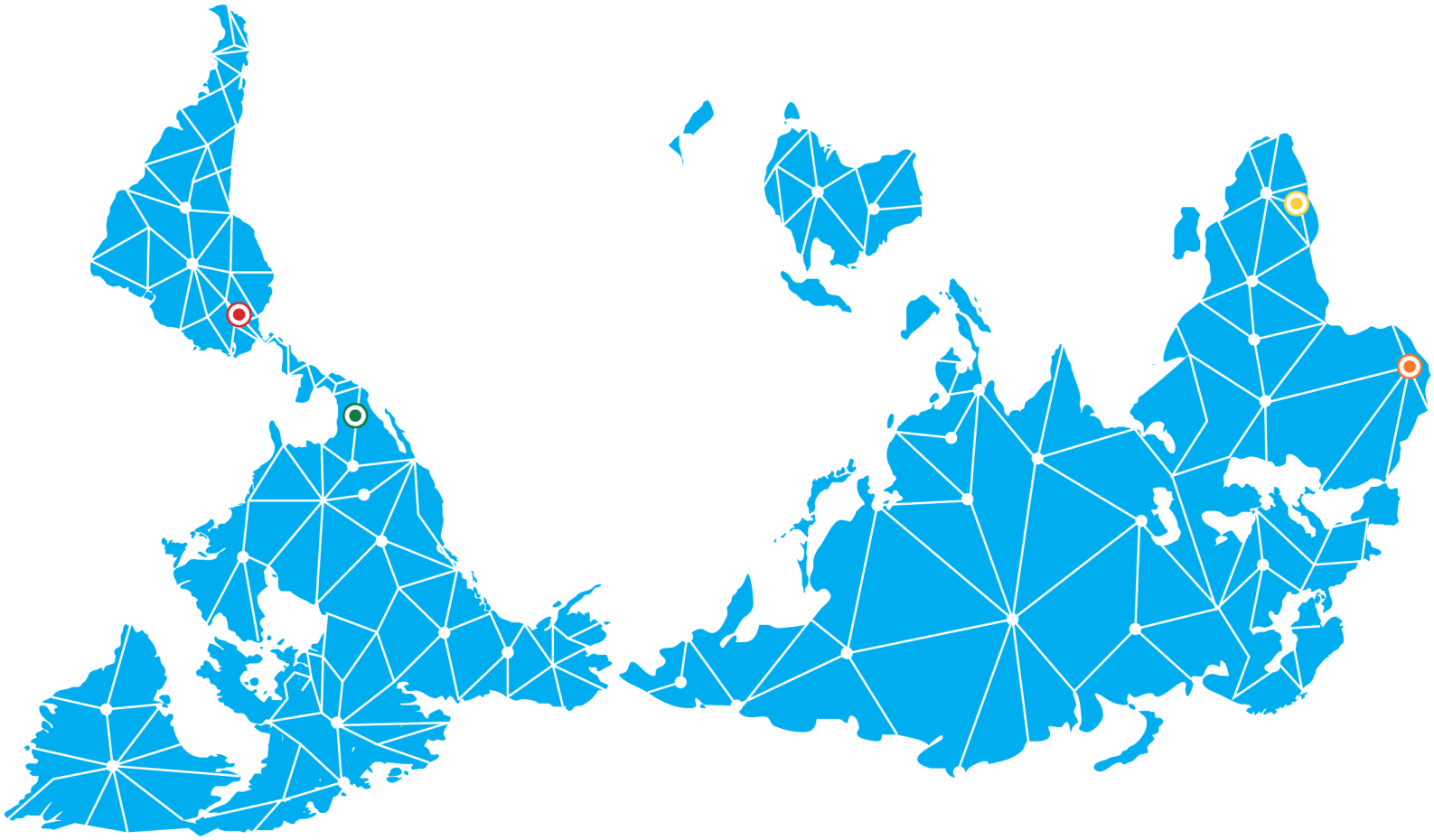
BIG DATA

TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES

2018

DATA-POP ALLIANCE



1. Socio-physical Vulnerability to Flooding in Senegal.
2. Characterizing and analyzing urban dynamics in Bogota.
3. Understanding the Relationship between Short and Long Term Mobility.
4. Large-Scale Mapping of Citizens' Behavioral Disruption in the Face of Urban Crime Shocks.

Partners:

DATA-POP ALLIANCE

Cloud to Street

FLOWMINDER.ORG

mit media lab

HARVARD HUMANITARIAN INITIATIVE

Funded by:

AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

BIG DATA
TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES
2018

DATA-POP ALLIANCE

UNDERSTANDING THE RELATIONSHIP BETWEEN SHORT AND LONG TERM MOBILITY

Sveta Milusheva - sveta.milusheva@gmail.com
Elisabeth zu Erbach-Schoenberg - elisabeth.zu.es@gmail.com
Linus Bengtsson - University of Southampton, Flowminer
Erik Wetter - University of Southampton, Flowminer
Andy Tatem - Department of Geography and Environment, University of Southampton , Flowminder

Funded by:

DATA-POP ALLIANCE FLOWMINDER.ORG AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

UNDERSTANDING THE RELATIONSHIP BETWEEN SHORT AND LONG TERM MOBILITY

Sveta Milusheva,
sveta.milusheva@gmail.com

Elisabeth zu Erbach-Schoenberg,
elisabeth.zu.es@gmail.com

Linus Bengtsson,
University of Southampton, Flowminer

Erik Wetter,
University of Southampton, Flowminer

Andy Tatem, Department of Geography and Environment, University of Southampton, Flowminder

Please cite this paper as:

MILUSHEVA, S., E. zu ERBACH-SCHOENBERG, L. BENGTSSON, E. WETTER and

A. TATEM (2017), "Understanding the Relationship Between Short and Long Term Mobility", AFD Research Paper Series, No. 2017-69, June.

Abstract

Populations are highly mobile, both in terms of long term movements of individuals relocating their place of residence as well as shorter term mobility such as commuting, seasonal travel and recreational trips. Working with call detail record data from Namibia and Senegal, we study population migration and its link to short term movement. We compare the short term mobility estimates extracted from call detail records to census data in the two countries and find a strong annual relationship, as well as distinct daily patterns in the relationship between long and short term movement. The relationship is strongest for holidays, and we find it to be consistent both across countries as well as across multiple years. In particular, we observe periods of increased travel on migration routes around holidays, with net short term travel in the opposite direction of the direction of migration before the holiday and net travel in the same direction after. Using the Namibia data set, which spans several years, we investigate the link between short term mobility and long term relocation on an individual level, allowing us to gain insights into the mechanisms of interaction of short and long term mobility. We find that it is common for individuals to both visit the place they will migrate to prior to migration and also visit their place of origin after migrating. Additionally, distance between the origin and destination of a migration has a significant influence on the probability of a short term trip associated with a long term move.

The Senegal dataset provides information on the full network of users, which we use to study the relationship between the location of network contacts and probability of traveling to those locations, investigating the importance of social contacts for mobility. We find that while the majority of social contacts in different regions can be explained by long term migration patterns between regions, which in turn are linked to short term movement patterns, social contacts can explain some of the additional short term movement not captured by the long term migration. We also find non-linear relationships between the probability of visiting a region and the number and strength of contacts, as well as between the duration of a visit and social contacts. These results can help inform evidence-based policies that target some of the negative externalities of short term population movement such as spread of infectious disease, increased congestion, and inadequate infrastructure.

*We would herewith like to give credit to MTC, whose assistance and support made the realization of this Study/Project possible. Anonymous mobile phone data have been made available by Orange and Sonatel within the framework of the D4D-Senegal challenge, and we want to thank Sonatel and Orange for their continued support.

Correspondence: sveta.milusheva@gmail.com, elisabeth.zu.es@gmail.com

Key words: Migration, Big data, Call detail Records

JEL Classification: O15, R23, C8, C88, C89

Original version: English

Accepted: July 2017

TABLE OF CONTENTS

1. Introduction

Migration is observed all over the world, both in terms of international migration between countries as well as internal migration within a country. Urbanization is continuing across Africa, although almost two thirds of people still live in rural areas (Henderson, Storeygard, and Deichmann 2014). Linked to migration and urbanization are short-term movements, which in contrast to permanent relocation on the scale of years are temporary relocations on shorter time scales such as days, weeks or months. These short term movements can be the result of economic activity (seasonal labour) (Atamanov and Berg 2012, Etzold et al. 2014, Rain 1999), disaster or conict related displacement (Bengtsson et al. 2011, Wilson et al. 2016, Poncelet et al. 2010, Doocy et al. 2015) or recreational travel or pilgrimages during holidays (Ahmed, Arabi, and Memish 2006, Rinschede 1992, Mokashi 1987). Permanent relocation can generate short term recreational travel to visit friends and family. Conversely, having visited friends or relatives that have moved away might inform individuals about economic opportunities away from their current place of residence, in which case short term movements could increase the likelihood of long term migration. Additionally, economic opportunities could lead to seasonal, temporary migration of some members of a household, which later causes permanent change of residence for the whole household. Intuitively, it seems migration and short-term mobility as well as the social networks of migrating individuals might influence each other.

Short term internal population movements have implications for a range of settings. They have been shown to have an impact on spread of diseases (Oster 2012, Prothero 1977, Balcan et al. 2009, Huang and Tatem 2013, Tatem and Smith 2010, Wesolowski et al. 2015b, Wesolowski et al. 2015a, Stuckler et al. 2011, Pison et al. 1993), one example being the relationship between mobility and malaria (Lynch et al. 2015, Osorio, Todd, and Bradley 2004, Siri et al. 2010, Yukich et al. 2013, Alkhalife 2003, Littrell et al. 2013). Increases in short term mobility will likely increase pollution, congestion in urban areas and have potential implications for economic activity. In order for policy makers to address some of these potentially negative externalities of population movement, it is necessary to understand short term mobility patterns as well as some of the factors that inuence them in order to be able to predict what short term movement uctuations will look like and to incorporate this knowledge into existing policies. Yet, the drivers of short term mobility have not been studied due to the lack of available data on ne spatial and temporal scales.

Migration is often measured in the context of a census as change of permanent residence on a given spatial scale, and thus migration data is available for many countries. Short-term movements on the other hand are harder to capture and the extent of seasonal mobility is often unknown or only studied in the context of travel surveys for particular populations. Mobile phone data provide a powerful source to measure these shorter term mobility patterns for whole countries (Wesolowski et al. 2012, Tatem et al. 2014, Wesolowski et al. 2015b). Call detail records (CDRs), collected by cell phone companies for billing purposes, provide geo-location and timing for calls and text messages, allowing researchers to observe changes in location for individuals within a country. Where CDR data are made available to researchers in an anonymised form, it is possible to determine country-wide mobility patterns with ne temporal and spatial resolution. These data are frequently used to estimate population mobility over short periods, however, it has been shown that they can also be used to estimate migration, where census based migration estimates are not available (Wesolowski et al. 2013). Here, we study the ip side of this question, using CDR data to study the relationship between short term population movement and long term migration and investigating the motivating factors for this relationship in order to predict short term movement when estimates are not available. Understanding the relationship between the two will allow policy makers to better plan for potential increases in short term movement in order to mitigate negative impacts of these increases in movement. In addition, for countries where cell phone data are not available, understanding the link between long and short term movement can inform policy makers in which cases it is valid to use available census or survey data on long term migration to predict short term movements.

In the context of this study, we focus on migration patterns in two countries Namibia and Senegal. Internal migration in Namibia is common, but varies by demographics, migration being slightly more likely for men and most common for the age group 20-34 (Namibia Statistics Agency 2015). In 2004, 27 percent of the Senegalese population was recorded as an internal migrant (Fall, Carretero, and Sarr 2010). Much of the migration is rural to urban, due to a reduction in the agricultural production index, though there is also circular migration, pastoral movements, and temporary rural to rural or urban to urban migration (Fall, Carretero, and Sarr 2010, Goldsmith, Gunjal, and Ndarishikanye 2004, Linares 2003, Adriansen 2008, Herrera and Sahn 2013). A study focused specically on migrants to Dakar finds that 87% of male migrants and 81% of female migrants visit their home locations, primarily for holidays, family ceremonies and religious festivals (Fall 1998). Similarly, for Namibia, rural to urban migration makes up the largest proportion of internal migration (Namibia Statistics Agency 2015) and the percentage of the population living in urban areas increased from 33% in 2001 to 43% in 2011. Of the people living in regions that contain urban centers, 40% are migrants from other regions. During major holidays, large outows out of the capital Windhoek can be observed, with most individuals traveling to the north of the country, most likely to visit relatives (Erbach-Schoenberg et al. 2016).

