAX models is to predict the receiving airtime top-ups, by importing independent variables such as natural disasters and public holidays. Moreover, all the models that were developed in the thesis, are depending on the previous-day value. Since the (S)ARIMAX models are used mostly for short-term predictions, then the prognostic horizon of the models corresponds to the next day.

4.3 Validation

For each (S)ARIMAX model, the data will be separated into training and test data. The former will be used to estimate the parameters, and the latter is necessary to evaluate the efficiency of each model. Following this process, one could gain some insights into the reliability of the models to forecast new unseen mobile transactions.

As validation methods for the models, there are three KPIs that will be used to comprehend the errors of each one of them. These are the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). All measures are negatively-oriented scores and scale-dependent error techniques, which means that they are scaled in the same unit as the variable that one wants to predict. The formulas are;

$$MAE = mean(|e_t|) \tag{10}$$

$$RMSE = \sqrt{mean(|e_t^2|)} \tag{11}$$

$$MAPE = mean(|100(\frac{e_t}{y_t})|), \tag{12}$$

where the e_t denotes the forecast error of a model and it is the difference between the actual observed value and the predicted one. It is defined as:

$$e_t = y_t - \hat{y}_{t|t-1} \tag{13}$$

All the models will have the same dataset's length and equal separation to training and test subsets. Since the data is limited, the separation of the training and test data are 80% and 20% respectively.

5 Results

For this research, three countries were selected as the main focus of the thesis. These are the Philippines, Indonesia, and Ethiopia.

According to World Bank, the Philippines is a middle-income country (MIC), Indonesia has the largest economy in Southeast Asia and is labeled higher than a MIC, and lastly, Ethiopia is a low-income country (LIC). Having three different income categories, it will be interesting to research the importance of receiving mobile top-ups on them. Moreover, since incomes from other nations are crucial for many LIC or MIC economies, prediction models for these countries could be a substantial opportunity to forecast the receiving top-ups. Apart from the models, there will be an explanatory analysis of Indonesia, Ethiopia, and the Philippines.

The analysis starts with all the home countries, which are the top fifteen countries existing in the dataset. From this investigation, the RQ1 will be answered, that is to find which natural hazards and holidays have an effect on the airtime top-up transfers. Moreover, the same research question will be answered separately for the three case studies.

5.1 Analysis of the origin countries

5.1.1 Regression model

In Section 4, the OLS regression model was explained. From the daily airtime top-ups data we have for fifteen receiving countries, the information was used and imported to this model. The formula is as follows:

$$\begin{aligned}\log T_{ijt} = \alpha_0 + \vartheta_t + H_t + O_t + \pi_{ij} \\ + F_t + W_t + L_t + EA_t + EP_t + S_t + ET_t \\ + V_t + D_t + I_t + HO_t + \varepsilon_{ijt}\end{aligned} \tag{14}$$

T_{ijt} is the dependent variable and refers to the daily airtime top-ups from host country i to home country j at time t. On the right part of the formula there are the independent variables, where ϑ_t corresponds to date dummies fixed effect, H_t, O_t the fixed effects of host country i and origin country j respectively, and the π_{ij} the fixed effects for each corridor (e.g. USACUB). The rest of the variables refers to dummies of the natural disasters, except the HO_t which represents the public, national, and religious

Dependent variable: daily airtime top-up transfers from host country (i) to home country (j)	
Independent variable	Coefficient
Flood	0.054*** (0.007)
Wildfire	0.000*** (0.000)
Landslide	-0.004** (0.038)
Earthquake	-0.082* (0.087)
Epidemic	0.023** (0.015)
Storm	0.087** (0.022)
Extreme temperature	0.000*** (0.000)
Volcanic Activity	-0.251** (0.037)
Drought	0.260** (0.025)
Insect Infestation	-0.142** (0.013)
Holidays in Home country	-0.023** (0.011)
Day dummies	Yes
Home Fixed Effect	Yes
Host Fixed Effect	Yes
Corridor Fixed Effect	Yes
Observations	82.890
R-squared, Adjusted R-squared	0.942

Table 5: The dependent variable corresponds to daily top-ups between corridors, and the independent variables represent the natural disasters and the public, national, religious holidays of the origin country. For each of the variables; the 1^{st} number corresponds to the coefficient and in parenthesis the std. error. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

holidays of the origin country.

When a regression model includes fixed effects for some of the independent variables, it means that the group means are fixed and non-random. These effects are used when there are factors that might influence the outcome of the analysis. By incorporating fixed effects we are controlling the differences across unobserved or observed predictors. For this analysis, they are useful since the regression model can understand most of the information from the dataset (see R^2 and Adj. R^2).

In table 5, the estimations can be obtained from the analysis. All the top fifteen receiving countries were employed, with their sending countries. The dependent variable refers to the daily top-ups and the independent variables are the natural disasters and a dummy variable if a specific day is a holiday or not. The regression model learned most of the information in the dataset, and this can be seen from the values of the R-squared and Adjusted R-squared, which are the same (0.942 for both of them).

It seems that the p-value for earthquakes is greater than 0.05, which is not a shred of strong evidence to conclude that a significant difference exists. Therefore, we cannot be precise if the earthquakes influence the airtime top-ups in general. On the other hand, floods, epidemics, storms, and droughts are the most severe natural disasters that when they occur, individuals transfer money back to their origin countries. Looking at the coefficients of each of these four values, droughts appear to have the highest impact on the top-ups.

Even though extreme temperatures and wildfires seem to be significant ($p - value < 0.01$), they do not have any effects on airtime top-ups ($coef. = 0.000$). Moreover, the last natural disasters that are statistically significant (landslides, volcanoes, insect infestations), have a negative correlation with the dependent variable. If one of them befalls with the rest of the independent variables constant, the receiving airtime top-ups of the origin countries are less likely to climb.

The last independent variable of the regression analysis refers to the public, national, and religious holidays of the origin country. The outcomes revealed that it is less likely to affect the airtime top-ups. However, as we will explain in the next sections, only some of the holidays influence the receiving countries. For the OLS regression analysis, all the holidays for the year 2020 were used. The model recognized that some days impact the top-ups, yet various holidays do not contribute to a rise in airtime transactions.

5.1.2 Impact of natural disasters and public holidays on top-ups

Before proceeding in the three case studies mentioned earlier, there will be an analysis of six other receiving countries.

Even though the regression model in the previous section showed that the public holidays do not influence the receiving airtime top-ups, this conclusion does not align with all these public, national, and religious holidays. As can be viewed in figure 1, there are particular days that international migrants transmit money back to their relatives. Eid al-Fitr and Eid ul-Adha are the most influential Muslim days. Bangladesh, Pakistan, and Nepal are the three countries that the receiving airtime top-ups increased on these specific days. Apart from the Eid holidays, other holidays affect these transactions. Some of the Christian ones, such as Easter and Christmas, do influence them. As for the former, this can be seen in the case of Nigeria. On Christmas, there is a slight peak of the receiving top-ups in Argentina and Nepal.

Due to the Covid-19 pandemic, several countries declared lockdowns in early 2020. For some of the countries existing in the dataset, the receiving airtime transactions dramatically increased after the announcement of these measures. Some examples worth mentioning are India, Nigeria, and Nepal. For these cases, international migrants residing abroad supported their homes to their origin countries.

Natural hazards occurred in all six countries presented in figure 1. For India and Bangladesh, disasters did not affect the receiving airtime top-ups. On the other hand, a drought with 35.000 affected individuals stroke in Argentina even before 2020. Along with the drought, a storm occurred. The receiving top-ups increased this day and the following. Nevertheless, from the figure, we can recognize that the top-ups that migrants transferred during the occurrence of natural catastrophes in Argentina are higher than the rest of the year. One could safely assume that the extended variance throughout this period could be caused by the disasters that happened.

During 2020, Pakistan had the same trend of receiving transactions. The only exception was after Pakistan day, where international migrants dramatically sent mobile top-ups to this country. It is difficult to interpret the intuition behind these actions during this period since three different factors could be the cause of this rise. Three natural disasters started weeks before this peak, yet one could assume that the first peak is explained by the public holiday on $23/03/20$. The second peak followed after the announce-

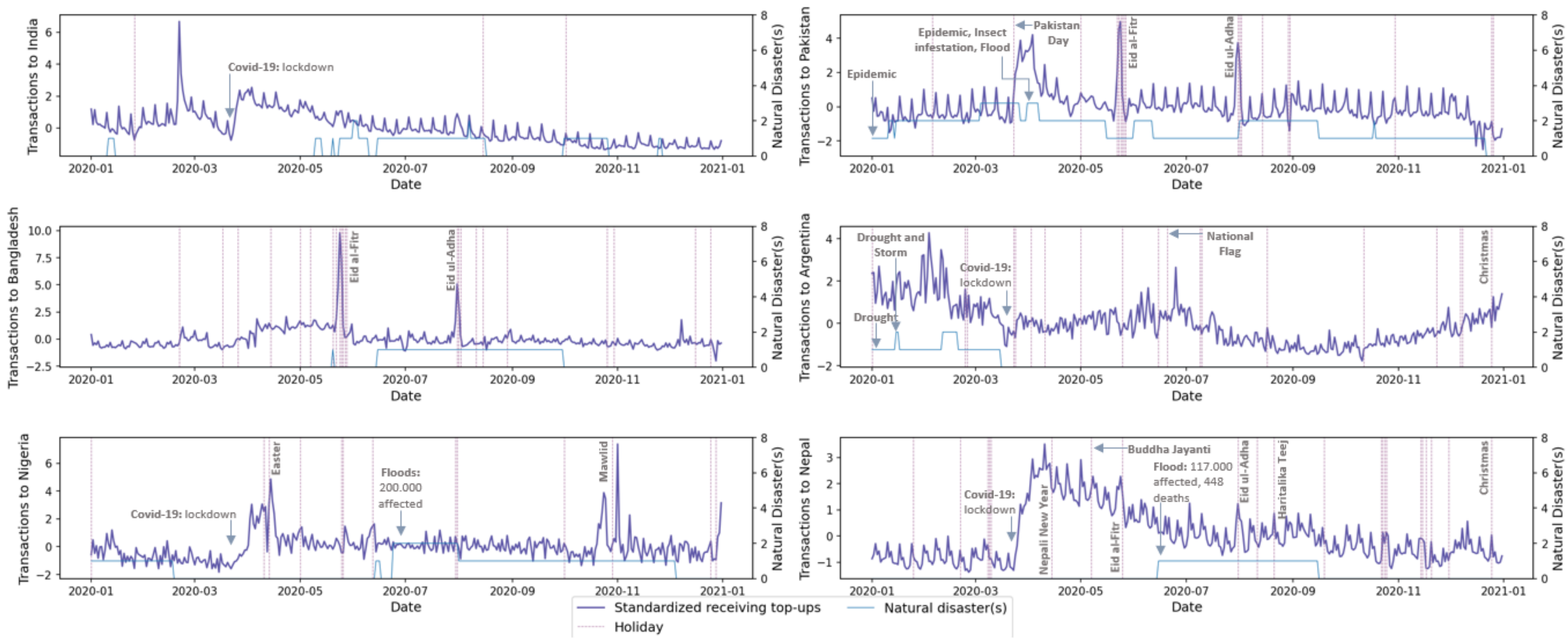

Figure 1: The receiving international airtime top-ups for six of the countries in the dataset.

ment of the lockdown due to Covid-19. At the same time, an epidemic and an insect infestation were taking place for months. However, a second flood occurred which led to a total of three hazards. Therefore, the intuition behind the second peak could be explained by the Covid-19 or the occurrence of these natural disasters.

