NBER WORKING PAPER SERIES

NEW CELLULAR NETWORKS IN MALAWI: CORRELATES OF SERVICE ROLLOUT AND NETWORK PERFORMANCE

Dimitrios Batzilis
Taryn Dinkelman
Emily Oster
Rebecca Thornton
Deric Zanera

Working Paper 16616
http://www.nber.org/papers/w16616

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2010

Research for this paper was funded by the NBER Africa Project on African Successes. We are grateful for research assistance from Eirin Forsund, Cecile Gaubert, Soren Larson, Ajay Shenoy and Ryan Wang. We also acknowledge the technical insight on cellular phone networks provided by Mung Chiang and Tian Lin in the Electrical Engineering Department at Princeton University. In addition, we thank the National Statistics Office in Malawi for assistance with data as well as the helpful staff at Zain and TNM, Malawi. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

New cellular networks in Malawi: Correlates of service rollout and network performance
Dimitrios Batzilis, Taryn Dinkelman, Emily Oster, Rebecca Thornton, and Deric Zanera
NBER Working Paper No. 16616
December 2010
JEL No. O20,O25,R39

ABSTRACT

Cellular technologies have become increasingly important in the developing world; infrastructure for mobile networks has expanded dramatically over the past two decades giving access to remote areas without previous phone service. Despite this expansion, relatively little is known about the correlates of the rollout of cellular phone networks or the performance of these networks. Since the rollout of cellular networks has been largely spearheaded by an active private sector in telecommunications, how demand-side and cost-side factors affect the timing of rollout and quality of network service is of particular interest. In this paper we use new data to estimate the correlates of cellular phone access and network performance across rural areas of Malawi. We compile a dataset which combines administrative data of the entire cellular network of Malawi with geographic and Census data to describe the rollout and the performance of the cellular network measured by the dropped call rate. We find that both demand-side and cost-side factors are important in determining the timing of network access, while demand-side factors appear most relevant for the dropped call rate, one metric of network quality.

Dimitrios Batzilis
University of Chicago
Chicago, IL 60637
batzilis@gmail.com

Taryn Dinkelman
Princeton University
357 Wallace Hall
Prospect Avenue
Princeton NJ, 08540
tdinkelm@princeton.edu

Emily Oster
University of Chicago
Booth School of Business
5807 South Woodlawn Ave
Chicago, IL 60637
and NBER
eoster@uchicago.edu

Rebecca Thornton
University of Michigan
611 Tappan St.
Ann Arbor, MI 48109
rebeccal@umich.edu

Deric Zanera
Malawi National Statistical Office
P.O. Box 333
Zomba
Malawi
dzanera@statistics.gov.mw

I. Introduction

Cellular phone technologies have become increasingly important throughout the developing world and in Africa in particular. Figure 1 shows growth in the population serviced by at least one cellular phone network in Africa over the last decade. The number of subscribers has more than tripled during this period and reached as high as 280 million at the end of 2007 (Buys et al., 2009). In contrast, growth in fixed line infrastructure has been much slower than growth in cellular phone coverage with approximately 90% of all African telephone subscribers consisting of cellular phone users (Paul Budde Communication Pty Ltd., 2009). There are challenges to the expansion of fixed line infrastructure, including widespread material theft and the need to cover large, low population areas. As a result, cellular network technology may be the best way to bring telecommunication services to most of Africa.

Popular media, as well as a growing body of research within economics, suggest that cellular technologies may be enormously important for improving productive efficiency, for providing a cheap method of transferring information and money across space and, ultimately, for promoting growth.[2] Despite this potential importance, we know relatively little about how this technology spreads and who receives access, especially in poor countries with limited infrastructure and limited ability to provide even very basic public goods. We also know little about how quality of cellular services varies within a country: while the presence of cell towers is necessary for cellular technology to have an impact on societies, reliability of these networks is also important. In most developed country settings we would expect that the timing of access to a new network provided by the private sector would correlate strongly with demand and cost factors. Further, we would expect network quality to be correlated with these factors, through network upgrading and strengthening. However, in a developing country setting it seems possible that other factors, such as political or ethnic influence, or the existence of complementary infrastructure could also drive placement and affect the quality of the service provided to an area.[3] How and where infrastructure is provided has potential welfare consequences.

In this paper we take a first step towards addressing these issues using data from Malawi. We focus on estimating the role of demand-side and cost-side factors in driving cell phone access and the performance

---

[2] See, for example, Aker (2009), Aker and Mbiti (2010), Arnquist (2009), Balakrishnan (2008), Hausman (2010). Jack and Suri (2009), Jensen (2007), Klonner and Nolen (2008), Krudy (2009), McGreal (2009) and Ngowi (2005)

[3] There is an established literature on the effects of corruption on public good provision and on economic growth more generally. See, for example, Alesina, Baqir and Easterly (1999), Tanzi and Davoodi (1998), Mauro (1995), Banerjee, and Somanathan (2007), Besley et al. (2004), Khwaja (2009), Easterly and Levine (1997), and Kimenyi (2006).

of these networks across rural areas. Understanding the role of these factors can put an upper bound on the importance of factors like political influence in driving access.