Working with call detail record data from Namibia and Senegal, we study the link between migration and short term movement through three dierent analyses. We start out by exploring the relationship between short term mobility measured using CDR data and long term migration measured using census data in the two countries. Using the Namibia data set, which spans several years, we then investigate the link between short term mobility and long term relocation

on an individual level, allowing us to gain insights into the mechanisms of interaction of short and long term mobility. The Senegal dataset provides information on the full network of users, which allows us to study the relationship between the location of network contacts and probability of traveling to those locations, investigating the importance of social contacts for mobility. By using data from two dierent countries, we have the opportunity to test how well models relating short and long term mobility translate from one country to another, which will be important in extrapolating our research to other country contexts. Finally, in the last section we discuss the implications and potential uses of the results presented in this paper.

2. Data

2.1 Census data

Data on migration were provided by the Agence Nationale de la Statistique et de la Demographie (ANSD) and the Namibia Statistics Agency for Senegal and Namibia respectively. For Namibia, the migration data used were collected in the context of the 2011 census and the data for Senegal come from a 10 pecent sample of the 2013 census data. For both countries we use a variable that denes migration as a change in usual residence from the previous year compared to the time of the census.[1]

The specic question asked in the census for Namibia is "where did NAME usually live since September 2010?" (Namibia Statistics Agency 2015). The answer from this question is compared with the answer to the question "where does NAME usually live?" (referring to at the time of the census done in 2011) and migration is inferred if the location has changed between 2010 and 2011.[2] The data for this question is only available at the region level for the 13 regions of Namibia. For Senegal, the specic question asked in the census was "where did you reside one year ago." If the location specied is different from the location the respondent states to be the current residence, then a long term migration is inferred between the two locations. While the Senegal census data is available at the commune level, which is a higher granularity, we aggregate the data up to the region level for the 14 regions of Senegal in order to work at a spatial level comparable to Namibia. This helps facilitate the comparison of results between the two countries.

Out of the 12.9 million individuals in the census that responded to the question of where they were located one year ago with a valid response that could be aggregated at the region level, there were 344,345 individuals that had a long term migration in 2013.[3] In Namibia, a migration from one region to another was recorded for 40,867 individuals (Namibia Statistics Agency 2015). The two countries are very similar in that 45% of all long term migrations occur between neighboring regions in both countries.

2.2 Mobile phone data

Call detail records (CDRs) are collected routinely by mobile phone providers for billing purposes. When a subscriber makes or receives a call or text, a record is created by the operator, containing a unique ID, a timestamp for the time of the communication as well as the location of the closest mobile phone mast through which the communication is routed. In some cases, network data is also available. Network data provides the IDs of both the caller and the receiver, thus allowing for the linking of individuals into a communication network. The data are anonymised before they are shared with researchers in order to preserve the privacy of users.

For Senegal, we use CDR data provided by Sonatel and Orange in the context of the Data for Development Challenge. The data consist of call and text records for Senegal between January 1, 2013 and December 31, 2013 for all of Sonatel's user base. In 2013, Sonatel had slightly over 9.5 million unique phone numbers in its network, representing a large portion of the 13.8 million Senegalese population. Out of the three telecom providers in Senegal, Orange Telecom had between 56 and 62 percent of the cell phone market in 2013 (ARTP 2013), though according to the *Listening to Senegal* survey done in 2014, 83% of those surveyed with a cell phone cite Sonatel as their primary provider and 89% cite having a Sonatel SIM card. The data contains information on all calls and texts made or received by an individual, namely the time and date of the communication and the location of the closest cell phone tower, which makes it possible to measure movement as people change locations based on making calls/texts from dierent towers.

---

[1] The Namibia data also includes information on lifelong migration, but this information is not used in the current study.

[2] The Namibia census agency notes that the wording of the question concerning the usual location lived since September 2010 could have prompted some respondents to provide their current residence rather than the former residence even if they lived somewhere else in 2010, because "since" could potentially include the current time. If that were to systematically happen, then long term migration would be underestimated with the census data. Implications of this for our analyses are considered in the discussion analysis.

[3] The 10% sample in the data was weighted in order to be representative of the full population. Individuals that responded that they were abroad one year ago are not included, only internal migrants who provided alocation that could be assigned a region within Senegal are included.

Each SIM card has an anonymised identication number associated with it, which allows linking together all communications and locations for a particular card over time. In addition, the data contains not only the anonymous ID of the SIM that makes or receives a call/text, but also the anonymous ID of the SIM that is receiving or making that call/text. With this information, the mobile phone data can be used to create both a contact location network of the locations where an individual has contacts that he or she has been in touch with over phone, as well as a locational network of the places the individual has physically visited.

For Namibia, CDR data were provided by MTC with the purpose of estimating internal mobility and its impact on malaria transmission. The data span a period of 3.5 years, from October 2010 to May 2014 and include all subscribers to MTC, corresponding to 72 Billion entries for a user base of 4.5 Million unique users. MTC's market share is high at 76% in 2012, and 95% of the population are covered by the network (http://www.mtc.com.na/coverage). Each data entry corresponds to a communication, either a call or text, made or received. Each entry contains a unique identier, time and date for the communication as well as the tower the communication was routed through. There is no network data available for Namibia. While both CDR data sets have a ner spatial resolution than region level, as the locations in the data are on the level of mobile phone towers, we here aggregate the data to the region level in order to match the census data and to further protect the privacy of users.

2.3 Measuring short and long term moves from CDR data

To dene short term movement, we rst calculate a daily location for each user at the region level. This daily location is dened based on the last call or text of the day, with missing days assigned the location of the closest day with data. In Namibia there are 13 regions, while Senegal has 14 regions, though two of the regions are excluded in the analyses as they experience large pilgrimages during the year. Pilgrimage regions are excluded since the large population movements observed are representative of a dierent driver of movement that is unrelated to the relationship between short and long term migration patterns. Results including the pilgrimage regions are included and described in detail in Appendix B. On the individual level, we dene a short term move as a change in a user's location from one day to the next. Days without calls and therefore without a dened location are interpolated using the user's location from the closest day with data.[4] We aggregated these individual level data to create daily region level mobility networks. In Senegal, there are a total of around 63.5 million short term trips that take place in 2013. On average that is around 174,469 trips per day.[5] In Namibia, the CDR data capture 64 million short trips over the whole period and 20 million short trips in 2013, which is an average of 55,245 trips per day. While the two countries were very similar in the percentage of long term migrations between neighboring regions, they dier some when it comes to short term movement. For Senegal, 63% of short term trips occur between neighboring regions, while for Namibia around 80% of trips are between neighboring regions.

There are some limitations to the data in calculating movement. Since we dene a short term move as a change in location from one day to the next, if someone lives on the border between two regions and commutes for work or other reasons, we might pick up this movement, which represents a dierent mechanism from the one studied in this paper. To minimize this issue we use the last call or text of the day to dene the daily location in order to increase the probability of capturing a user at their place of residence. There is still the possibility that an individual might not have any evening calls, meaning that the last call of the day will be early in the day but as we can see in Figure 1, over 80% of calls/texts take place between 6pm and midnight in Namibia. Therefore, we partially mitigate the issue of capturing commuting trips by using the last call of the day to dene the daily location. Additionally, while using a very low granularity and dening movement at the region level does not take advantage of the high granularity available with mobile phone data, it helps in decreasing the level of commuting that is picked up with our denition of movement.

An additional limitation of the data is that we can only observe a user's location for days when he or she has a phone communication. For days with no data, we ll in the missing days based on the location of the closest days with data. For individuals that are missing many days of data, this potentially means interpolating a large number of days, and it is possible the person is located somewhere else, which we cannot capture. Figure 2 shows a histogram of the percentage of SIMs in the data with a given number of days of activity in 2013 in Senegal and in Namibia. We can see that for both countries, there are a number of individuals that have less than 30 days of data in 2013, but there are also many that have data for every single day, or for most days in the year. The data in Namibia are a bit more polarized with more than 50% of individuals falling in the lowest or highest category of usage and fewer in the middle, while for Senegal the distribution is more spread out[6]. Nevertheless, the countries are very similar on average, with individuals having 155 days of data in Senegal and 151 days of data in Namibia.

---

[4] Filling in missing days in this way means that if there is a period of time where the person does not use the SIM card and right before and right after this period the person is seen in dierent locations, then it is assumed the change in location occurs halfway between the two days when data is available.

[5] Note that we do not have movement for January 1, 2013 for Senegal, so the number of short term movesis based on 364 days. This is due to the fact that to calculate movement on a given date we need to have a location for the day before.

[6] Note that we only ll in undened locations between the rst day of activity and last day of activity for a user. This way users who are only active for a very short period will not be included in the data outside of that period.

Given the length of the Namibia data set, we can also observe long term migrations in the CDR data. While there are many ways of dening long term migration, we use the denition of an individual living in one region for at least 6 consecutive months followed by at least six consecutive months in a dierent region. This denition was chosen to match the definition of place of residence in the census (Namibia Statistics Agency 2015) and is similar to the methodology used by Blumenstock (2012), though we use 6 months rather than 12 months as the cuto.[7]

2.4 Example of Movement Patterns in Senegal and Namibia

Figure 3 looks at movement into the regions where the capitals are located in both countries in order to exemplify what the movement looks like. For Senegal, the capital Dakar is located in the Dakar region, while for Namibia the capital Windhoek is located in the region Khomas. The two graphs on the left show movement based on the percent of long term migrants from the census data entering from each region into the capital region. In both countries, we see a relatively even distribution of individuals migrating into the capital, with around the same number of people entering from most regions. In Namibia, there are two regions from which there are very few migrants into the capital and one from which there are more migrants than the rest, but generally, it's a relatively even distribution. In Senegal, we see similarly a relatively even distribution in the western regions of the country that are closer to Dakar, with magnitudes similar to those in Namibia. There are more regions with fewer migrants though in Senegal, with very low numbers coming from the ve regions to the west. This is in line with the geography of the countries since having the capital in the middle of the country, like in Namibia, means that it is relatively easier to get there from anywhere in the country. In Senegal, the capital is the western most tip of the country, and therefore it means those living in the eastern part of the country have to travel much further, which limits the number of migrants.

The two graphs on the right based on the CDR data show the percent of people entering from each of the other regions. When we look at the short term moves into the capitals, we see that distance plays an even bigger role in where the majority of short term moves come from. We see that for Namibia, over 60% of short term visits come from regions that border the capital. In Senegal, this is even starker since there is only one region that borders Dakar (Thies), so we see almost half of all short term movement coming from that region. This could represent several dierent factors. First, it is easiest to travel from a neighboring region, so we might see the most short term movement coming from the one neighboring region. Additionally, it is possible that there might be more commuting from the neighboring regions, and while this is spread out in Namibia across multiple neighboring regions, in Senegal it is concentrated since there is only one.[8] As people travel into Dakar from other regions, they all have to pass through Thies in order to get to Dakar, so it is possible some people spend a night in this region if they are coming from very far away and make a phone call while there; therefore, due to the denition of movement, we are picking up their movement from Thies to Dakar and not from the farther region that the person is originally coming from.

2.5 Short term movements and census migration

We rst look at the relationship between short term movement and migration on an annual basis, summing all short term moves going from region x to region y in a year and summing all long term migration trips from region x to region y in the census data. We use 2013 for both Senegal and Namibia in order to be consistent with the timing that we are using to compare the two countries. The observations are directional, with separate observations for movement from x to y and for movement from y to x. We run regressions of total short term movement on total long term migration, clustering the data at both the origin and destination levels using two-way clustering.