5.2 Case study: Indonesia

The first case study of the thesis corresponds to Indonesia. To begin with, there is an analysis of the data where we find which are the top sending countries to Indonesia. After that, there is a correlation of the domestic and international top-ups, the effects of natural disasters on airtime top-ups, and lastly the prediction model.

5.2.1 International and domestic flows: The impact of natural disasters and public holidays on them

The second question will be answered in this section. The aim is to understand the behavior of international migrants when a natural disaster occurs back in their home country. What insights can we gain for the behavior of the migrants? Are the natural disasters founded in the previous section, affecting the receiving airtime top-ups? These questions can be answered in these subsections in each of the three cases studies.

Indonesia is one of the fifteen countries in the dataset that have domestic and international flows for 2020. In figure 2 there is a detailed plot with the most significant natural disasters that happened in Indonesia, as well as the labeled national-religious holidays that influence airtime top-ups.

Overall, the domestic remittances are greater than the international receiving top-ups for most of the half-year. However, it is obvious from the plot that there is no data for the domestic flows within Indonesia after the first days of July. Our partner was not aware of why this problem occurred, therefore we could not compare these flows with the international top-ups after this point.

It is understood from the figure that when the year started, a flood occurred with more than 75.000 individuals affected and 10 deaths. Although the domestic flows resemble a steady trend with ups and downs for 2020, it is safe to say that after this disaster, these inflows reached the highest peak of this year. However, the response from individuals living abroad was not equivalent to the support within the country.

After some time from the first natural disaster, a flood stroke in Indonesia with more than 100.000 people affected. In that case, both international and domestic receiving top-ups were higher than in previous days. Moreover, on 4^{th} of April 2020, fewer individuals were influenced by the previous disasters. It seems that international migrants sent top-ups directly

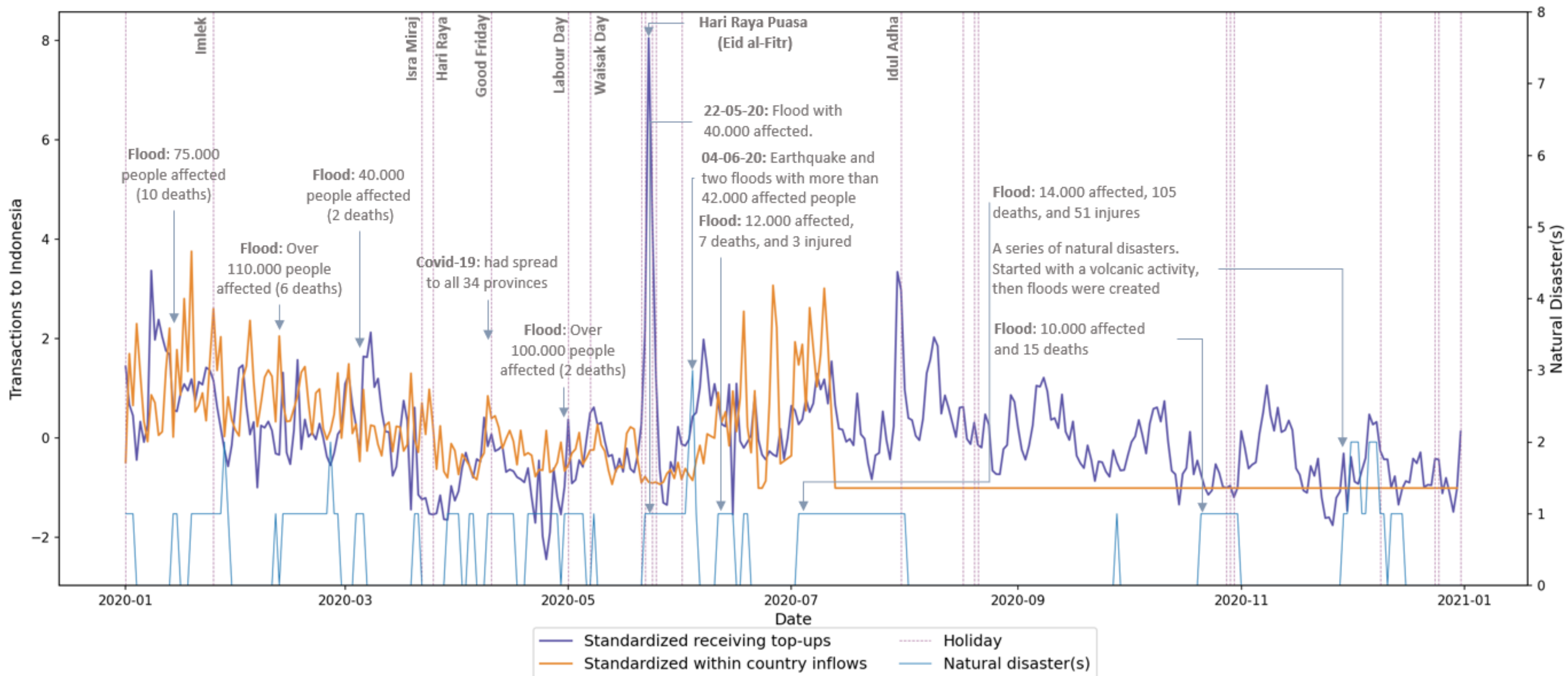

Figure 2: The international and domestic top-ups in Indonesia.

after the natural disaster. Even if this flood influenced less population than the previous disasters, it seems that the cities-provinces that were stroke by it, are important for Indonesia (e.g. the capital). One other reason could be that these affected cities have greater mobile infrastructures than others. Therefore, it is easier for international migrants to transmit top-ups to their homes. Given that the airtime top-ups data is not geotagged, we cannot say which cities or provinces were hit by the natural disasters.

Through the first week of May, there were two holidays in Indonesia and a flood with 100.000 people affected. One can perceive that the international receiving top-ups started to rise during this period, with the most important peaks on Labour and Waisak days.

The highest peak of the receiving international airtime top-ups appeared on Eid day. Given that more than 80% of Indonesia's population declared to be Muslims on 2010 [8], it is not surprising to observe these spikes each year that will follow.

After Eid al-Fitr, the international top-ups dropped for some days. However, they began to rise when a flood happened on the same day as Eid. On 4^{th} of June, three natural disasters occurred where the remittances appear to increase following them. More than 42.000 people were affected by the earthquake and the floods occurred during this time. However, not only the international top-ups started to grow, but the domestic flows as well. From the figure, it can be seen that during June these flows having fluctuations and peaks after a flood on $11/06/2020$.

The last peaks of the domestic flows happened at the beginning of a natural hazard on 2^{nd} of July. The effects from this flood were the worst during this time (the most deaths and many injures compared to the other disasters). Moreover, the second most important public holiday that influences the receiving international top-ups, is Idul Adha. Even though this peak is during the flood that took place on $02 - 07 - 2020$, we assume that the transactions happened because of this holiday.

Not all natural catastrophes enhance receiving remittances. For instance, in the last week of October 2020, 15 deaths followed a flood with 10.000 affected people. However, the international top-ups are steady and did not increase.

At the end of the year, a volcano caused problems to more than 15.000 individuals, yet the deaths or the homeless individuals are not available.

[8] https://worldpopulationreview.com/countries/indonesia-population

After the volcanic activity, two floods succeeded with 110.000 affected people. From figure 2, it is obvious that the receiving airtime top-ups increased steadily. The highest peak can be found in the occurrence of the last two natural disasters that happened simultaneously.

Given the analysis of the international and domestic top-ups, one comprehends the severity of the floods that occurred in Indonesia. In the previous section, we showed that when a flood occurs, most of the time the remittances grow. Comparing the analysis for Indonesia and this result from the regression model, it is safe to say that the floods indeed have a positive impact on the airtime top-ups.

5.2.2 Top sending countries to Indonesia

Figure 3 displays the top four sending countries of Indonesia for 2020. Since for each sending country the range was different, the data were standardized using z-normalization as mentioned in Section 3.

It can be seen that Malaysia holds the leading role in sending international top-ups to Indonesia. Hong Kong, Saudi Arabia, and the United States appear to follow a similar pattern. Moreover, they do not have a significant weight on the receiving top-ups for Indonesia.

Comparing the digital remittances and the official remittance flows could derive useful conclusions. According to World Bank's report, Malaysia and Saudi Arabia consist of more than 60% of the receiving remittances of Indonesia for 2020. This finding aligns with the results for the airtime top-ups since Malaysia has a vital role in receiving international flows in Indonesia.

5.2.3 Prediction model for Indonesia

The last research question will be answered in this section. For each of the case studies, a model was developed to predict the next day receiving top-ups.

In Section 4, the components and the methodologies of the ARIMAX and SARIMAX models were presented. For the case of Indonesia, we discovered that the data is stationary, and there is no seasonality (ARIMAX is suitable). However, since the data was only for one year, it is difficult to say if there is seasonality for all the years that the company has information for Indonesia. For instance, there could be a yearly seasonality, yet one cannot answer if that is the case with this restricted data.