Our analysis takes advantage of new data on cell phone access and performance which we collected in Malawi. Our first outcome is cell phone network access. There are two cell phone providers in Malawi. We obtained detailed data from each of the two them on their exact tower locations and date of construction; our data cover every tower in the country. Using this information, we use geographic information system (GIS) software to construct coverage maps for the country over time. We focus on cellular phone access within a relatively small area, called an Enumeration Area (EA) of which there are about 9,200 in Malawi. Our data allow us, for each EA, to define coverage by year and to determine the first year in which the area had significant coverage. In addition, for the last two years of our data (in 2008 and 2009) we have some measures of network performance (dropped calls), which allow us to observe one aspect of the quality of service access provided to an area. In both cases, we link these data with Malawian Census data from 1998 (demographic proxies for demand) and to information on the geography of the country (variables which proxy for cost).

The first cellular towers in Malawi were built in the mid-1990s, and coverage was initially focused around the major cities (i.e., Lilongwe and Blantyre) and other cities covering tourist areas near Lake Malawi (i.e., Salima and Mangochi). However, increases in coverage were rapid: by 2004, 57% of the land area of the country had access and this figure increased to 86% by 2008. Conditional on having cell coverage, network quality varies. At the cell-tower level, the fraction of dropped calls reached up to 30% in some months of 2008 and 2009, although median monthly dropped call rates are below 3% in most months.

We perform three analyses to describe the placement and performance of cellular phone infrastructure in Malawi. In our simplest analysis, we regress network coverage (either a binary measure indicating whether more than 50% of the area is covered by a cell tower or a continuous measure of coverage) on the independent variables of interest for three years (2000, 2004 and 2008). This gives us a sense of which variables are important in determining access and how their importance changes over time. Second, we generate Kaplan-Meier survival curves and estimate Cox proportional hazard models to describe each EA's hazard out of the uncovered state. This allows us to use all of the information on coverage timing together. Finally, we estimate OLS regressions on the percent of dropped calls for towers covering each EA for the month of August 2008 to see whether the performance of the network is related to some of the same demand-side and cost-side variables used in the rollout analysis.

We begin with the role of demand. Our analysis suggests that area-level proxies for demand measured in 1998 are strongly correlated with the timing of access between 2000 and 2008. A number of proxies for area income (which we do not observe directly in the Census) are correlated with the timing of access. Areas with higher population density in 1998 receive coverage earlier. Higher levels of employment in agriculture correlate with delayed access, and higher levels of education in 1998 correlate with earlier access in the subsequent 10 years. The results are similar if we use a binary variable measuring coverage or a continuous measure of the number of towers covering these areas. These correlates correspond with demand-side factors that engineers and other cellular phone industry representatives report as important for defining cell phone market potential.

Following this, we move to estimate the correlation between cost-side geographic factors and coverage, in particular the altitude and slope of the location. Both of these variables affect the size of the potential market served by a particular tower (since line of sight may be more or less obstructed depending on terrain) and affect the cost of building and maintaining towers. We find some evidence that these factors matter. Places with higher slope are consistently less likely to have access; results on altitude are more mixed. Not surprisingly, isolation matters: areas that are far from a road are significantly less likely to get access to the network throughout the period. This last correlate may also capture aspects of consumer demand.

We find similar results when we estimate hazard models: areas with higher population density receive coverage earlier, as do areas with more education and less agriculture. Areas further from roads and in higher altitude areas have a tendency to get coverage later. The evidence on geographic correlates is more mixed in the hazard models.

Turning to the results on network quality, we have two related measures of dropped calls for each cellular network provider in Malawi. Since we only have network traffic data for the later period, we focus on describing how the performance of each network (measured by percent of dropped calls per month) varies in the cross-sectional data. Our measure provides one way of measuring network congestion. *A priori*, the relationship between market demand and cell network performance could be negative or positive: cell networks may be more congested in areas with more users per tower (i.e., higher market demand), but companies may respond to such congestion by differentially strengthening their infrastructure in these high demand areas.

In Malawi, the latter relationship appears to be the case for both companies. Using the traffic data for each firm separately, we find demand-side factors but not cost-side factors are important for predicting the percent of dropped calls. EAs with higher population density and a higher fraction of educated adults have fewer dropped calls. We also find that areas with more extensive prior network coverage have a much lower percent of dropped calls in the month. Not surprisingly, the density of the network matters for network performance in Malawi.

Taking the two analyses together, we find that within a Traditional Authority (TA), a larger unit of which there are about 300 in the country, demand and cost factors can explain between 40% and 60% of the cross-EA variation in cell tower coverage. For the measures of congestion that we use, our control variables account for 85% of the variation in dropped call rates across EAs within the same Traditional Authority. Although this obviously does not rule out roles for other factors like political influence or complementary infrastructure in determining access to this particular service, it does suggest that these other factors do not play a predominant role in allocation in Malawi.