To see how the relationship between short and long term movement shifts from day to day, we can look at the relationship between census data and short term mobility on a daily level. We run regressions of daily short term movement on long term migration in the census data for each day of 2013 for both Namibia and Senegal. To investigate travel behavior around holidays further, we look at the net ows on each route to take into account the direction of ows. Around holidays there are two possibilities, with day to day net short term movement being either in the same direction as the net long term movement or in the opposite direction. Movement in the opposite direction might correspond to individuals

that have migrated traveling back to their place of origin or having family returning after a visit. Movement in the same direction could correspond to family from their place of origin traveling to see them or the migrant returning after a visit to the place of origin. We explore this by regressing the net daily movement between locations on the net long term migration.

---

[7] A region is assigned to each month based on the location where the most days were spent in that month, where a day is assigned a location according to the last call or text of the day and missing days are assigned to the location of the closest day with a call or text.

[8] Though we have already discussed strategies that we have used in order to minimize the amount of commuting that we pick up in the data.

2.6 Drivers of the relationship between short term movements and census migration

In addition to looking at how short and long term movement are related, we also provide evidence for some of the motivating factors behind this relationship. While a number of factors may exist, we focus on investigating three processes:

1. Individuals might travel to a new location and afterwards migrate to it. This could be because they might assess the new location before moving by visiting someone that has already migrated or investigating economic opportunities in person.[9]
2. Individuals travel to their location of origin after migration, most likely to visit contacts, such as friends and family.
3. Individuals that are not necessarily migrants travel along routes with high migration because they want to visit social contacts that have chosen to migrate

In order to explore processes one and two, we can use the data from Namibia which allows us to follow SIMs over a long period of time. Using this data, we can observe long term migrations for individuals in the data and also measure any short term moves that these individuals might make prior and post migration. If an individual has migrated from location x to location y, we count a short term move prior to migration as the individual having traveled from x to y as well as from y to x prior to the long term migration. Similarly, a short term move after the migration is counted as an individual traveling from y to x and from x to y after the long term migration.[10]

We study all of the long term moves that take place in the Namibia dataset and identify all short term moves taken by the same individual between the origin and destinations prior to the long term move and after the long term move. We break down visits before and visits after by the origin and destinations of individuals, since the likelihood of a visit might depend on the specic locations. To more explicitly explore the inuence of location, and specically the inuence of distance on the likelihood of a visit before or a visit after, we estimate two logistic models. In the rst, we have a dummy for whether or not the migrant made a visit before migrating on the left hand side, and we have dummies for the region of origin and dummies for the region of destination, along with controls for month and year fixed effects. The same model is also run including distance. These two are analogously estimated for visits after.

We then turn to explore process three. First, we compare the annual relationships between short term movement, long term movement, and social contacts. For these comparisons we use total ows instead of directional ows, summing the number of people going from x to y and from y to x. For social contacts, we sum the total number of social contacts between regions x and y.[11] We then use regression analysis to look at how short term movement, long term movement and social networks are related to each other.

On the individual level, we can study how having a social contact in a region inuences individuals' likelihood of visiting the region. We rst estimate a conditional logit model with a dummy on the left hand side for having visited a particular region. On the right hand side we have the number of people in that region that the individual interacts with as well as the total number of calls/texts that the person has with people in that region, the latter measuring the strength of the relationship[12]. Additionally, we include a dummy for whether the person has contact with at least one person in the region. The reason for this is that the relationship between contacts and visits might not be linear, with the rst contact being more important in making the decision to visit or not and each additional contact inuencing the decision to a lesser degree, leading to diminishing returns for each additional contact. The dierence in the likelihood of visiting when a person has one contact versus two in a location is potentially much larger than the dierence in the likelihood of visiting when a person has 10 versus 11 contacts in a location. We test this by also running a specication where we include a squared term for number of contacts as well as a squared term for number of calls/texts.

3. Results

3.1 Relationship between Migration and Short Term Movement

For Senegal we observe around 127 short term moves between regions x and y in 2013 for each long term migrant going from region x to region y recorded in the census. In Namibia, this number is higher, with around 506 short term moves per migrant in the census. In both cases, the R^2 is between 0.3 and 0.4, with long term migration patterns being able to explain around 30-40% of variation in short term movements happening on an annual basis. We find a strong positive

---

[9] Based on this analysis, it will not be possible to disentangle whether the motivation for a prior visit iseconomic opportunity or visiting a contact, though in many cases, it is probably a combination of the two.

[10] This denition is relatively restrictive as indirect travel, for example if an individual does not travel between x and y directly but instead stops at an intermediary location for a day or more, will not be counted.

[11] A social contact between x and y is counted as a person whose home location is x making or receiving calls from a person whose home location is y. The home location is based on the location where the most number of days are spent in 2013.

[12] Using the assumption commonly made that individuals are more likely to call and text individuals with whom they have tronger ties.

relationship between census migration numbers and short term movement measured through CDRs for both countries. Table 1 shows the results from the annual regression.

On a daily level, the relationship between movement and census migration remains relatively consistent for both countries, but with signicant jumps occurring at distinct points (Figure 4). On average, there are around 0.35 daily moves per long term migration in Senegal and around 1.38 daily moves per long term move in Namibia. The spikes in the relationship between short and long term movement in both countries are related to the major holidays, marked as vertical red lines. This implies that people are more likely to make trips between migrant places of origin and destination around holidays.

Next, we look at net movement to investigate the directionality of these increases in travel on migration routes during holiday periods. We nd that before a holiday the coecient between net short term movement and net long term movement is negative while after a holiday we observe a positive relationship (Figure 5). This signies net short term movement in the opposite direction of the long term migration patterns before a major holiday, whereas after the holidays the increased travel is in the same direction as the long term move.[13]

Using the daily coecients of the regression on the net movement data to describe the relationship between short and long term movement not only tells us about the direction of the short term movement, but additionally highlights the timing of the movements. We observe dierences in the length of the time period during which the holiday associated movements occur. For example, looking at Easter in Namibia (the rst red vertical line), there is one distinct spike, meaning most people travel on the same day, whereas for Christmas in Namibia, travel is more spread out. Visually, we can see this spread of the travel volume over a longer period by the coecient decreasing more steadily, from the beginning of December for the travel before Christmas and similarly gradually increasing after Christmas. In contrast, for Easter we observe a sudden spike in the coecient just before and after the holiday. We also see a dierence in travel synchronization when comparing travel before and after Christmas, with the return travel period being shorter. This highlights that individuals tend to travel back at a similar time, whereas travel before Christmas is less synchronized. Finally, in addition to the holiday patterns, we see very distinct weekly patterns of movement. These are especially pronounced in Namibia, where it seems there is a steady ow of individuals back and forth every week between places of origin and destination.

Several years of data are available for Namibia making it possible to study whether the observed patterns of holiday related increases in travel as well as dierences in travel timing and synchronization for dierent holidays are consistent across dierent years. Figure 6 shows daily regression results of net short term movement on net long term movement for 2011, 2012, and 2013. We nd that patterns of short term movement around holidays remain consistent from year to year when taking into account the slight change of timing for Easter.

3.2 Relationship Between Short and Long Term Movement for Migrants

In the previous subsection, we saw a strong relationship between long term movement in the census data and short term movement in the CDR data. We now show results for some processes that are likely to inuence this relationship.

Looking at the short term movements made by individuals for which we observe a long term migration from the 3 years of mobile phone data available for Namibia, we find that about 26% of individuals migrate having both never visited the place of destination prior to moving and never visit their place of origin after migration.[14] The most common of the four possible cases is for individuals to make short term trips on the migration route both before and after migration, with 38% of individuals falling into this category.[15] Of individuals that only visited either before or after, visiting only after the move is more common with 21% of migrants choosing to do this, versus 15% of migrants visit only before the move. Therefore, we find that while short term movement after a long term migration slightly dominates, much of the relationship between short and long term movement can also be explained by short term movements prior to migration.

Figure 7 shows the percentage of migrants that do not make any short term visits, neither before nor after the move, by region of origin as well as destination. This highlights that most remote regions, such as 1 (in the most eastern part of the country) or 4 (southern most part of Namibia) are the places most likely to never see a visit (and least likely to see a visit both before and after as shown in Appendix Table A2). In contrast, central regions with several nearby regions, such as regions 3 and 11, have the lowest probability for no short term visits associated with a long term move.

We explicitly explore distance as an underlying factor for this nding, as intuitively it is much easier to visit between two regions that neighbor each other, compared to visiting between regions on opposite sides of the country. Remote regions such as 1 and 4 have fewer neighbors thus making it necessary to travel longer distances to any other regions

[13] We should note that at an annual level, there is no signicant relationship between net short term movement and net long term movement, see appendix table A1. This is intuitive since the daily correlations go back and forth between negative and positive and average out to zero.

[14] An example of this type of behavior would be a wife whose husband has migrated prior, who then migrates to follow her husband with their children and no longer goes back to the place of origin since the family is now all together in the new destination.

[15] In the previous scenario, the husband who first migrated might have visited rst to ensure that he can find work, migrates and then visits home to see his wife and children until they are able to migrate as well to the new destination.

in the country. Upon ranking pairs of regions based on number of visits between them, the top five with the lowest number of visits before and after each have region 4 as one of the regions in the pair. In contrast, the six pairs with the highest percentage of visits both before and after are 11 to 12, 6 to 3, and 11 to 10 (and their reciprocals), which are all neighboring regions.

Table 2 shows results in the form of odds ratios for the probability of visiting a particular region based on a logit model, controlling for region of origin, region of destination, month and year fixed effects. In the rst two columns, we have the results for visiting a particularfregion that is the destination prior to migration. In columns 3 and 4, we have the results for visiting a particular region that is the origin after migration. In all cases, the region of comparison is region 13. The table shows that controlling for distance between pairs of regions has very signicant eects on the likelihood of visiting. We see that when distance is not controlled for, it looks like there is a signicant dierence between visiting region 13 and visiting regions 2 and 4. Once distance is controlled for, the probability of visiting region 13 is actually not signicantly dierent from visiting those two regions. The results also show that for some regions such as 1 and 5, the probability of visiting is much lower than region 13 until distance is controlled for, and then we see that once we take account of distance, people are more likely to visit those regions as compared to region 13. Many of these results are also present when looking at visits to the origin after a migration. Finally, the log likelihood on distance is less than one, signifying that the larger the distance between origin and destination, the lower the likelihood of a visit before and similarly for a visit after.

3.3 Relationship Between Social Networks and Visiting Patterns

Table 3 shows regressions exploring the relationship between short term movement, long term migration and social contacts. The first column shows the results from a regression of total short term movements in 2013 on total long term migration. These results are in line with the results presented in Column 1 of Table 1, though a larger R^2 indicates that it is possible to explain more of the variation in short term movement with the long term movements if we do not take into account directionality of movements.[16]

The second column of Table 3 shows total short term movements regressed on number of social contacts between regions x and y. The relationship between contacts and short term movement is positive and highly signicant, with a slightly higher R^2 compared to regressing short term movement on long term migration, implying that social contacts potentially provide slightly more information in helping to explain short term movements.

The hypothesis that social contacts might provide additional information as compared to long term migration patterns is explored in Column 4, where we show the results from a regression of total contacts on long term movements. The coecient on long term movement is large and highly signicant, and more importantly, there is a very high R^2 of 0.68. This means that a very large portion of the variation in number of social contacts can be explained by long term migration. Nevertheless, there is still around 30% of the variation that cannot be explained by long term migration and instead is based on other factors that influence the formation of social networks. These non-migrant social networks can also influence the likelihood of making a short term trip.

Finally, in Column 3, short term movement is regressed on both long term migration and number of social contacts. Including both factors in the regression, they each become smaller in terms of the size of the coecient, but both remain signicant at the 0.01 level. Additionally, we see that the R^2 increases to be above 0.5, implying that including both long term migration and social network information provides additional information for explaining the variation in short term movement. This analysis shows that while a large portion of the social network between dierent regions can be explained by long term migration patterns, the social network provides additional information that can help in determining short term movement patterns because people have some social contacts in locations not determined by long term migration patterns.

The first part of this analysis focuses on social networks and movement aggregated at the region level. We now turn to results from the individual level analyses. Columns 1 and 2 of Table 4 show the results from the conditional logit models as odds ratios. In Column 1, we see that indeed both the number of contacts and the strength of contacts measured by number of calls and texts are signicant in explaining the likelihood of visiting a region. They are also greater than 1, indicating that each additional call or text and contact lead to a higher likelihood of visiting. The results also show that there is an extra eect of having at least one contact, with a large coecient of 5.655 on the dummy for having any contacts.