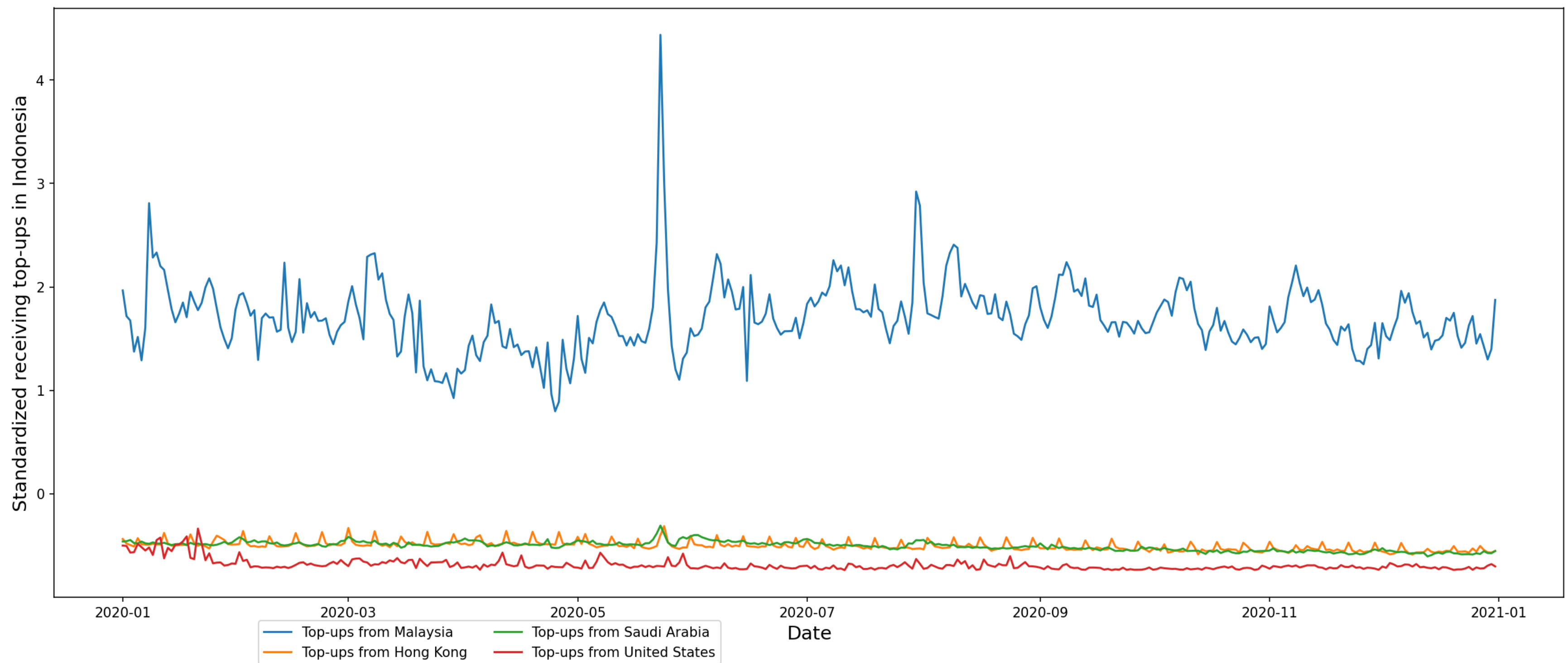

Figure 3: The top sending countries to Indonesia.

Model	AIC	BIC	RMSE	MAE	MAPE
(3,1,2)	673.14	706.11	0.822	1.097	0.696
(1,1,2)	672.77	698.41	0.404	0.539	0.347

Table 6: Indonesia: The two candidate models for the prediction model.

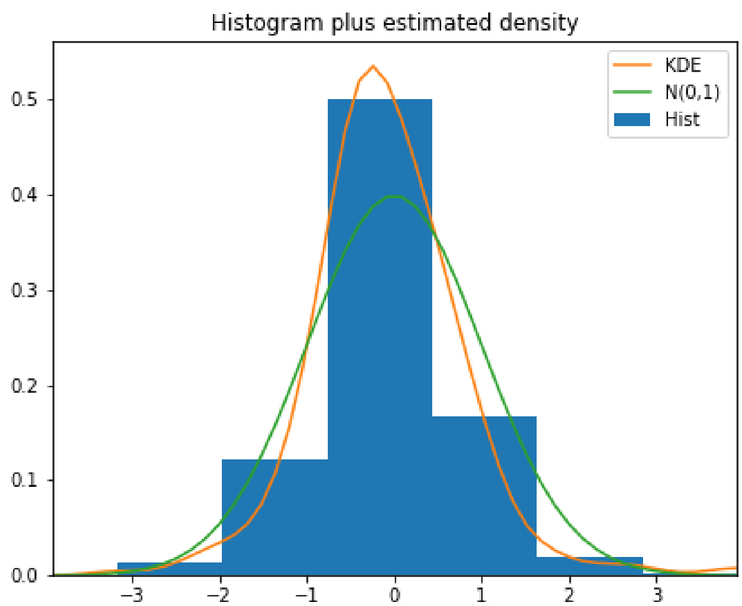

Figure 4: The histogram of the fitted model for Indonesia.

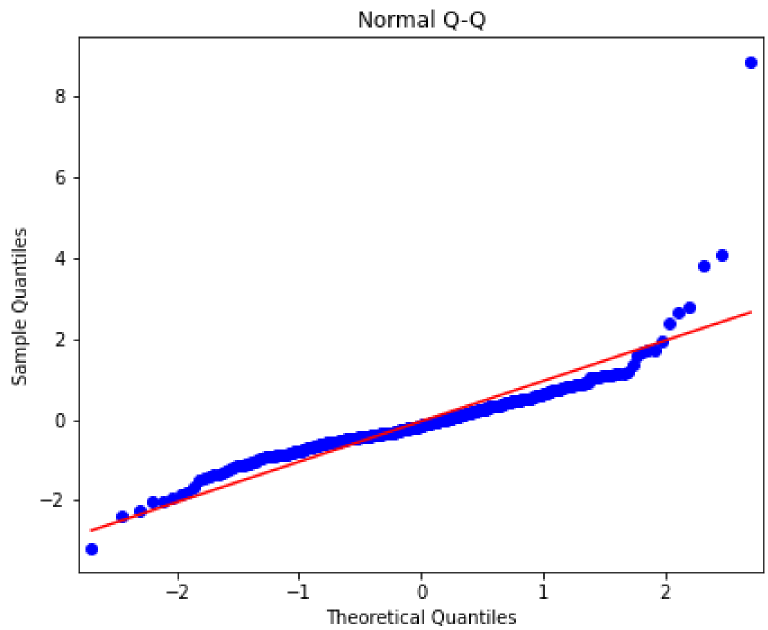

Figure 5: The QQ plot of the fitted model for Indonesia.

Applying the "Auto Arima" as mentioned in the Methodology section, we can discover the optimal parameters of the model. Among multiple iterations that happened using this library for the data corresponding to Indonesia, it concluded that the fittest values for the order (p,d,q) were (3,1,2) with an $AIC = 676.60$. However, when these parameters were fitted on the model, the latter did not perform well. Choosing manually the parameters from the ACF and PACF plots, the best parameters seem to be (1,1,2). Indeed, from table 6, it can be seen that both the AIC and BIC of this model are less than the previous one. Besides, the validation errors are lower than the ones from the first model.

The next step is to observe the QQ and histogram plots of the values. With these plots, one can determine if there is a normal distribution of the data. In figures 4 and 5, the results are presented for the chosen model with order $(1, 1, 2)$. The histogram shows that the data are normally distributed. Given that there is a single peak, this considers being a unimodal distribution.

As for the QQ plots, the ideal would be to see all the blue points to form a straight line as the red one. Hence, it would indicate that the data is

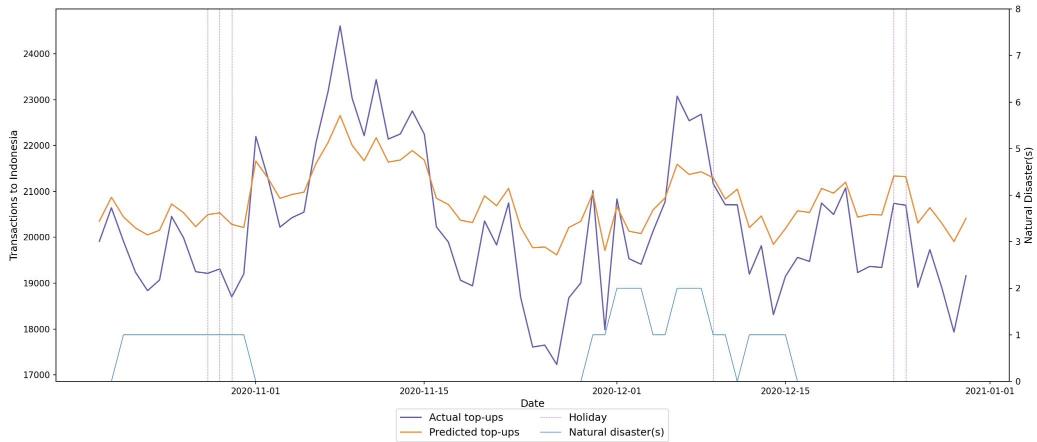

Figure 6: The performance of the model in the out-of-sample data for Indonesia.

distributed normally. In the case of figure 5, the blue points match the red line to a sufficient degree, yet there are some outliers. Nevertheless, it is a fairly safe assumption that the data is normally distributed and the model could be used for predictions.

After passing the diagnostic tests, the fitted model was tested in out-of-sample data. In table 6, the MAE, MSE, and RMSE validation methods are presented for the model (1,1,2) that was used for the predictions. All the measures were tested with the normalized data and not the actual values of the data points.

In figure 6 the orange line denotes the predicted values and the dark blue line the ground truth data. Along with these factors, the data for natural disasters and public holidays are presented in the plot.

From the same figure, it seems that the fitted model can predict some of the peaks that exist in the data, but the performance does not seem to be the optimal one. The model follows the trend and the fluctuations of the ground truth line. However, even if it reflects the same trend, it cannot predict the actual values in some cases. For other days, the predicted values are close to the real ones.