This shows that most important for a visit is to have at least one contact, with each additional contact increasing the likelihood a little bit. Including the squared terms in Column 2 shows that, as hypothesized, there are diminishing returns to contacts and calls since the coecients are less than 1. This implies that as individuals have more and more contacts and make more calls or texts, the impact of the increase on the likelihood of visiting decreases.

The results of the duration analysis are shown in Columns 3 and 4 of Table 4. Each additional call or text to a place and each individual contact contribute to the person remaining for longer (0.03 of a day for each call and 0.36 for each social

[16] This result is reasonable since in the directional regression we regress short term movement from x to y on long term movement from x to y, but realistically, the long term move of an individual from x to y might cause both short term movement from x to y and from y to x, which does not get captured in the directional regression, but does get captured in the regression looking at total movements.

contact). Interestingly, when it comes to duration, there does not seem to be the large discrete jump observed for the probability of visiting. Instead, the coecient on the dummy for at least one contact is not signicantly dierent from the coecient on total social contacts, so the additional effect of having at least one contact is equivalent to the eect of having one extra contact. However, when including the square terms as shown in Column 4, the coecient on the dummy becomes negative, so once the diminishing returns of extra contacts are taken into account, there is no additional positive effect on the length of a stay of having at least one contact. Together, all of the results shown in this table point to a very strong non-linear relationship between the location of social contacts and both likelihood of visiting and the length of the visit.

4. Discussion

In this paper, we study the relationship between short term movement and long term migration patterns and dive into the factors that inuence this relationship. In the rst part of the analysis, we nd a strong positive relationship between short term movement and long term migration which conrms previous descriptive research showing that people tend to travel back and forth between the place they originate from and their destination (e.g. Fall 1998). A factor not previously investigated is the consistency of the results across multiple countries and we nd the travel patterns to be remarkably similar between the two countries studied, even though Senegal and Namibia dier on a number of factors, such as population size, density and GDP.[17] Results are not only consistent at the annual level, but also for daily regressions. The pattern of increased movement on migration routes around holidays is very consistent, despite the two countries having different holiday patterns due to different majority religions in the two countries (Senegal is majority Muslim while Namibia is majority Christian).

The first regressions we show in Table 1 have an R^2 of between 0.3 and 0.4 for the two countries. With long term migration in the census data as the only variable, we are able to explain around a third of the variation in short term movement. While this demonstrates the importance of long term migration patterns in determining short term movement, as discussed in the introduction, there are many other factors that aect short term movement in addition to long term migration. People might travel for tourism, for business or seasonal work, or for pilgrimages. These all represent dierent mechanisms that will impact the amount of short term movement. Appendix B provides a rst exploration of the impact of pilgrimages on short term movement. While the aim of the appendix is to demonstrate why the two regions with large pilgrimages in Senegal are excluded from the analyses in this paper, it also provides some insight into the impact of large events like pilgrimages on short term movement patterns. Additional mechanisms of short term movement can be explored in future work.

Studying net movements allows us to investigate which pattern of movement is more common: individuals that have migrated traveling back to their place of origin, or contacts (such as family and friends) from their place of origin traveling to visit them. The net movement analysis does not refute the possibility that both types of travel are happening, but instead shows which type of movement is more common. The results show travel in the opposite direction of migration before the holiday and in the same direction after. This is in line with the anecdotal evidence that individuals that have migrated are likely to visit their home location for holidays and return to their new permanent place of residence after the holiday (Fall 1998).

Some additional patterns we observe are that travel is less synchronized for longer holidays when individuals have more time o (for Senegal the two Eids and for Namibia, Christmas), with individuals not necessarily traveling on the same date. This is especially the case for Namibia around Christmas where we see an extended period of movement in the opposite direction of long term movement before the holiday. In addition to increased travel on migration routes around Christmas we also see a weekly pattern, with people moving back and forth between the place of origin and destination on a weekly basis. This signies that some individuals that have migrated long term choose to travel to their place of origin every weekend. This pattern is present in both countries, but is more marked in Namibia.

We consistently see Namibia having higher levels of short term movement per long term migrant, and there are several things that could explain this. First, as mentioned in the data section, due to the wording of the question in Namibia, it is possible that respondents misinterpreted the question leading to an underestimate of long term migration. This underestimate would mean that when running regressions, the coecient on long term migration will be larger.

Additionally, in both Senegal and Namibia we are only able to look at short term movement for those individuals represented in the data. If the proportion of individuals represented is smaller in Senegal, then there will be a smaller number of short term moves captured, which would again lead to the dierence in magnitude of the coecient on long term migration we see between the two countries. In Senegal, there are around 9.5 million unique SIMs in 2013 and the population was around 13.8, leading to a ratio of 0.69. In Namibia, on the other hand, there are around 2.9 million SIMs in 2013 and the population was around 2.1 million, leading to a much larger ratio of 1.39 SIMs per person. This likely means that a larger proportion of the population is covered by the data in Namibia, however measuring this eect from

[17] The Namibia population was around 2.5 million, population density was around 3 persons per sq.km., and GDP per capita was $4,673.60 in 2015. In Senegal, the population was around 15.1 million, population density was 79 people per sq.km, and GDP per capita was $899.60 in 2015 according to The World Bank (2017).

the count of unique SIMs alone is not possible due to factors inuencing the SIM to population ratio, such as multiple SIM ownership. Individuals frequently own more than one SIM card to make use of promotions or to transfer credit to other users (Stork 2011).[18]

Nevertheless, despite some dierences in the magnitude of short term moves per long term migrant, we generally see very similar patterns between Senegal and Namibia. We also finnd that the pattern is very consistent from year to year, as highlighted by comparing results from different years for Namibia. These two ndings are important from a policy maker's point of view. While CDR data are an excellent data source for measuring the timing and magnitude of short term movement, mobile phone data are not available to researchers and policy makers in every country context. Additionally, where this data is available, it generally only spans a limited time period. The analyses here show the possibility for policy makers to use information on short term movement from periods for which mobile phone data is available and to combine them with knowledge of main holidays in order to foresee the expected movement patterns that will be critical for many settings, ranging from infrastructure planning to controlling infectious diseases. Additionally, the analyses demonstrate the possibility of using the patterns seen in countries where CDR data is available to extrapolate to other country contexts where the CDR data is not available. In those contexts, if census data is available, it can be combined with information on public holidays to predict fluctuations in short term movement, and this information can be similarly applied to different policy questions.

For example, our first analysis looking at the relationship between short term movement in the CDR data on a daily basis and long term migration from the census data shows that there are very consistent patterns in the relationship, with short weekly spikes representing the weekly movement patterns and large spikes around holidays. Specically, we are able to identify the periods of movement, which vary with each holiday, but are consistent across years. These periods are times at which large numbers of individuals travel on particular routes, linked to the long term migration routes. In studying issues such as spread of malaria due to travel in areas where the disease is at an elimination stage, this type of information can inform interventions such as mobile surveillance points on high trac routes or increased malaria screening after individuals return home. To exemplify how this information can be used, we more formally identify these periods of large uxes of movement between places of origin and destination by highlighting days for which the absolute value of the coecient is larger than two standard deviations of the absolute daily coecient for Senegal and larger than three standard deviations for Namibia.[19]

Figure 8 shows the graphs of the coecients from the daily regressions of net short term movement on net migration (as in Figure 5), but we now highlight the regions of time when the coecient is larger than two or three standard deviations. This marks the timing of travel, which can provide additional information to guide evidence-based strategies to fight malaria. While it is known that there is a lot of travel around holidays, this analysis shows the duration of the period of increased travel, which might be more dicult to determine from alternative data sources. We can see that for some holidays, the travel happens extremely quickly over the span of a weekend, while for other holidays, it is much longer. This can help policy makers in better targeting resources in order to facilitate decreasing the malariaburden at a lower cost or plan for potential congestion along highly traveled routes.

Looking at more specic aspects of the relationship between short and long term movement patterns provides additional insights that can be exploited by policy makers in crafting policies to address issues related to short term population movement. As might be expected, we saw that the farther origin and destination are, the less likely people are to visit both before and after a long term migration. This is intuitive since the farther the distance, the more dicult it becomes to visit in terms of monetary and time cost, and we are able to quantitatively conrm this. We also saw that social contacts play an important role in determining the location of short term visits and that contacts in other regions are not solely determined by long term migration patterns.

By modeling the relationship between short and long term movement and understanding some of the factors underlying this relationship, we can then consider how changes in long term migration might lead to new patterns of short term movement. This has implications for policies related to population movement, since while previous studies have used mobile phone data to classify sources and sinks of malaria, they do not necessarily provide information on how those sources and sinks might change if the movement patterns shift. Therefore, again, with only limited access to mobile phone data, it is possible to use this data more eectively in order to craft targeted policies that can account for changes in the short term movement even when real-time access to the mobile phone data to measure short term movement is not available.

Finally, in thinking about population movement and applications such as the spread of infectious disease or access to adequate infrastructure, not only does the movement matter, but the length of time that an individual spends in a location can matter as well. For the application of spread of malaria, the longer the person remains in an area, the larger the probability that the person might become infected with malaria (if they are in a high malaria area) and the higher the probability that they might infect mosquitoes that spread the disease if an infected individual enters a low malaria area.

---

[18] Transfer of credit is only possible between prepaid SIM cards.

[19] We use a larger cuto for Namibia due to the larger weekly patterns of movement

The combination of the relationship between long term migration and social contacts and the relationship between social contacts and duration of a visit help to provide insights into how long individuals spend in dierent locations, which could be used to inform risk calculations or assessment of necessity of additional resources.

While CDRs provide extremely detailed information from calls and texts that can help us measure short term mobility, there are nevertheless important limitations. The first is that we can only measure movement for those individuals that use the mobile phone providers for which we have data access, and we have to make the simplifying assumption that those using other providers are not dierent in some fundamental ways and will have similar movement patterns. This might not necessarily be the case, especially if factors like socioeconomic status or ethnicity inuence the choice of phone provider. Additionally, while in the paper we have interchangeably talked about individuals and SIMs, we only have data on SIMs and we do not know how many SIMs might correspond to a given individual. This is especially problematic for Namibia, where we see a ratio of SIMs to the population of 1.39. Nevertheless, multiple SIM ownership might not be a problem for measuring short term movement because of the fact that some users are only active for a short period of time[20], which suggests that there might be sequential multi-SIM ownership, with each individual using only one SIM card at a time but switching SIM cards frequently. For this analysis where we only need two subsequent days with data to observe a short term trip taken, this sequential multiple SIM ownership will not be a problem. Another limitation of the data is that we can only infer movement from people's location based on calls and texts and we have no information on locations at times when they are not using their phone. It is possible that phone usage could be correlated with movement, for example individuals always use their phone when they travel somewhere to call home and let family know they are safe, but they might not call once they are home since they see their family, and so we might miss some movements. If this type of correlation is systematic, it could lead to bias in the short term movements that is measured so that we miss some moves that happen between certain location and not others.

The long term migration data is also mainly helpful in considering short term movements linked to visiting social contacts, but does not make it possible to study additional processes inuencing short term movement such as pilgrimages, which are excluded from the current paper. A limitation of the social network analysis is that given that for Senegal only one year of data is available, it is not possible to determine whether the contacts living in a different region migrated there previously or are originally from this area, or if the person is the one that migrated and the contacts are living in the place of origin. These are all important caveats that could inuence our results and impact the interpretation of the results.

Using unique data for two dierent countries, we have exploited the long temporal span of the Namibia data and the detailed network structure of the Senegal data to study the processes behind the relationship between short and long term movement. In addition, this is one of the first papers to combine mobile phone analyses from two different country contexts within sub-Saharan Africa. While it is necessary to study the factors of movement separately due to the different strengths of the two datasets, we are able to conduct similar analyses using census data for both country settings. As more and more work is conducted using mobile phone data in particular contexts, it will become important to examine how much of the analysis and results can be extrapolated to other country settings and to future years.

[20] Table A.1 for Namibia shows that there are many SIMs only active for a very short period of time

5. References

Adriansen, Hanne Kirstine (2008). "Understanding pastoral mobility: the case of Senegalese Fulani". *The Geographical Journal* 174.3, pp. 207-222.

Ahmed, Qanta A, Yaseen M Arabi, and Ziad A Memish (2006). "Health risks at the Hajj". *The Lancet* 367.9515, pp. 1008-1015.

Alkhalife, Ibrahim S (2003). "Imported malaria infections diagnosed at the Malaria Referral Laboratory in Riyadh, Saudi Arabia". *Saudi medical journal* 24.10, pp. 1068-1072.

ARTP, Autorité e Régulation des Télécommunications et des Postes (2013). *Observatoire de la telephonie mobile Tableau de bord au 31 décembre 2013*. url: http : / / www . artpsenegal.net/images/TB_mobile_dec-13.pdf.

Atamanov, Aziz and Marrit van den Berg (2012). "International labour migration and local rural activities in the Kyrgyz Republic: determinants and trade-os". *Central Asian Survey* 31.2, pp. 119-136.

Balcan, Duygu et al. (2009). "Multiscale mobility networks and the spatial spreading of infectious diseases". *Proceedings of the National Academy of Sciences* 106.51, pp. 21484-21489.

Bengtsson, Linus et al. (2011). "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti". PLoS Med 8.8, e1001083.

Blumenstock, Joshua E (2012). "Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda". *Information Technology for Development* 18.2, pp. 107-125.

Doocy, Shannon et al. (2015). "Internal displacement and the Syrian crisis: an analysis of trends from 2011-2014". *Conflict and health* 9.1, p. 33.

Erbach-Schoenberg, Elisabeth zu et al. (2016). "Dynamic denominators: the impact of seasonally varying population numbers on disease incidence estimates". *Population health metrics* 14.1, p. 35.

Etzold, Benjamin et al. (2014). "Clouds gather in the sky, but no rain falls. Vulnerability to rainfall variability and food insecurity in Northern Bangladesh and its eects on migration". *Climate and Development* 6.1, pp. 18-27.

Fall, Abdou Salam (1998). "Migrants' long-distance relationships and social networks in Dakar". *Environment and Urbanization* 10.1, pp. 135-146.

Fall, Papa Demba, María Hernández Carretero, and Mame Yassine Sarr (2010). *Country and Research Areas Report*.

Goldsmith, Peter D, Kisan Gunjal, and Barnabe Ndarishikanye (2004). "Ruralurban migration and agricultural productivity: the case of Senegal". *Agricultural economics* 31.1,pp. 33-45.

Henderson, J Vernon, Adam Storeygard, and Uwe Deichmann (2014). "50 years of urbanization in Africa: Examining the role of climate change". *World Bank Policy Research Working Paper 6925*.

Herrera, Catalina and David E Sahn (2013). "Determinants of internal migration among Senegalese youth". *Cornell Food and Nutrition Policy Program Working Paper 245*.

Huang, Zhuojie, Andrew J Tatem, et al. (2013). "Global malaria connectivity through air travel". *Malar* J 12.1, p. 269.

Linares, Olga F (2003). "Going to the city...and coming back? Turnaround migration among the Jola of Senegal". *Africa* 73.01, pp. 113-132.

Littrell, Megan et al. (2013). "Case investigation and reactive case detection for malaria elimination in northern Senegal". *Malar* J 12, p. 331.

Lynch, Caroline A et al. (2015). "Association between recent internal travel and malaria in Ugandan highland and highland fringe areas". *Tropical Medicine & International Health* 20.6, pp. 773-780.

Mokashi, Digambar Balkrishna (1987). *Palkhi: an Indian pilgrimage*. SUNY Press.

Namibia Statistics Agency (2015). *Namibia 2011 Census Migration Report*.

Osorio, Lyda, Jim Todd, and David J Bradley (2004). "Travel histories as risk factors in the analysis of urban malaria in Colombia". *The American journal of tropical medicine and hygiene* 71.4, pp. 380-386.

Oster, Emily (2012). "Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa". *Journal of the European Economic Association* 10.5, pp. 1025-1058.

Pison, Gilles et al. (1993). "Seasonal migration: a risk factor for HIV infection in rural Senegal". *JAIDS Journal of Acquired Immune Deciency Syndromes* 6.2, pp. 196200.

Poncelet, Alice et al. (2010). "A country made for disasters: environmental vulnerability and forced migration in Bangladesh". *Environment, forced migration and social vulnerability*. Springer, pp. 211-222.

Prothero, R Mansell (1977). "Disease and mobility: a neglected factor in epidemiology". *International Journal of Epidemiology* 6.3, pp. 259-267.

Rain, David (1999). "Eaters of the dry season: Circular labor migration in the West African Sahel".

Rinschede, Gisbert (1992). "Forms of religious tourism". *Annals of tourism Research* 19.1, pp. 51-67.

Siri, Jose G et al. (2010). "Signicance of travel to rural areas as a risk factor for malarial anemia in an urban setting". *The American journal of tropical medicine and hygiene* 82.3, pp. 391397.

Stork, Christoph (2011). *Namibian Sector Performance 2011*. url: http://www.researchictafrica. net/publications/ Evidence_for_ICT_Policy_Action/Stork_C_-_2011_Namibian_ Sector_Performance_Review.pdf (visited on 05/15/2017).

Stuckler, David et al. (2011). "Mining and risk of tuberculosis in sub-Saharan Africa". *American journal of public health* 101.3, pp. 524-530.

Tatem, Andrew J and David L Smith (2010). "International population movements and regional Plasmodium falciparum malaria elimination strategies". *Proceedings of the National Academy of Sciences* 107.27, pp. 12222-12227.

Tatem, Andrew J et al. (2014). "Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning". *Malaria journal* 13.1, p. 52.

The World Bank (2017). *World Development Indicators*. http://data.worldbank.org/datacatalog/world-development-indicators. [Accessed 31 May 2017].

Wesolowski, Amy et al. (2012). "Quantifying the impact of human mobility on malaria". *Science* 338.6104, pp. 267-270.

Wesolowski, Amy et al. (2013). "The use of census migration data to approximate human movement patterns across temporal scales". *PloS one* 8.1, e52971.

Wesolowski, Amy et al. (2015a). "Impact of human mobility on the emergence of dengue epidemics in Pakistan". *Proceedings of the National Academy of Sciences* 112.38, pp. 11887-11892.

Wesolowski, Amy et al. (2015b). "Quantifying seasonal population uxes driving rubella transmission dynamics using mobile phone data". *Proceedings of the National Academy of Sciences* 112.35, pp. 11114-11119.

WHO (2010). "Water woes in Senegal's holy city". *Bulletin of the World Health Organization* 88.1. http://www.who.int/ bulletin/volumes/88/1/10-020110/en/.

Wilson, Robin et al. (2016). "Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal Earthquake". *PLOS Currents Disasters*.

Yukich, Joshua O et al. (2013). "Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study". *Malar J* 12, p. 33.

Figure 1: Timing of the Last Call of the Day in Namibia.

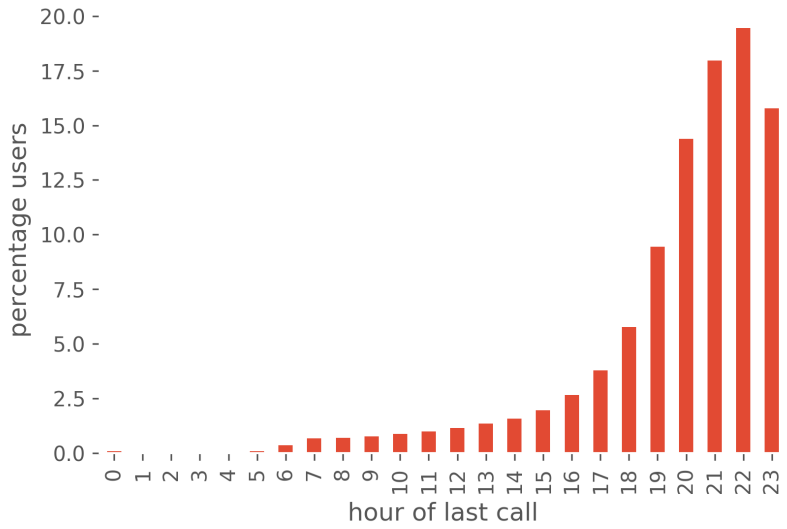
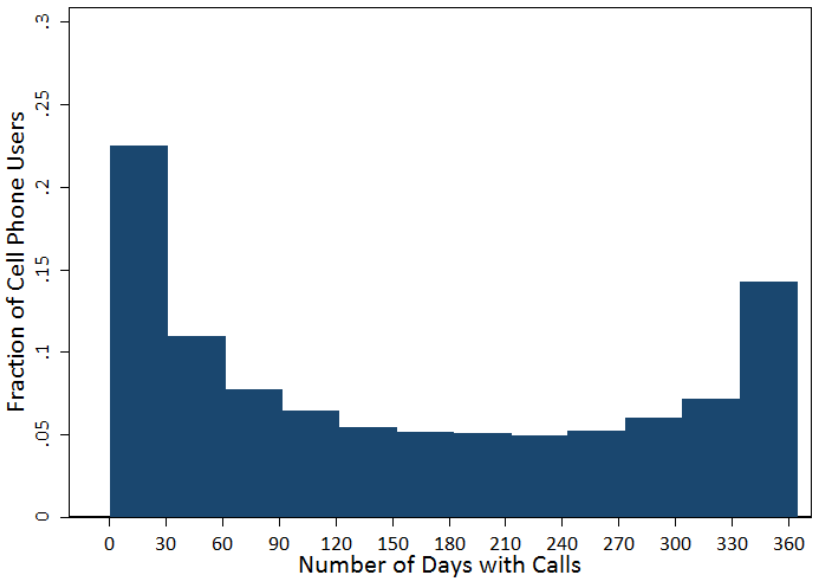


Figure 2: Histogram of Number of Days with Cell Phone Usage



(b) Namibia, 2013

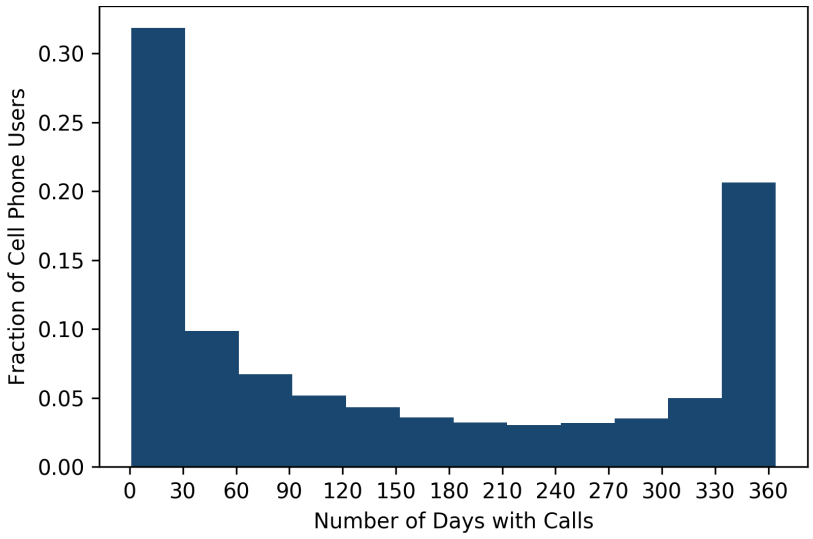
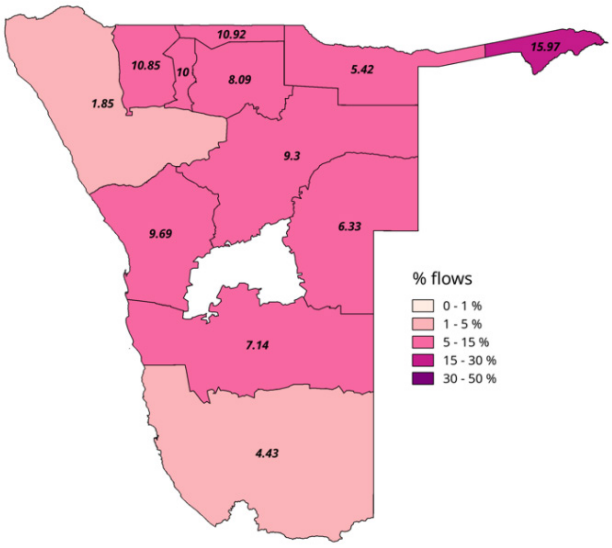
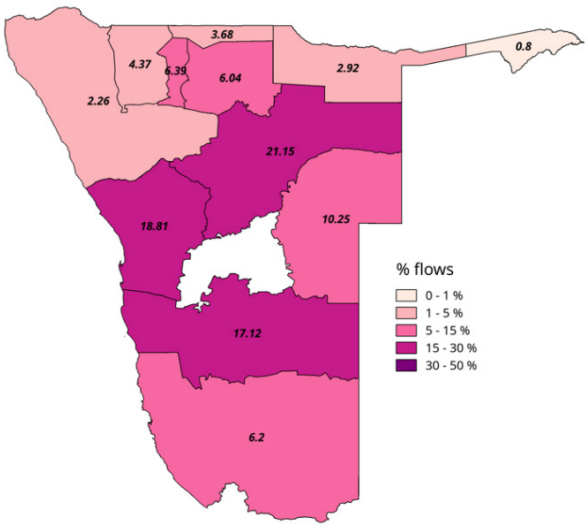