5.3 Case study: Ethiopia

5.3.1 International receiving top-ups, domestic flows, and the impact of natural disasters and public holidays on them

In figure 7, the most important natural disasters and the public holidays are shown for Ethiopia. It should be mentioned that there were missing values for the domestic top-ups in Ethiopia.

Ethiopian Christmas is celebrated on 7^{th} of January, and it can be seen that both international and domestic flows reach a peak on this day. At the same time, there were two disasters taking place. During this period, the domestic top-ups have their ups and downs, and likely the epidemic and the insect infestation could explain this behavior. However, international migrants do not seem to support their origin country during these two disasters.

The next spike of the domestic remittances can be explained by the fact that the government of Ethiopia declared a state of an emergency $(08/04/20)$ due to many Covid-19 cases. However, they did not proceed to lockdown.

From the airtime top-ups data, it seems that the rest of the natural disasters did not have a significant role in receiving remittances. International top-ups grow when a public holiday occurs (Ethiopian Easter, Eid al-Fitr, Eid Al Adha, Ethiopian New Year, and Moulid). However, the domestic flows do not rise during these days.

Even though there was a flood that befell on $20/04/2020$, with more than 210.000 individuals affected, the receiving remittances did not change. Only the domestic flows appeared to peak at the end of the flood, but it is not as remarkable as others.

Lastly, on 25^{th} of December, the top-ups within Ethiopia were increased. Even though the most population of Ethiopia is Orthodox Christian (43.5%) [9] and Muslims consist the 34%, the receiving international remittances do not seem to grow on Orthodox Christmas day. On the other hand, one can assume that Muslims living abroad, transmit money more frequently than Christians actually do during these days.

[9] https://culturalatlas.sbs.com.au/ethiopian-culture/ethiopian-culture-religion

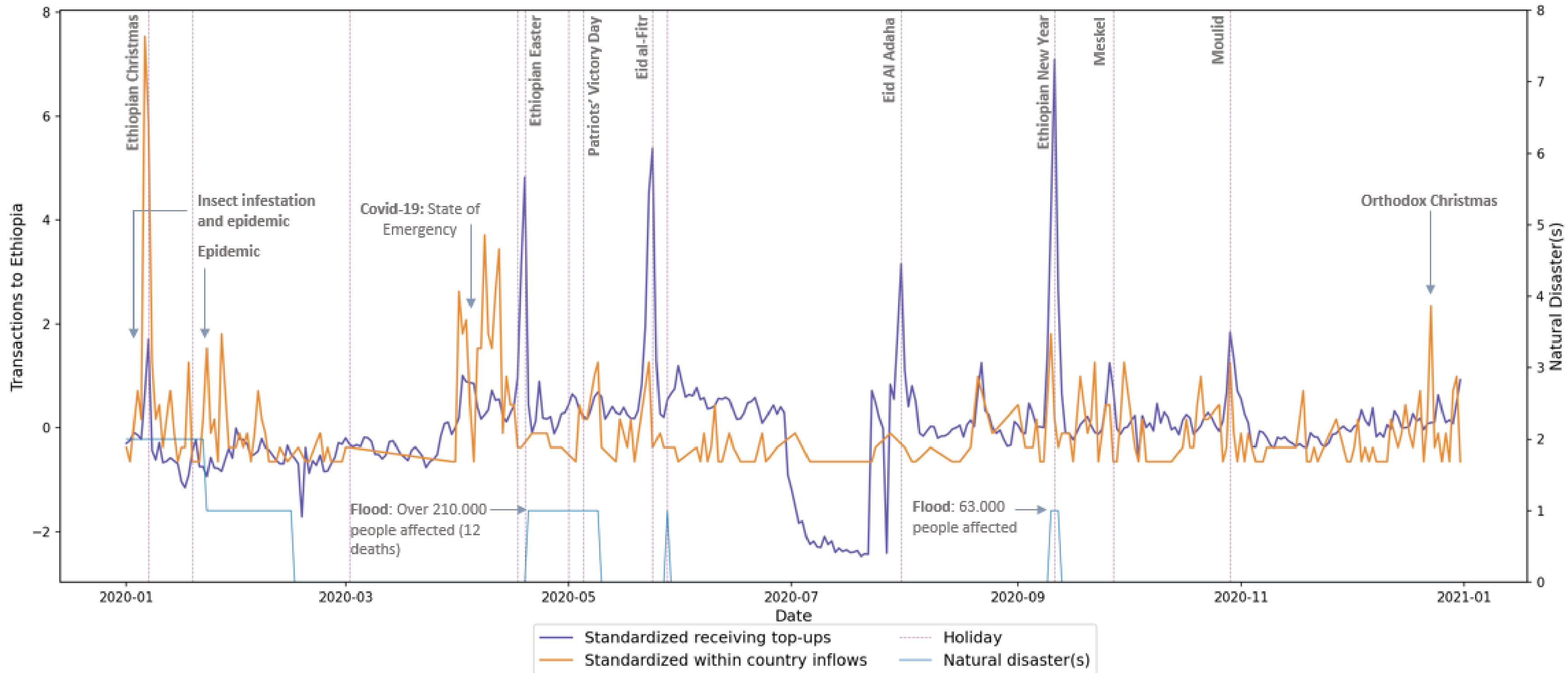

Figure 7: The international and domestic top-ups in Ethiopia.

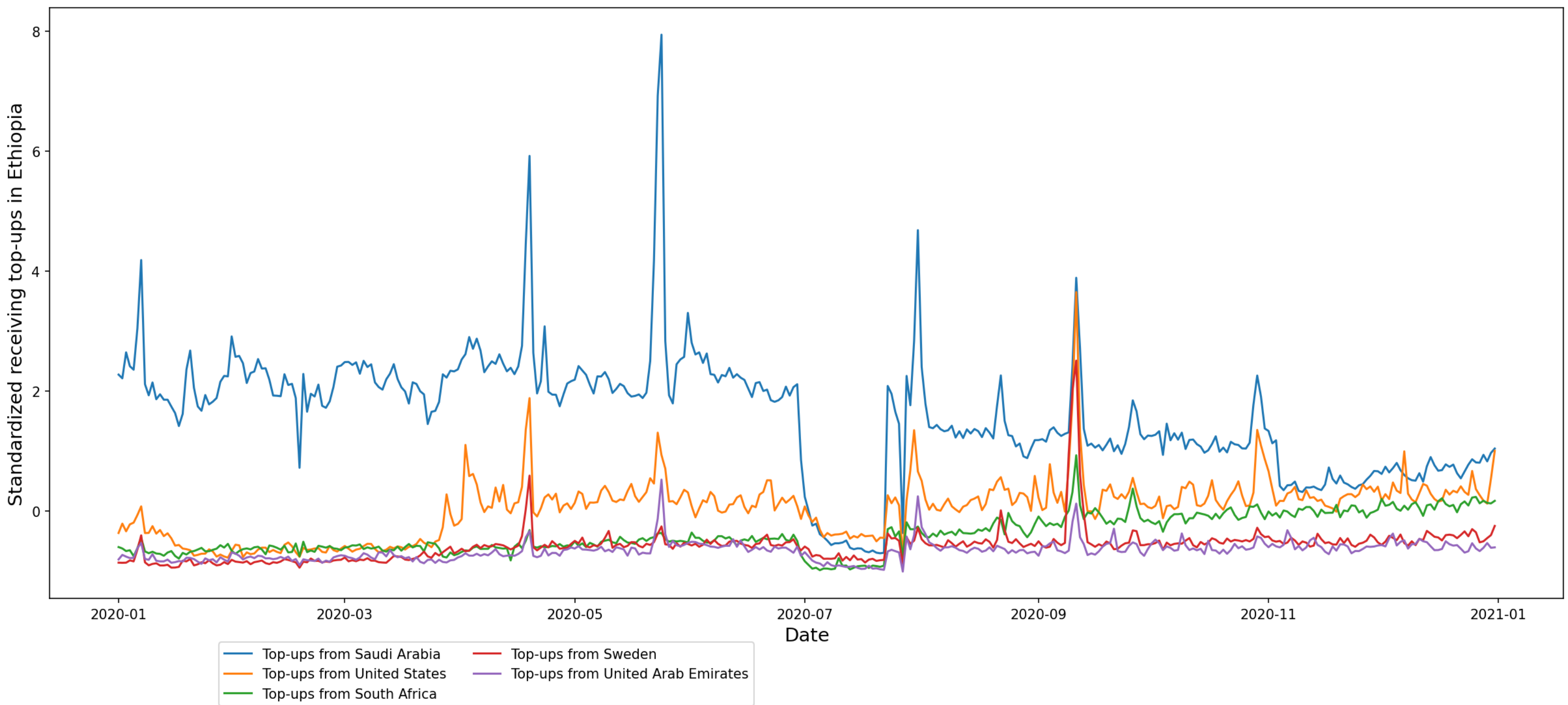

Figure 8: The top sending countries to Ethiopia.

Model	AIC	BIC	RMSE	MAE	MAPE
(0,1,2)	594.63	616.61	0.212	1.037	0.113
(1,1,1)	596.80	618.80	0.243	1.193	0.137

Table 7: Ethiopia: The two candidate models for the prediction model.

5.3.2 Top sending countries to Ethiopia

The top five countries that have the most influence on the receiving airtime top-ups in Ethiopia for 2020 are the following; Saudi Arabia, United States, South Africa, Sweden, United Arab Emirates. Figure 8 reveals that almost for the whole year, Saudi Arabia was the most promising country transferring top-ups to Ethiopia. However, for the last two months of the year, the five top sending countries have an equivalent trend and there are not many differences in the daily sending amounts from each one of them.

There are specific significant days that international migrants transmit a vast amount of airtime top-ups to Ethiopia. For instance, $19/04/2020$, $24/05/2020$, $31/07/2020$, $11/09/2020$, and $29/10/2020$ are some days where the receiving top-ups from most of the sending countries have a peak. As previously noted, these spikes are highly correlated with the holidays of Ethiopia.

5.3.3 Prediction model for Ethiopia

The daily data for Ethiopia did not have seasonality as in the case of Indonesia. Therefore, an ARIMAX model was developed for this case study.