Figure 3: Movement into the Two Capital Regions

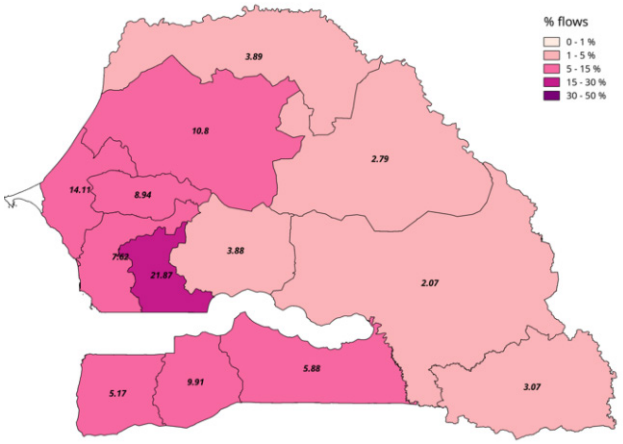
(a) Based on Census, Namibia



(b) Based on Cell Phone Data, Namibia



(c) Based on Census, Senegal(b) Based on Cell Phone Data, Namibia



(d) Based on Cell Phone, Senegal

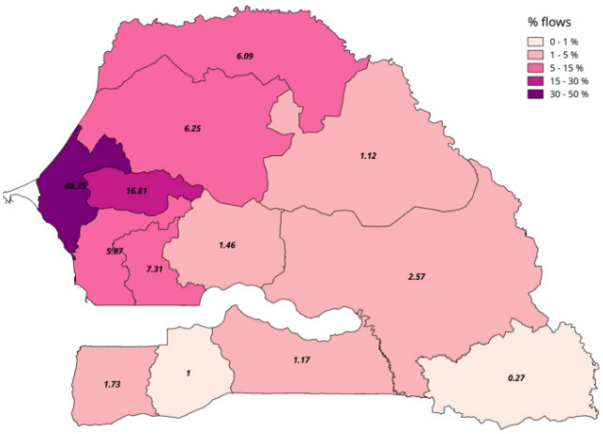
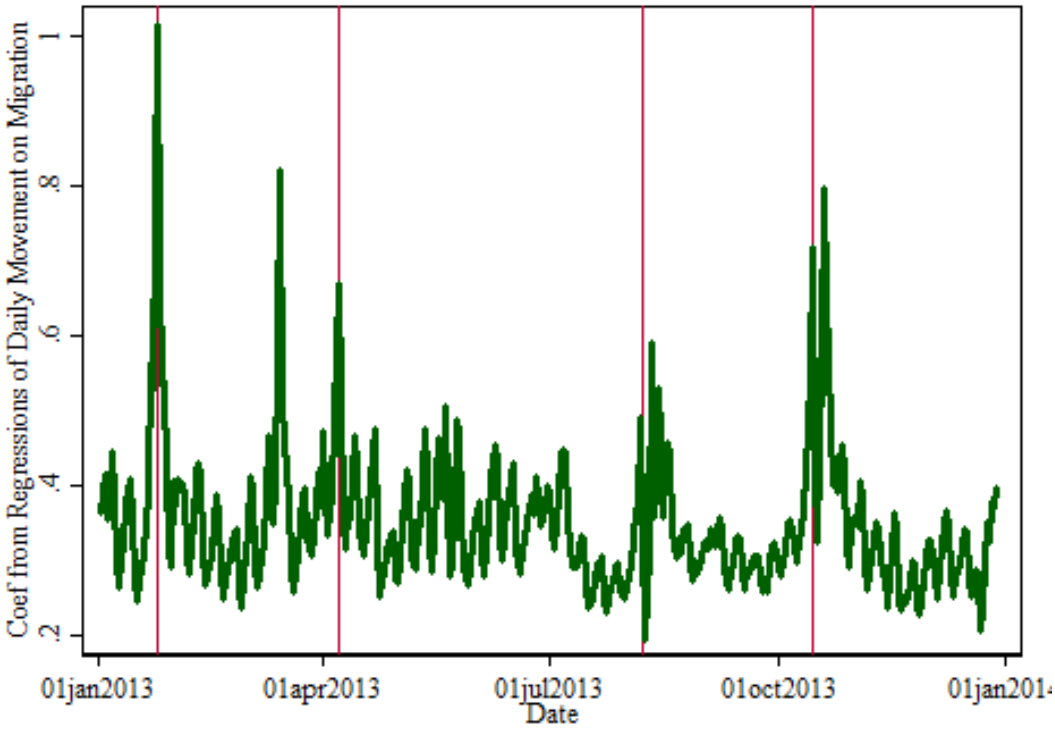


Figure 4: Short Term vs Long Term Movement

(a) Senegal, 2013



(b) Namibia, 2013

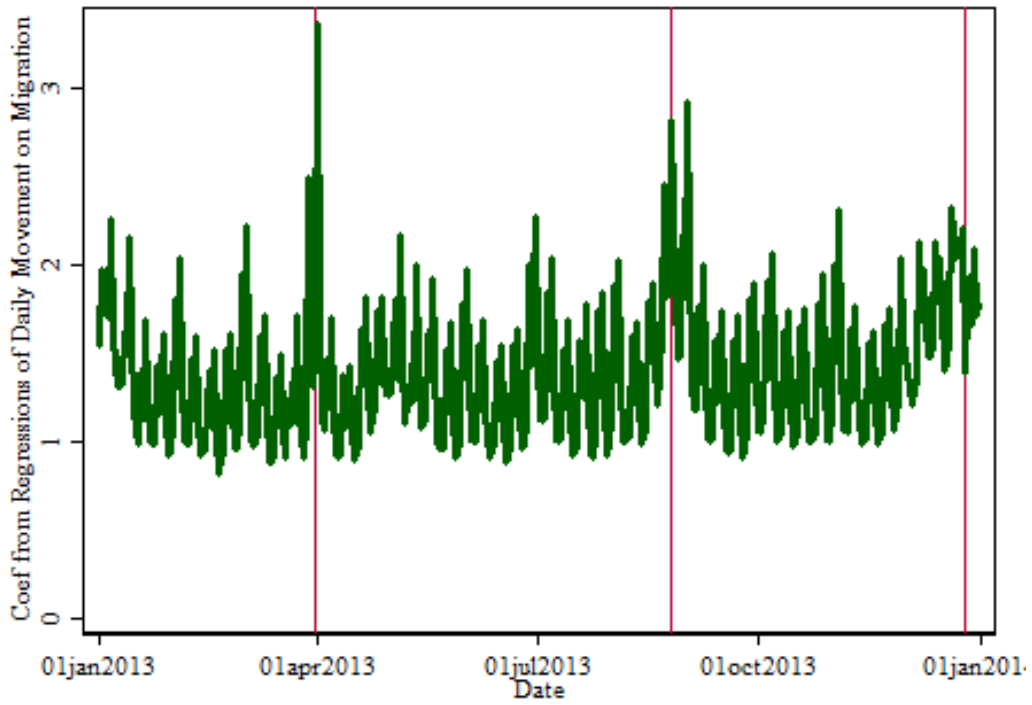
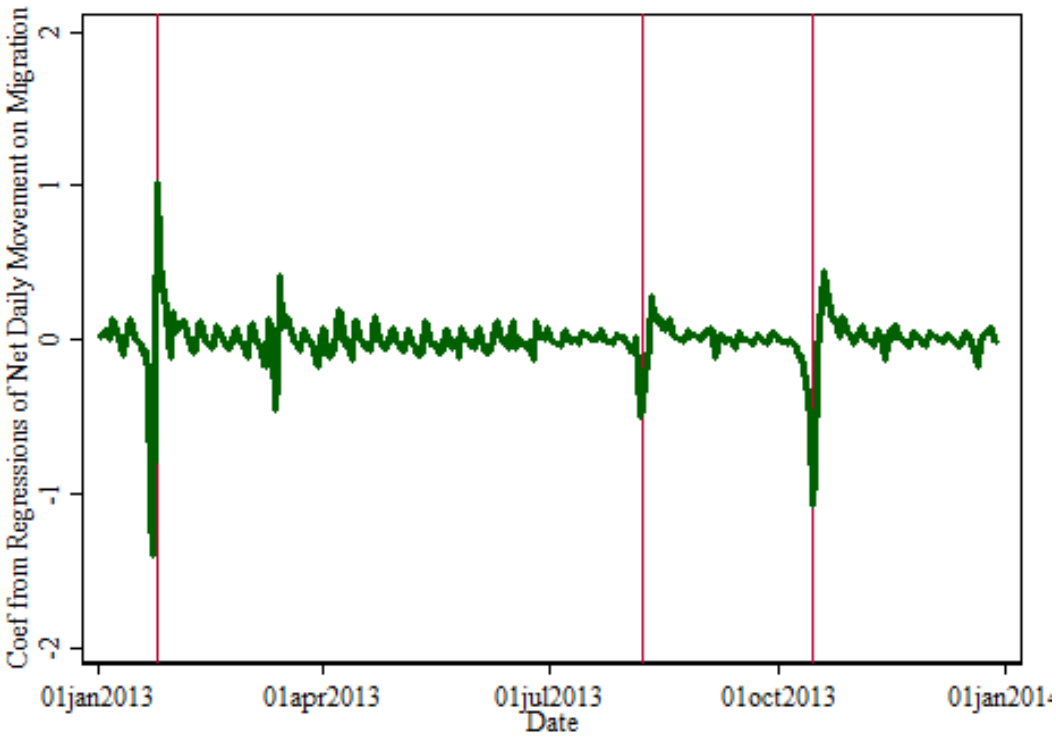


Figure 5: Net Short Term vs Long Term Movement

(a) Senegal, 2013



(b) Namibia, 2013

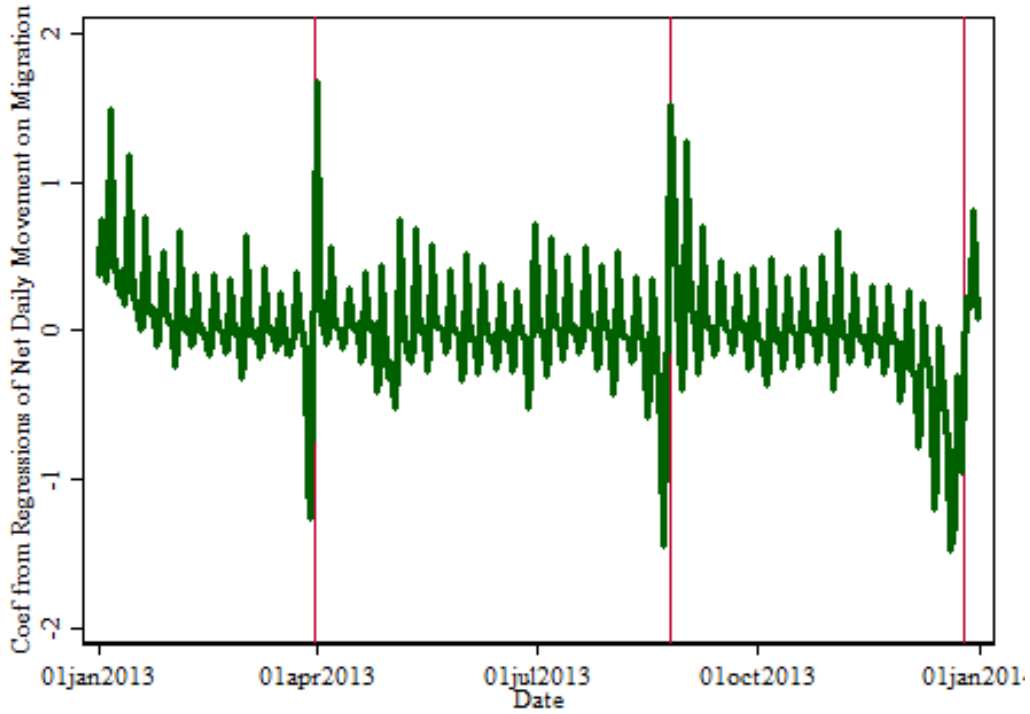
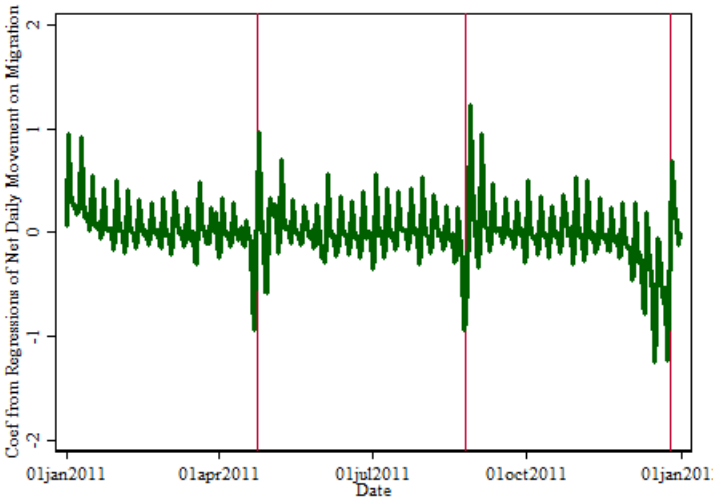
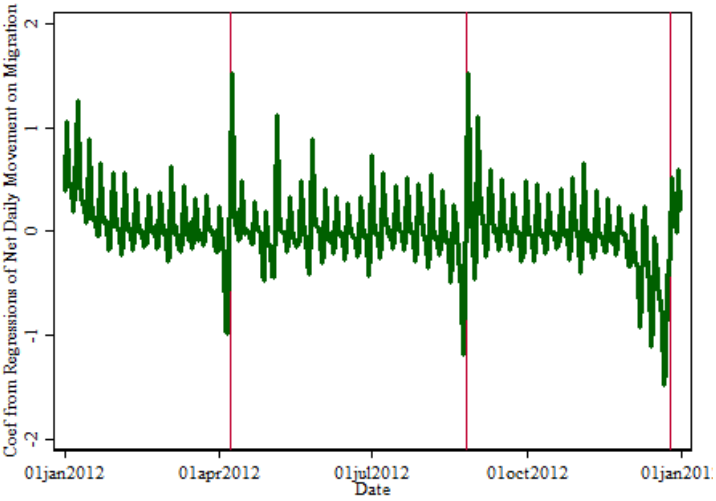


Figure 6: Net Short Term vs Net Long Term Movement for Namibia

(a) 2011



(b) 2012



(b) 2013

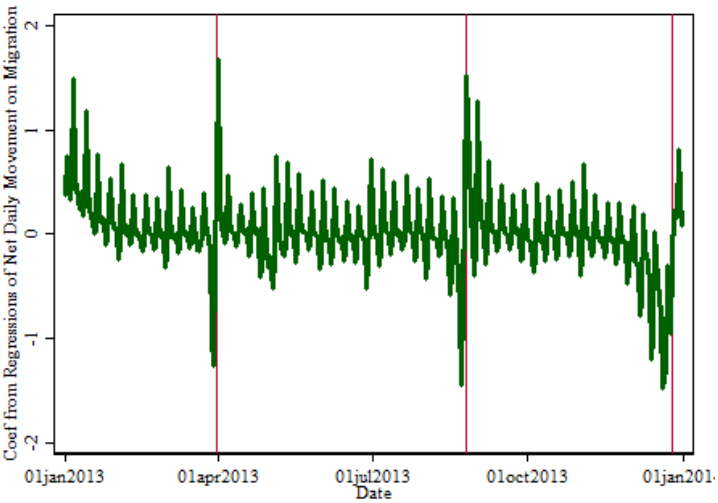


Figure 7: Percentage never visiting by admin unit

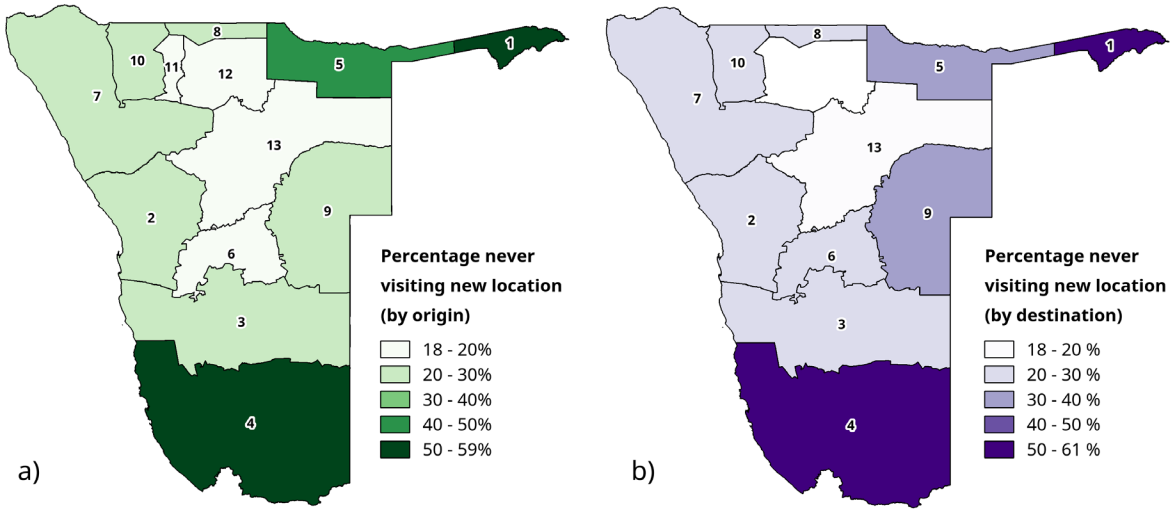
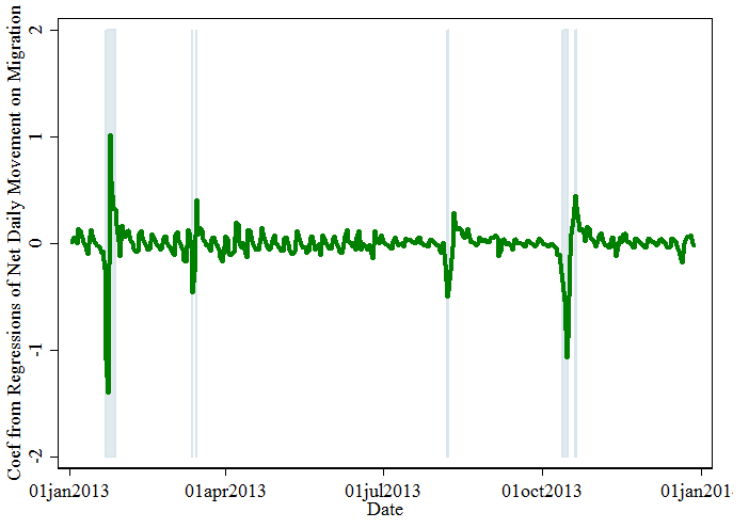


Figure 8: Timing for Targeting Travelers

(a) Senegal, 2013



(b) Namibia, 2013

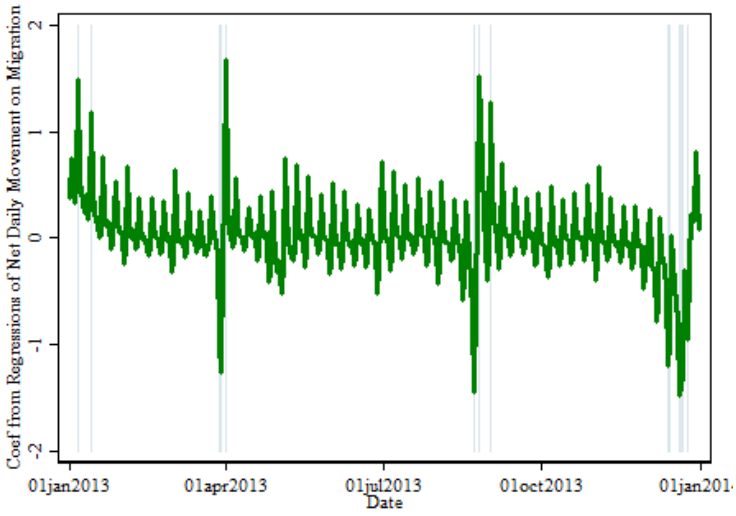


Table 1: Annual Regressions of Short term on Long Term, Senegal and Namibia

	(1) Senegal	(2) Namibia
Migrants entering	127.3*** (17.22)	505.5*** (164.8)
Constant	57,336*** (17,715)	-8,044 (25,175)
Observations	132	156
R-squared	0.328	0.394
Robust standard errors in parentheses, two-way clustered *** p<0.01, ** p<0.05, * p<0.1		

Table 2: Logit Model of Probability of Visiting a Given Region Before and After Migration to It

	(1) Visit Before	(2) Visit Before with Distance	(3) Visit After	(4) Visit After with Distance
Distance		0.996*** (0.000242)		0.996*** (0.000319)
Region 1	0.285*** (0.136)	1.701** (0.453)	0.227*** (0.0144)	1.527*** (0.229)
Region 2	0.620*** (0.0676)	0.804 (0.170)	0.606*** (0.0142)	0.792*** (0.0284)
Region 3	0.861 (0.275)	1.019 (0.136)	0.794** (0.0926)	0.872*** (0.0255)
Region 4	0.223*** (0.0515)	1.059 (0.316)	0.226*** (0.00666)	1.068 (0.126)
Region 5	0.658* (0.166)	1.342** (0.191)	0.576*** (0.0665)	1.268*** (0.0694)
Region 6	1.006 (0.200)	1.570** (0.282)	1.146** (0.0731)	1.740*** (0.0860)
Region 7	0.644 (0.180)	1.296 (0.283)	0.709*** (0.0274)	1.664*** (0.142)
Region 8	0.932 (0.454)	1.186 (0.259)	1.037 (0.0441)	1.580*** (0.0699)
Region 9	0.526* (0.174)	0.589** (0.135)	0.566*** (0.0563)	0.608*** (0.0181)
Region 10	0.787 (0.368)	1.045 (0.271)	0.984 (0.0547)	1.460*** (0.0630)
Region 11	1.674 (0.863)	1.372 (0.363)	1.690*** (0.148)	1.617*** (0.0200)
Region 12	1.223 (0.448)	1.144 (0.216)	1.248*** (0.0700)	1.319*** (0.0400)
Month FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	139,103	139,103	139,103	139,103
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Table 3: Annual Regressions of Short term on Long Term Movement and Social Network

	(1) Short Term	(2) Short Term	(3) Short Term	(4) Contacts
Long Term Migration	173.7*** (26.51)		86.46*** (27.25)	1,018*** (133.5)
Person Contacts		0.144*** (0.0239)	0.0857*** (0.0251)	
Constant	11,617 (48,058)	141,985** (52,676)	52,983 (39,450)	-482,603*** (147,952)
Observations	66	66	66	66
R-squared	0.449	0.466	0.502	0.681
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Table 4: Individual Level Relationship Between Short Term Movement and the Social Network