Finding the best model was considerable challenge since all the possible combinations had a small difference between the AIC and BIC values. Nevertheless, in table 7 the two best possible solutions for this case study are presented. The ARIMAX model that best straddled the requirements of the goodness of fit, was the one with the first order $(0, 1, 2)$. From the results, this model had slightly lower values than the $(1, 1, 1)$. For the latter, the parameters were chosen manually by looking at the ACF and PACF plots.

For the figures 9 and 10 which correspond to the histogram and QQ respectively, it is clear that the data is not so much normally distributed. In the QQ plot, the blue points do not align with the red line. As mentioned before, different (p,d,q) order combinations were tested. However, the diagnostic tests were not better than these two figures. The histogram

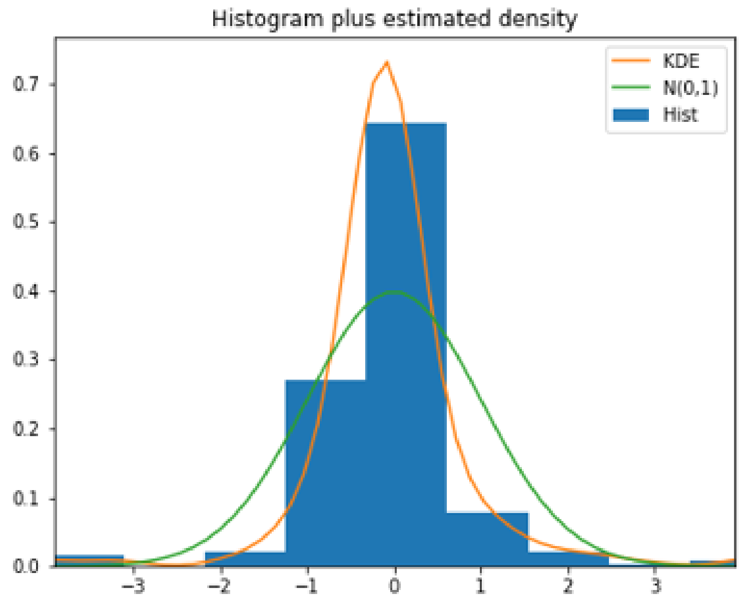

Figure 9: The histogram of the fitted model for Ethiopia.

and QQ plots for the chosen model (0,1,2) performed better among other candidate models.

Figure 11 presents the performance of the fitted model to the out-of-sample data. There are no natural hazards that happened during this period in Ethiopia. Therefore, there is no evidence if the model predicts correctly the airtime top-ups when a disaster occurs. Moreover, the model was not able to predict accurately the peak that happened on $29/10/2020$. It manages to predict that the top-ups will increase, however this was seen after one day from the actual peak. The model predict the airtime top-ups with higher accuracy for December compared to the previous two months.

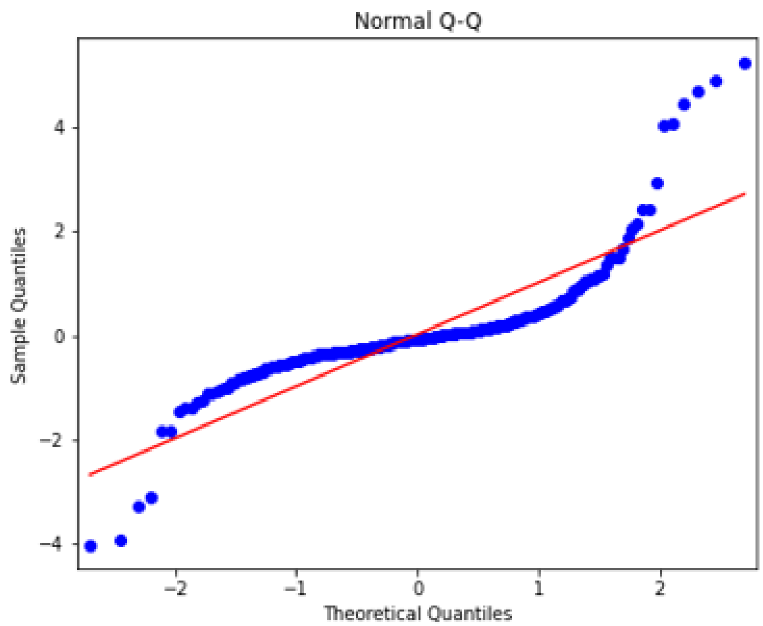

Figure 10: The QQ plot of the fitted model for Ethiopia.

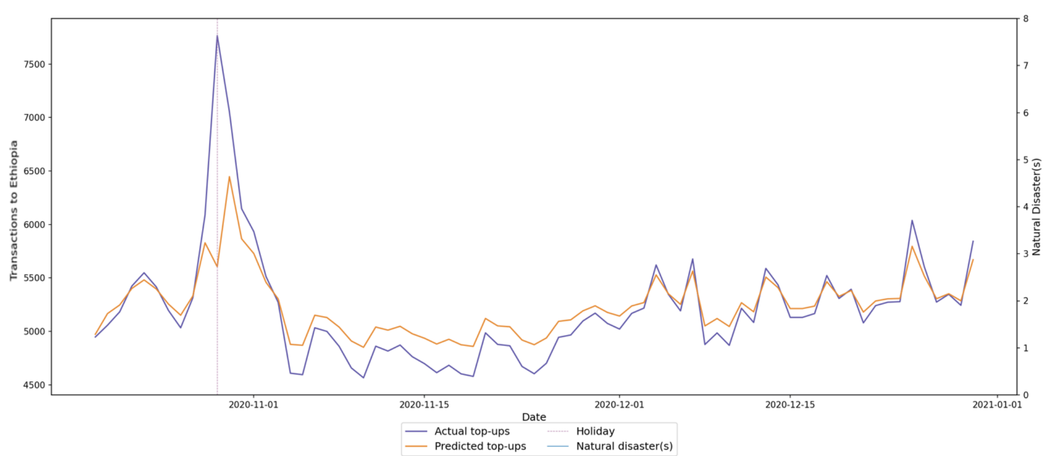

Figure 11: The performance of the model in the out-of-sample data for Ethiopia.

5.4 Case study: Philippines

5.4.1 International receiving top-ups, domestic flows, and the impact of natural disasters and public holidays on them

According to World Bank, the Philippines were among the top five countries that received the most remittances in 2020. It is interesting to correlate these results, with the airtime top-ups during this period. As in the case of Indonesia and Ethiopia, the outcomes from this section answer the second research question.

In figure 12 the domestic and international flows are presented in the Philippines during 2020. In the first quarter of the year, the top-ups sent within the country are higher than the ones received from abroad. The domestic flows dropped significantly after the public health emergency that the government announced on March 2020 [10]. On the other hand, international migrants supported their households back to the Philippines. This can be seen by the huge amount of top-ups that were being sent the following months, compared to the first months of the year where Covid-19 has not spread yet.

As with the previous case studies, there are some national or religious holidays that the receiving airtime top-ups rise. On Good Friday, Labor Day, Eid al-Fitr, Eid ul-Adha, Ninoy Aquino, National Heroes' Day, and Christmas are days where the international top-ups had peaks. On the other hand, the domestic flows do not have a significant role on some of these days (e.g. Good Friday, Labor Day, etc.).

It should be mentioned that the majority of the population (almost 81%) is Catholic, and 5% Muslim [11]. Even though Muslims in the Philippines consist a small portion of the population, the second-highest spike of the receiving international top-ups throughout 2020 happened on Eid al-Fitr which is a Muslim holiday.

Apart from the influence of holidays on airtime top-ups, it can be seen from the plot that natural disasters are impacting the receiving remittances. For instance, two storms happened in four continuous days, the occurring on 28^{th} and 31^{st} of October. The former influenced Philippines with 880.000 affected individuals, yet the consequences of the second storm were more

[10] https://time.com/5945616/covid-philippines-pandemic-lockdown/

[11] https://geriatrics.stanford.edu/ethnomed/filipino/introduction/religion.html

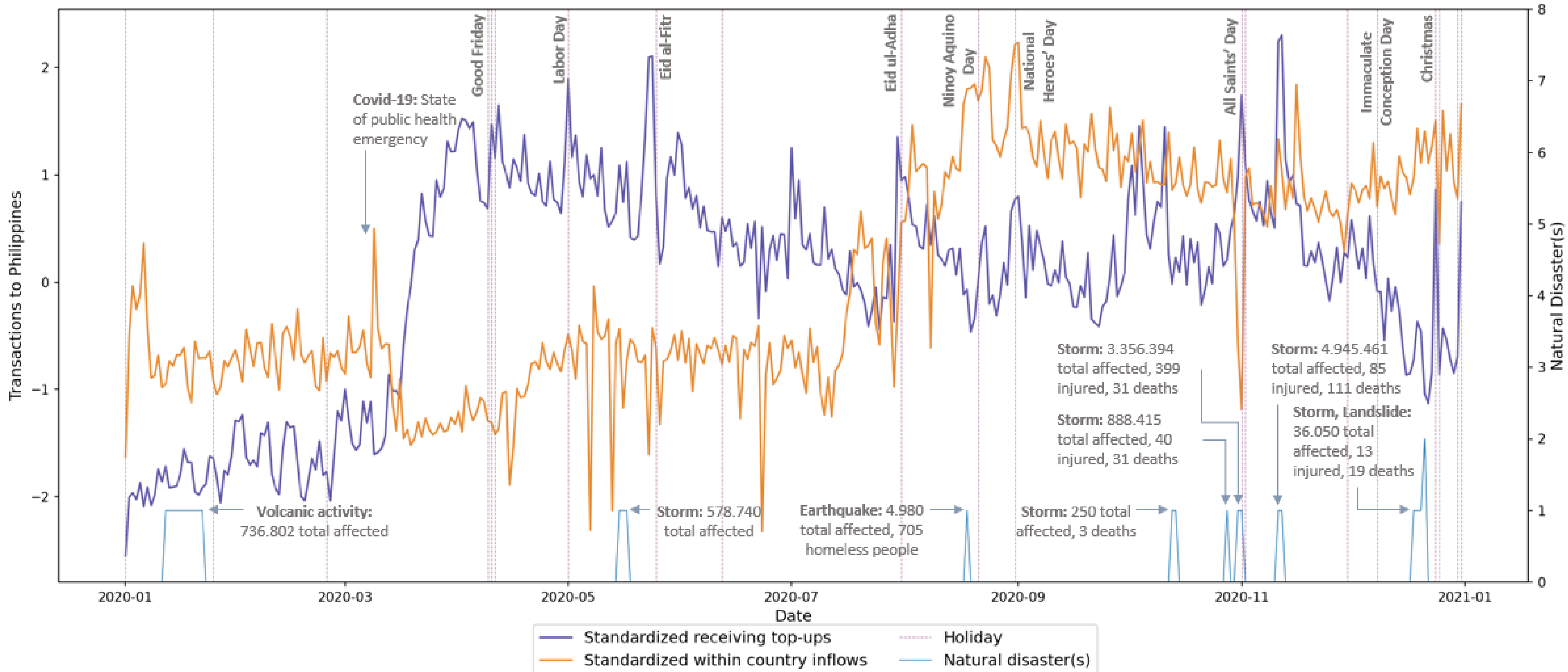

Figure 12: The international and domestic top-ups in Philippines.

than 3 million affected individuals and 31 deaths. The migrants living abroad supported their people back to the Philippines and this can be explained by the sharp spike the same day of this storm. Yet, the top-ups within the country had a notable fall when this disaster occurred.

The major and most recent natural disaster that happened in the Philippines, was a storm on $11/11/2020$ with almost 5 million affected individuals. Both international and domestic flows increased this day.

In conclusion, the Philippines was a country with severe storms that occurred in 2020. Countless people accepted damaging consequences, yet international migrants assisted them by transmitting airtime top-ups. From these subsections of the three countries, we could answer the second research question and understand the behavior of the migrants supporting their families. Moreover, our results from the regression model in the previous section showed that the storms affect the receiving remittances the most. This finding aligns with the analysis in this section. The results contribute to the literature since no prior research has been conducted utilizing such a huge amount of airtime top-ups data before.

5.4.2 Top sending countries to the Philippines

In figure 13 the top five sending countries to Philippines are displayed. The most influential country transferring airtime top-ups to the Philippines is Saudi Arabia. The United Arab Emirates and Hong Kong are the next two countries that have an impact on the sending transactions. After that, Malaysia and United States are two countries without a significant performance in the overall receiving remittances to the Philippines.

According to World Bank's report for 2020, the top countries for sending remittances to the Philippines were the United States with 40%, Singapore with 7%, and Saudi Arabia 6%. The total amount of remittances from the United Arab Emirates was merely 4.3%. Comparing these results with the top sending countries of airtime top-ups to the Philippines, it can be understood that there are some differences when using airtime top-up services. For instance, Saudi Arabia contributed to a small portion of the remittances in the Philippines, yet for the airtime top-ups, it was the major leader of the receiving top-ups.

Due to Covid-19 strict lockdown on March 2020, it can be seen that only the airtime transfers from Saudi Arabia started to rise. The United Arab Emirates and Hong Kong had small growth after March, but not something

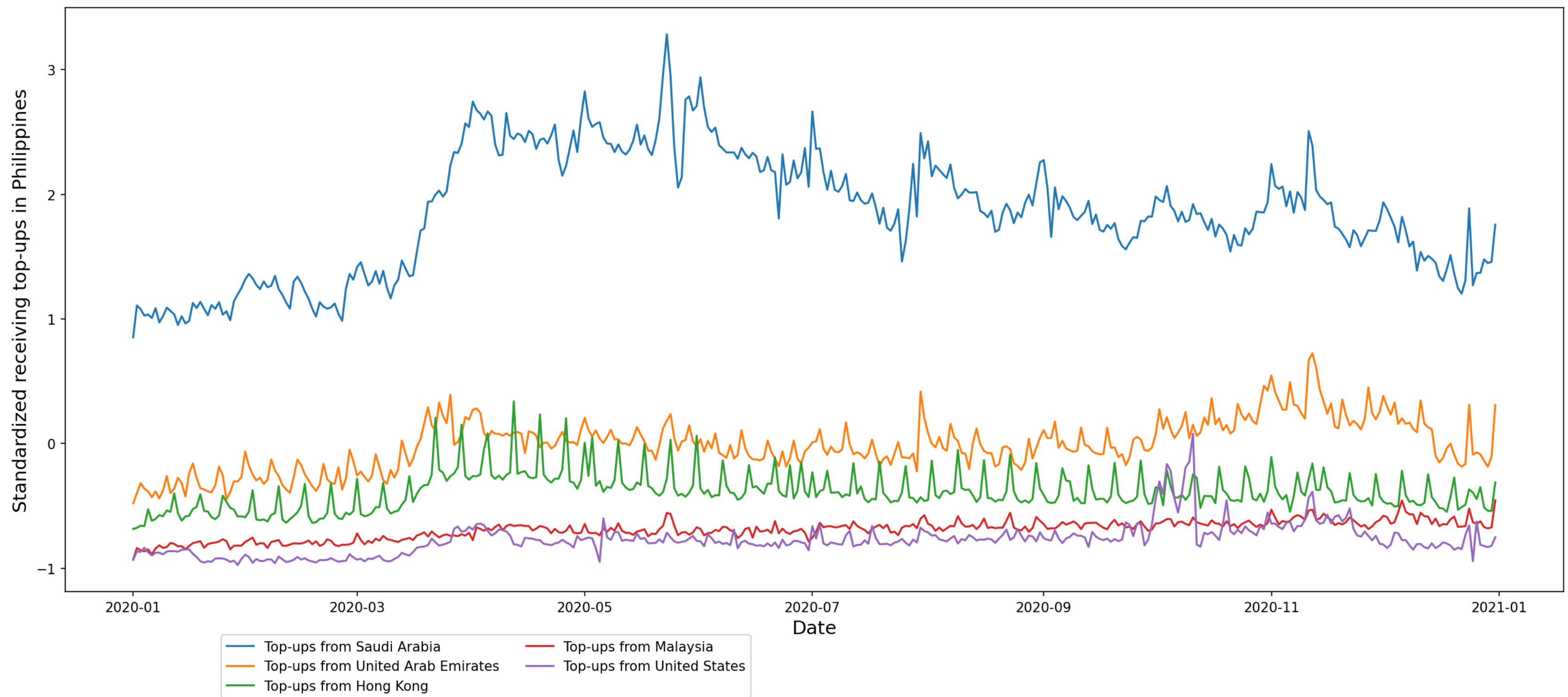

Figure 13: The top sending countries to Philippines.

Model	AIC	BIC	RMSE	MAE	MAPE
(3,0,2)x(0,0,0,12)	146.13	179.22	0.118	0.636	0.102

Table 8: Philippines: The results of the SARIMAX model.

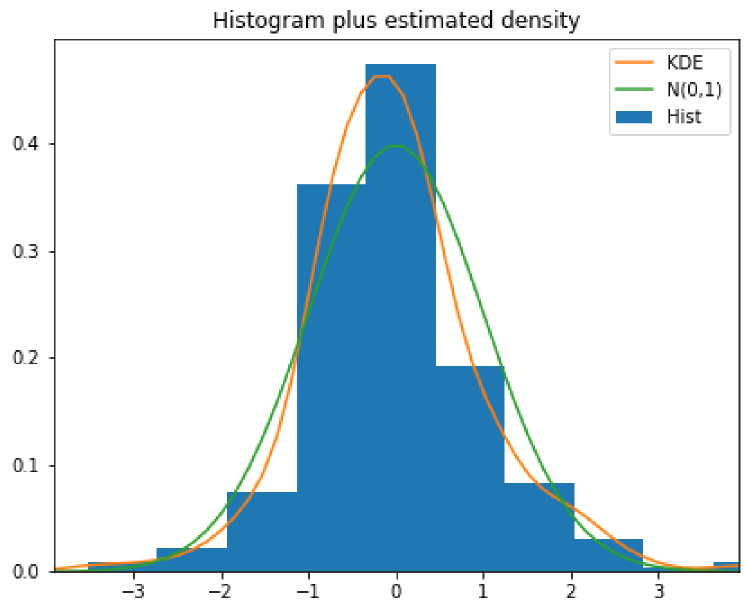

Figure 14: The histogram of the fitted model for Philippines.

withstanding. Lastly, Malaysia and United States did not have a significant impact on the receiving airtime top-ups in the Philippines.

5.4.3 Prediction model for the Philippines

In the case of the Philippines, the data moderately showed to have a seasonality during 2020. The "Auto Arima" library revealed the same result. Nevertheless, the best SARIMAX model that was found for the Philippines is as follows: $(3, 0, 2)x(0, 0, 0, 12)$.

The decision for the best model was challenging because it was time-consuming to test all the potential candidates of the parameters. In general, when one adds the seasonal parameters to obtain a qualified model, the time for training each model surprisingly increases. Possibly there are more solid models than the one that was chosen, yet it was not feasible to explore

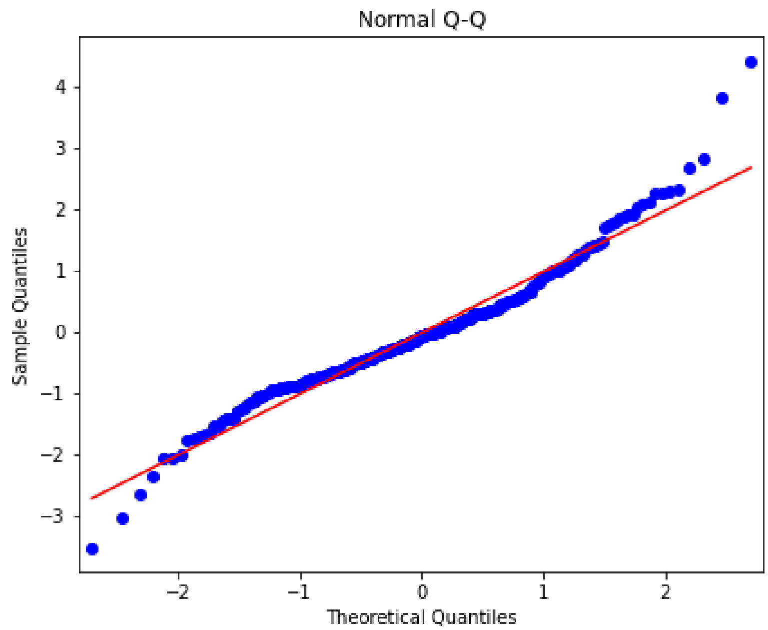

Figure 15: The QQ plot of the fitted model for Philippines.