	(1) Probability of Visiting	(2) Probability of Visiting	(3) Interpolated Time in Region	(4) Interpolated Time in Region
Total Calls to Region	1.002*** (0.000261)	1.004*** (0.000278)	0.0267*** (0.00273)	0.0256*** (0.00249)
Total Calls to Region2		0.999998*** (1.63e-07)		-4.25e-06*** (1.04e-06)
Total Social Contacts in Region	1.144*** (0.00379)	1.144*** (0.00374)	0.355*** (0.0366)	0.536*** (0.0273)
Total Social Contacts in Region2		0.99986*** (7.24e-06)		-0.000709*** (0.000103)
Dummy at Least 1 Contact	5.655*** (0.150)	5.533*** (0.147)	0.380*** (0.0841)	-0.218*** (0.0554)
Constant			-0.323 (0.355)	-2.009*** (0.249)
Observations	212,520	212,520	451,902	451,902
R-squared			0.246	0.265
Region FE	YES	YES	YES	YES
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

A. Additional Tables and Figures

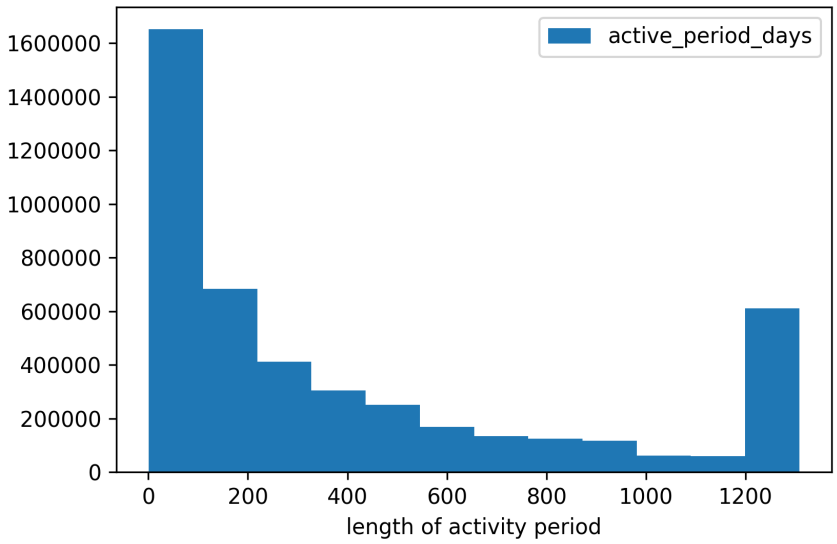
Table A1: Annual Regressions of Net Short term on Net Long Term, Senegal and Namibia

	(1) Senegal	(2) Namibia
Net Migrants	-1.079 (1.635)	-2.097 (2.289)
Constant	-1,194 (1,160)	616.9** (244.8)
Observations	66	78
R-squared	0.010	0.022
Robust standard errors in parentheses, two-way clustered *** p<0.01, ** p<0.05, * p<0.1		

Table A2: Percent of Migrants Making a Trip both Before and After or Never Based onOrigin and Based on Destination

Region	Origin		Destination	
	(1) Visit Before and After	(2) Never Visit	(3) Visit Before and After	(4) Never Visit
1	12.6	58.9	14.0	55.9
2	36.3	26.1	31.5	27.5
3	49.8	24.2	48.9	25.4
4	17.5	56.6	14.4	60.7
5	19.8	40.4	23.9	35.3
6	44.9	18.4	38.1	22.9
7	38.5	25.5	37.5	23.1
8	34.3	26.9	40.8	24.0
9	39.5	29.0	37.8	33.0
10	31.8	29.7	36.7	27.3
11	49.1	18.2	54.2	14.9
12	41.7	18.9	46.4	16.9
13	44.2	19.0	43.4	18.1

Figure A.1: Length of Active Period over 3 Years for Namibia



B. Relationship Between Short Term Movement and Migration when Pilgrimage Regions Included in Senegal

As mentioned in the text, there are two important pilgrimage sites in Senegal. The first is Tivaouane in the Thies region and the second is Touba in the Diourbel region. The pilgrimage site that draws the most people is Touba, where for the Grand Magal over a million people come to the city. In 2013, this Magal occurred twice, on January 1 and on December 21. Additionally, there is a smaller pilgrimage that occurred in June, 2013. In Tivaouane, there is a pilgrimage for the Prophet's birthday, Maouloud, on January 24.

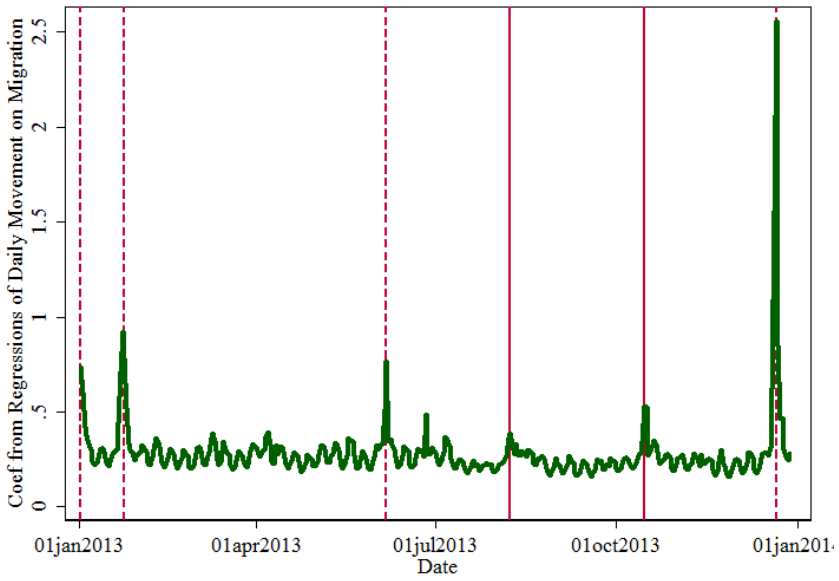
These pilgrimages represent a mechanism driving short term movement that is dierent from the one discussed throughout this paper. In the paper, we focus on long term migration patterns as a driver of short term migration, whereas in the case of pilgrimages the underlying driver is an important site that draws large numbers of people for specic events. Therefore, if we were to include the regions of these pilgrimages, especially since they draw such large numbers of people, it causes a distortion in the relationship between short term movement and long term migration patterns measured in the census

Figure B.1 shows results of daily regressions when all regions of Senegal are included in the analysis, including the two regions that receive large pilgrimages. In these graphs, in addition to solid red lines marking the most important Muslim holidays, dashed red lines mark the various large pilgrimages that occur. In the top panel we see that rather than some of the biggest holidays jumping out as large peaks in the correlation, we are seeing the largest peak as the Grand Magal in December. This is because there is some migration in the census data to the region where Touba is located; therefore, when running the regression for the day of the pilgrimage we see a very strong positive relationship as a large number of people enter the region. Yet, this does not represent a pattern of migrants that are visiting their place of origin or returning after a visit, but instead represents lots of pilgrims entering the region from places where there is also some migration.

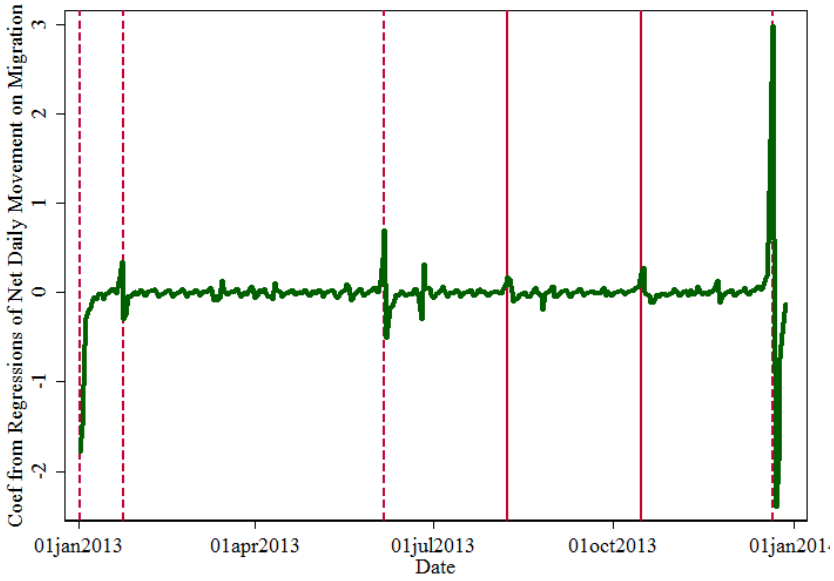
The dierence in the mechanism driving the short term movement becomes even clearer in the bottom panel. Rather than the expected relationship we saw between short and long term net movement in the paper, with individuals going in the opposite direction from where they migrated before a holiday (large negative coecient) and then going in the same direction as the migration after the holiday (large positive coecient), we see that for the pilgrimages there is a large positive coecient on the day of the pilgrimage and a large negative coecient on the day after the pilgrimage. This is because the typical migration pattern is for individuals to migrate into these areas.[21] Therefore, when regressing net short term movement on long term migration, we get a large positive coecient when a pilgrimage occurs and people enter the area and a large negative coecient as they exit the following day. When including the two pilgrimage sites, it masks the relationship we are trying to capture, which is long term migration and the expansion of social networks through long term migration leading to short term moves between two regions. This is especially true for the Prophet's birthday, which is not only a day when an important pilgrimage occurs to Tivaouane, but is also a public holiday when there is generally no work or school. This time o leads to lots of short term movement along migration routes, which we are able to capture in the main gures of the paper when we exclude the two regions with large pilgrimages.

The large movements that occur when there are pilgrimages are still important and can have important consequences for spread of disease, congestion, and economic activity, but they represent a dierent mechanism from the one studied in this paper. Additional work can explore this mechanism and potentially how to predict the size of the inux of pilgrims as well as the locations where pilgrims are likely to come from

Figure B.1: Senegal Results with Pilgrimage Regions
(a) Short Term vs Long Term Movement



(b) Net Short Term vs Net Long Term Movement



[21] Based on conversations with census ocials in Senegal, this is especially true for Touba in recent years since the city oers many social services, which can lead individuals to migrate there during economic downturns in the country. Additionally, the city has a policy of free land and free water, which also has led some to migrate there (WHO 2010).

Papiers de Recherche de l'AFD

Les *Papiers de Recherche de l'AFD* ont pour but de diffuser rapidement les résultats de travaux en cours. Ilss'adressent principalement aux chercheurs, aux étudiants et au monde académique. Ils couvrent l'ensemble des sujets de travail de l'AFD : analyse économique, théorie économique, analyse des politiques publiques, sciences de l'ingénieur, sociologie, géographie et anthropologie. Une publication dans les Papiers de Recherche de l'AFD n'en exclut aucune autre.

L'Agence Française de Développement (AFD), institution financière publique qui met en oeuvre la politique définie par le gouvernement français, agit pour combattre la pauvreté et favoriser le développement durable. Présente sur quatre continents à travers un réseau de 72 bureaux, l'AFD finance et accompagne des projets qui améliorent les conditions de vie des populations, soutiennent la croissance économique et protègent la planète. En 2014, l'AFD a consacré 8,1 milliards d'euros au financement de projets dans les pays en développement et en faveur des Outre-mer.

Les opinions exprimées dans ce papier sont celles de son (ses) auteur(s) et ne reflètent pas nécessairement celles de l'AFD. Ce document est publié sous l'entière responsabilité de son (ses) auteur(s).

Les *Papiers de Recherche* sont téléchargeables sur : http://librairie.afd.fr/

AFD Research Papers

AFD Research Papers are intended to rapidly disseminate findings of ongoing work and mainly target researchers, students and the wider academic community. They cover the full range of AFD work, including: economic analysis, economic theory, policy analysis, engineering sciences, sociology, geography and anthropology. *AFD Research Papers* and other publications are not mutually exclusive.

Agence Française de Développement (AFD), a public financial institution that implements the policy

defined by the French Government, works to combat poverty and promote sustainable development. AFD operates on four continents via a network of 72 offices and finances and supports projects that improve living conditions for populations, boost economic growth and protect the planet. In 2014, AFD earmarked EUR 8.1bn to finance projects in developing countries and for overseas France.

The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the position of AFD. It is therefore published under the sole responsibility of its author(s).

AFD Research Papers can be downloaded from: http://librairie.afd.fr/en/

AFD, 5 rue Roland Barthes
75598 Paris Cedex 12, France
ResearchPapers@afd.fr
ISSN 2492 - 2846

BIG DATA

TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES

2018