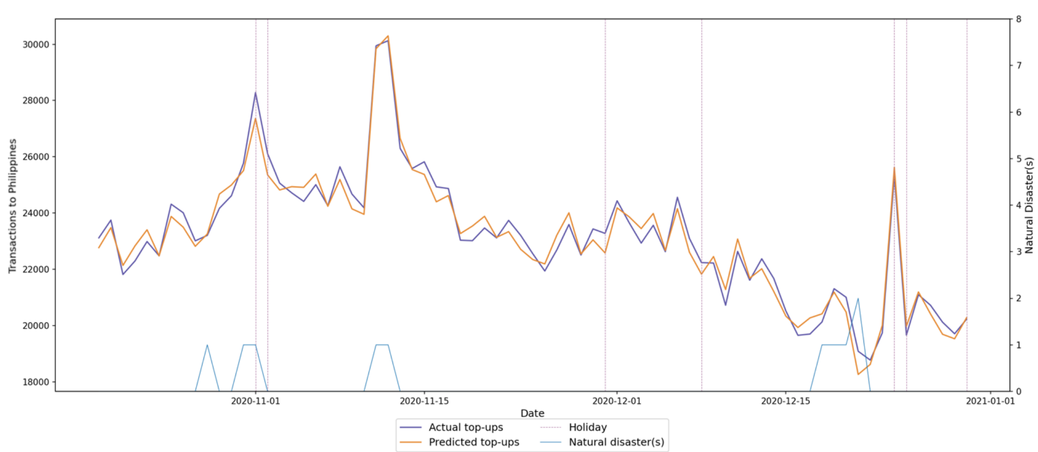

Figure 16: The performance of the model in the out-of-sample data for the Philippines.

all of the combinations.

As shown in the previous case studies, the figures 14 and 15 display the diagnostic tests for the fitted model. In the QQ and histograms plots, it can be seen that the data is normally distributed. However, some of the data points are outliers in the overall outline of the fitted model.

In figure 16, the performance of the model is presented with the orange line comparing to the ground truth data (dark blue line). The model can predict accurately the peaks existing on the holidays and when a natural disaster occurs. In general, the SARIMAX model appears to perform adequately. However, since there is a chance that another model would have lower AIC, BIC, and validation measures, it means that this result is not the optimal one. Therefore, further investigation in this SARIMAX case should happen to test the rest of the parameters' combinations.

6 Discussion

6.1 Effectiveness of the results

The airtime top-up credits reflect a new era of mobile transactions. Limited papers are analyzing that kind of data, therefore the contribution of this thesis to the literature could be valuable for future researches.

One of the main questions in the thesis was about the influence of natural disasters and the public holidays on top-ups. There is not rich literature on this topic, particularly for the effects of the disasters on them. Given that the airtime top-ups dataset corresponded to the top fifteen receiving countries, we could safely assume that the results reflect approximately the actual influence of the disasters and holidays on these mobile transactions. In the first research question, this relationship was answered.

The first finding was obtained by developing a regression model. Storms, droughts, flood, and epidemics appear to be the major natural catastrophes that affect the airtime top-ups. After the regression model, these results were tested individually for Indonesia, Ethiopia, and the Philippines which were the case studies of the thesis. We observed that there are peaks in the data reflecting the effects of these natural hazards. Moreover, for each case study, the severity of the disasters was mentioned to comprehend what was the physical impact that they had (total affected individuals, deaths, etc.).

Apart from the disasters, some of the public, national, and religious holidays of each origin country seem to influence the receiving top-ups. For Muslims, it was important to transfer mobile credits to their relatives, mainly on Eid days. For Christians, Easter and Christmas were the two most vital days to transmit top-ups to their families.

Given that some natural hazards and public holidays affect the airtime top-ups, we derived some inferences about the behavior of the international migrants. Individuals residing outside of their home countries, assist their families in case of severities. Apart from that, they also support relatives and friends to celebrate important days. Hence, these outcomes answer the second research question which was analyzed in the second subsection of each case study.

The last research question was about the potentiality of external factors that may be useful to predict the airtime top-ups. For each case study, a prediction model was developed with the predicted variable corresponding to the mobile credits. The independent variables that were implemented in the model and that could be useful to predict the top-ups, were about natural disasters and public holidays. To the best of our knowledge, there have not been studies utilizing daily data of airtime credits to predict the next day values. These findings act as a contribution to the literature, and other researchers could find the outcomes serviceable for the future development of other potential machine learning models.

6.2 Limitations

During the thesis, a variety of different shortcomings contributed negatively to the smooth process of the analysis.

The data itself was one of the most vital problems that needed to be faced. The limited data that the company provided us, was not adequate to analyze in-depth the airtime top-ups. Furthermore, since the natural data was geotagged, it will be much more informative and reliable if the company shared with us that kind of top-ups data.

One of the main research questions that were initially addressed for the thesis, was to develop a variety of machine learning models. Eventually, the aim was to compare these models and conclude which one was the most effective in predicting the airtime top-up transactions. As mentioned in Section 2, we decided to experiment with an LSTM model with limited data. The model did not work successfully since the amount of information was not sufficient enough to train the model.

Another limitation was the lack of crucial information about the natural catastrophes in the EM-DAT dataset. There were specific factors that could be used as input to the ARIMAX and SARIMAX models for the predictions. These values correspond to the total affected individuals, the total homeless people, or even the deaths to show the severity of a natural disaster. Many missing points occur in the dataset and there was not an option to fill these gaps. Since the effects of a disaster are shown during the first days of its occurrence in a region, the aim was to positive skew that kind of information and add the values in the models as independent factors. However, due to the lack of multiple missing values, this occurred with no success.

The seasonality could not be accurate for the analysis of the three countries. Given that the data corresponding to one year, it was hard to interpret if there is seasonality. Moreover, the SARIMAX model for the Philippines requires improvements. As discussed in Section 5, there could be another potential model that performs better.

Finally, in Section 2 we mentioned that there are factors that influence the remittance flows and the airtime top-ups (exchange rates, GDP, etc.). These external factors could not be handled for the models, since both ARIMAX and SARIMAX models require sequentially daily data for all the datasets that are used. It was a challenge to find daily data for GDP per country, migration flows, or exchange rates, yet there are no such comprehensive datasets available online.

6.3 Future work

Based on the findings in this thesis, future research could be done in this field. First of all, obtaining extra airtime top-up credits data would be helpful to extend this study. Besides, that was one of the main limitations of this thesis. If one could gain data for more years, the results would be more detailed than the ones presented in this thesis. Moreover, the seasonality effects will be obvious, and the prediction models (such as ARIMAX and SARIMAX) could be used more precisely.

By sharing more data by our partner or other companies utilizing airtime top-up data, the development of new models can take place. This limitation was mentioned in the Sections 2, 5, 6 and it was one of the main obstacles in this thesis. Therefore, other short and long-term ML models could be developed to predict the next days' top-up values.

As discussed in this thesis, the natural disasters dataset has geotagged information. If the company shares that kind of mobile top-ups information in the future, new research could commence. The results would reveal precisely the influences of natural disasters on the receiving top-ups in specific cities or provinces of a country.

As a last potential future work is to import the severities of natural disasters in the models. We stated that this could not be done using the ARIMAX and SARIMAX models.

7 Conclusion

In this thesis, we examined the influence of the natural disasters and holidays of the origin countries, on the airtime top-up credits for 2020. The natural hazards dataset was about the most severe ones that happened in this period. The holidays corresponded to public, national, and religious days. Specifically, three research questions were defined and answered in this thesis.

For the first research question, we examined the relationship between natural disasters and public holidays with mobile top-ups. The findings from the development of an OLS regression model revealed that droughts, storms, floods, and epidemics are the most critical disasters. Based on the effects that they produced, the receiving airtime top-ups increased for the countries that these natural catastrophes occurred.

Religious days appeared to have a positive correlation with the rise of mobile credits to a variety of origin countries. Eid al-Fitr and Eid ul-Adha were the two Muslim days that international migrants transmit top-ups to their home countries to support their families in these days. Moreover, Orthodox holidays such as Easter and Christmas were the most influenced religious days that the receiving mobile credits grew. For the second research question, we attempted to find some insights about migration and the correlation with the airtime top-ups. We discovered that international migrants support their families back to their home countries when a dreadful disaster occurs. As mentioned earlier, the religious days for Muslims and Christians appear to be the most meaningful holidays for migrants. These conclusions suggest that this group of people have a supportive role in both happiness and sorrow. More insights can be found in future studies about migrants' behavior employing airtime top-ups.

Finally, ARIMAX and SARIMAX models were developed for the three case studies of Indonesia, Ethiopia, and the Philippines. The exogenous factors of natural disasters and public holidays appeared to have a substantial part in the models. However, improvements need to take place and there were several limitations throughout the development of them. To the best of our knowledge, this thesis presented for the first time short-term prediction models using airtime top-ups. These models could be used to determine the next-day values of the mobile credits for the three countries mentioned before.

References

[1] Faisal Abbas, Amjad Masood, and Arifa Sakhawat. What determine remittances to pakistan? the role of macroeconomic, political and financial factors. *Journal of policy modeling*, 39(3):519–531, 2017.

[2] Richard H Adams Jr. The determinants of international remittances in developing countries. *World development*, 37(1):93–103, 2009.

[3] Richard H Adams Jr and John Page. Do international migration and remittances reduce poverty in developing countries? *World development*, 33(10):1645–1669, 2005.

[4] Adebayo ADEDOKUN. Forecasting remittances as a major source of foreign income to nigeria: evidence from arima model. *European Journal of Humanities and Social Sciences Vol*, 24(1), 2013.

[5] Reena Agarwal and Andrew W Horowitz. Are international remittances altruism or insurance? evidence from guyana using multiple-migrant households. *World development*, 30(11):2033–2044, 2002.

[6] Selçuk Akçay. Remittances and misery index in turkey: is there a link? *Applied Economics Letters*, 25(13):895–899, 2018.

[7] Héctor Allende, Claudio Moraga, and Rodrigo Salas. Artificial neural networks in time series forecasting: A comparative analysis. *Kybernetika*, 38(6):685–707, 2002.

[8] Nari Sivanandam Arunraj, Diane Ahrens, and Michael Fernandes. Application of sarimax model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems (IJORIS)*, 7(2):1–21, 2016.

[9] Marlene Attzs. *Natural disasters and remittances: Exploring the linkages between poverty, gender and disaster vulnerability in Caribbean SIDS*. Number 2008/61. WIDER Research Paper, 2008.

[10] Bilgecag Aydogdu, Hanif Samad, Bai Shiqi, Sami Abboud, Ilias Gorantis, and Albert A Salah. Analysing mobile remittances via international airtime top-up transfer data. *Manuscript submitted for publication*, 2021.

[11] World Bank. *Phase II: COVID-19 Crisis Through a Migration Lens*, volume Migration and Development Brief,no. 33. World Bank, Washington, DC., oct 2020.

[12] World Bank. *Resilienc: COVID-19 Crisis Through a Migration Lens*, volume Migration and Development Brief,no. 34. World Bank, Washington, DC., may 2021.

[13] Giulia Bettin and Alberto Zazzaro. The impact of natural disasters on remittances to low-and middle-income countries. *The Journal of Development Studies*, 54(3):481–500, 2018.

[14] Joshua E Blumenstock, Nathan Eagle, and Marcel Fafchamps. Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157–181, 2016.

[15] George J Borjas. Economic theory and international migration. *International migration review*, 23(3):457–485, 1989.

[16] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[17] Catherine Bragg, Glenn Gibson, Haleigh King, Ashley A Lefler, and Faustin Ntoubandi. Remittances as aid following major sudden-onset natural disasters. *Disasters*, 42(1):3–18, 2018.

[18] Peter J Brockwell, Richard A Davis, and Matthew V Calder. *Introduction to time series and forecasting*, volume 2. Springer, 2002.

[19] Jeffrey H Cohen. Remittance outcomes and migration: Theoretical contests, real opportunities. *Studies in comparative international development*, 40(1):88–112, 2005.

[20] Clara Dismuke and Richard Lindrooth. Ordinary least squares. *Methods and Designs for Outcomes Research*, 93:93–104, 2006.

[21] Nathan Eagle, Yves-Alexandre de Montjoye, and Luís MA Bettencourt. Community computing: Comparisons between rural and urban societies using mobile phone data. In *2009 international conference on computational science and engineering*, volume 4, pages 144–150. IEEE, 2009.

[22] Tingting Fang and Risto Lahdelma. Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. *Applied energy*, 179:544–552, 2016.

[23] International Monetary Fund. The macroeconomic effects of global migration. In *World Economic Outlook: The great lockdown*, chapter 4, pages 77–101. International Monetary Fund, apr 2020.

[24] Carlos R García-Alonso, Esther Arenas-Arroyo, and Gabriel M Pérez-Alcalá. A macro-economic model to forecast remittances based on monte-carlo simulation and artificial intelligence. *Expert Systems with Applications*, 39(9):7929–7937, 2012.

[25] Susan W Gikungu, Anthony G Waititu, and John M Kihoro. Forecasting inflation rate in kenya using sarima model. *American Journal of Theoretical and Applied Statistics*, 4(1):15–18, 2015.

[26] Paola Giuliano and Marta Ruiz-Arranz. Remittances, financial development, and growth. *Journal of development economics*, 90(1):144–152, 2009.

[27] World Bank Group. Leveraging economic migration for development, sep 2019.

[28] D Guha-Sapir, R Below, and P Hoyois. Em-dat: International disaster database.(cred), centre for research on the epidemiology of disasters, 2014.

[29] Mrinal Gupta, Ayman Abdelmaksoud, Mohammad Jafferany, Torello Lotti, Roxanna Sadoughifar, and Mohamad Goldust. Covid-19 and economy. *Dermatologic therapy*, 2020.

[30] Sanjeev Gupta, Catherine A Pattillo, and Smita Wagh. Effect of remittances on poverty and financial development in sub-saharan africa. *World development*, 37(1):104–115, 2009.

[31] Andrew C Harvey. Forecasting, structural time series models and the kalman filter. 1990.

[32] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[33] UNDESA IOM. Migration and human mobility, May 2012.

[34] George Edwards Johnson and W Edward Whitelaw. Urban-rural income transfers in kenya: an estimated-remittances function. *Economic Development and Cultural Change*, 22(3):473–479, 1974.

[35] Jake Kendall, Bill Maurer, Phillip Machoka, and Clara Veniard. An emerging platform: From money transfer system to mobile money ecosystem. *Innovations: Technology, Governance, Globalization*, 6(4):49–64, 2011.

[36] Jeroen Klomp and Kay Valckx. Natural disasters and economic growth: A meta-analysis. *Global Environmental Change*, 26:183–195, 2014.

[37] Loïc Le Dé, JC Gaillard, and Wardlow Friesen. Poverty and disasters: do remittances reproduce vulnerability? *The Journal of Development Studies*, 51(5):538–553, 2015.

[38] Everett S Lee. A theory of migration. *Demography*, 3(1):47–57, 1966.

[39] Joseph Mawejje and Paul Lakuma. Macroeconomic effects of mobile money: evidence from uganda. *Financial Innovation*, 5(1):1–20, 2019.

[40] Martin McKee and David Stuckler. If the world fails to protect the economy, covid-19 will damage health not just now but also in the future. *Nature Medicine*, 26(5):640–642, 2020.

[41] Tigran A Melkonyan and David A Grigorian. Microeconomic implications of remittances in an overlapping generations model with altruism and a motive

to receive inheritance. *Journal of Development Studies*, 48(8):1026–1044, 2012.

[42] Cynthia Merritt. Mobile money transfer services: the next phase in the evolution of person-to-person payments. *Journal of Payments Strategy & Systems*, 5(2):143–160, 2011.

[43] IOM UN Migration. World migration report 2020.

[44] Nicolas Mirabaud. Migrants' remittances and mobile transfer in emerging markets. *International Journal of Emerging Markets*, 2009.

[45] Sanket Mohapatra, George Joseph, and Dilip Ratha. Remittances and natural disasters: ex-post response and contribution to ex-ante preparedness. *Environment, Development and Sustainability*, 14(3):365–387, 2012.

[46] Sanket Mohapatra and Dilip Ratha. Forecasting migrant remittances during the global financial crisis. *Migration Letters*, 7(2):203, 2010.

[47] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[48] Britni Must and Kathleen Ludewig. Mobile money: cell phone banking in developing countries. *Policy Matters Journal*, 7(2):27–33, 2010.

[49] Emi Nakamura. Inflation forecasting using a neural network. *Economics Letters*, 86(3):373–378, 2005.

[50] Guy P Nason. Stationary and non-stationary time series. *Statistics in volcanology*, 60, 2006.

[51] United Nations. Inequality in a rapidly changing world. `https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/02/World-Social-Report2020-FullReport.pdf`, 2020.

[52] Paul Newbold. Arima model building and the time series analysis approach to forecasting. *Journal of forecasting*, 2(1):23–35, 1983.

[53] Benjamin Ngugi, Matthew Pelowski, and Javier Gordon Ogembo. M-pesa: A case study of the critical early adopters' role in the rapid adoption of mobile money banking in kenya. *The Electronic Journal of Information Systems in Developing Countries*, 43(1):1–16, 2010.

[54] Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus and covid-19 pandemic: a review. *International journal of surgery*, 2020.

[55] OECD. *International Migration Outlook 2020*. OECD Publishing, Paris, 2020.

[56] OECD Economic Outlook. A brighter outlook but recovery will be gradual, December 2020.

[57] Ďurka Peter and Pastoreková Silvia. Arima vs. arimax–which approach is better to analyze and forecast macroeconomic time series. In *Proceedings of 30th international conference mathematical methods in economics*, volume 2, pages 136–140, 2012.

[58] John T Pohlman and Dennis W Leitner. A comparison of ordinary least squares and logistic regression. 2003.

[59] Hillel Rapoport and Frédéric Docquier. The economics of migrants' remittances. *Handbook of the economics of giving, altruism and reciprocity*, 2:1135–1198, 2006.

[60] Stephen C Rea and Taylor C Nelms. Mobile money: the first decade. *Institute for Money, Technology and Financial Inclusion Working Paper*, 1, 2017.

[61] Sharon Stanton Russell. Migrant remittances and development. *International migration (Geneva, Switzerland)*, 30(3-4):267–287, 1992.

[62] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[63] Moises Neil V Serino and Donghun Kim. How do international remittances affect poverty in developing countries? a quantile regression analysis. *Journal of economic Development*, 36(4):17, 2011.

[64] Shunrong Shen, Haomiao Jiang, and Tongda Zhang. Stock market forecasting using machine learning algorithms. *Department of Electrical Engineering, Stanford University, Stanford, CA*, pages 1–5, 2012.

[65] Bai Shiqi. Exploring the value of international mobile credits transfer. *Department of Information and Computing Science, Utrecht University*, 2020.

[66] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1394–1401. IEEE, 2018.

[67] Melkamu Wondimu Habtamu Solomon. The effect of remittances on economic growth in ethiopia: A time series analysis (1980/81-2017/18). 2020.

[68] Oded Stark and David E Bloom. The new economics of labor migration. *The american Economic review*, 75(2):173–178, 1985.

[69] Enang Bassey Udah. Remittances, human capital and economic performance in nigeria. *Journal of Human Development and Sustainability*, 4:103–117, 2011.

[70] Carlos Vargas-Silva and Peng Huang. Macroeconomic determinantsof workers' remittances: Hostversus home country's economic conditions. *Journal of International Trade & Economic Development*, 15(1):81–99, 2006.

[71] Dean Yang. Coping with disaster: The impact of hurricanes on international financial flows, 1970-2002. *The BE Journal of Economic Analysis & Policy*, 8(1), 2008.

[72] Dean Yang. Migrant remittances. *Journal of Economic perspectives*, 25(3):129–52, 2011.

[73] Hui Zou and Yuhong Yang. Combining time series models for forecasting. *International journal of Forecasting*, 20(1):69–84, 2004.