II. Background

IIA. Country Background

Malawi is a small, landlocked country in Southern Africa. It is extremely poor, with a median daily wage for a casual worker of around US$0.40, and GDP per capita (PPP adjusted) of around US$800 in 2008 (CIA, 2009). In 1998, almost two-thirds of the population lived below the national poverty line and one-third of the population lived on less than US$0.25 per day (Benson et al., 2002). The large majority of Malawians work in agriculture (82% in 1998) and the country has a high population density and a youthful population (Benson et al., 2002). Education levels are low, as are school enrollment rates (although these have been rising in recent years); and the infant mortality rate is 83 per 1,000 (CIA, 2009).

Most types of public goods, infrastructure, and service provision are lacking in most areas. Roads are in poor condition: only 45% of national roads are paved. Although Malawi has an 800 km railway line, this line does not connect with lines in neighboring Zambia or Tanzania (as the line gauge differs). The only connection the country has to a port is through Mozambique's railroad, part of which has been closed since 1983 due to the civil war in that country.[4] Less than 60% of the population obtains water from a protected source while only 4.5% of households use electricity for lighting.

[4] CIA World Factbook (2009), Railways Africa: http://www.railwaysafrica.com/category/africa-update/sadc/malawi.

Despite high poverty rates and limited infrastructure, Malawi compares favorably with many other African countries. In the mid-1990s, the country transitioned from a dictatorship to a multiparty democracy, which has survived since then. Malawi has experienced no major civil conflict over this period, and elections have been held on schedule with only limited accusations of irregularities. Relative to other areas in Africa, tribal tensions have been limited, although post-democracy tensions have been growing.[5]

IIB. Cellular Telephones in Malawi

Overview of Providers The first cellular operator, TNM (Telekom Networks Malawi Limited), was initially majority-owned by the government of Malawi. This operator was licensed by the government agency MACRA (Malawi Communications Regulatory Authority) a government-appointed regulatory board, in 1995.[6] In the early years, growth in the network was concentrated in the four main urban areas and in tourist destinations; network coverage in rural areas was slow and coverage rates were low. The histogram in Figure 2 illustrates some of this rollout as it occurred in rural parts of the country. It shows the fraction of Census Enumeration Areas (EAs) that gain new access to a cellular network in each year from 1995 to 2009 (we define "access" in more detail below). The figure clearly shows the initial slow coverage of EAs in early years; many more rural EAs are covered by a cellular network from 2003 onwards.

The slow growth and poor coverage prompted reform of this sector in 1998, at which point a second private operator, Celtel, was awarded a government license. Celtel was purchased in 2008 by Zain, which is headquartered in Kuwait and has a strong regional presence in over 20 other African and Middle Eastern countries. By late 1999 both operators were active in establishing their networks across the country, and by 2007, the number of cellular phone subscribers (pre-paid and post-paid) had risen to over 1 million people, or 25% of the population (ITU, 2008). Many more than these subscribers have access to this telecommunications service, as most consumers use a cell phone on a pre-paid basis. Currently, over 70% of the country has access to the cellular phone network and Zain claims to serve 70% of the current Malawian market.

[5] For a detailed analysis and history of tribalism and ethnicity in Malawi see Vail (1989).
[6] The Communications Sector Policy Statement establishes MACRA, the Malawi Communications Regulatory Authority, as the body responsible for regulating telecommunications, posts, broadcasting and the radio frequency spectrum.

Licensing Process As in other countries, companies must have a license to operate a cellular network in Malawi and these licenses are typically awarded after a tendering process. TNM and Zain were both granted licenses to provide cellular services by MACRA. License fees are required for participation in this market, both at the initiation of the license agreement and on an annual basis after that (annual fees vary between 5% and 7% of company revenue). Additional fees apply for new services (e.g., internet service) that each company chooses to provide. Since 1998, there have been two subsequent tenders announced for the awarding of a third license; however, no third operator has entered the market yet and so we confine our discussion to the combined activities of TNM and Zain in the rest of this paper.[7]

One important aspect of the license agreements is that each company was required to build cellular phone towers in certain target areas before specific deadlines. Malawi's 1998 Communications Sector Policy Statement commits MACRA to "ensure extension of modern telecommunication services to rural areas" and do this "according to a defined program covering rural areas" . In discussions with engineers and managers of both companies in Malawi, many of the target areas in rural parts of the country would not have been viewed as commercially viable sites, at least initially. During the period in which we analyze the rollout, these two companies were therefore expanding the network partly into areas that appeared profitable and partly into areas which were important to connect to meet license obligations. We observe both types of areas in our sample.

Cellular Phone Costs and Access Although cellular phone infrastructure has expanded rapidly to cover almost the entire country, ability to access the network in Malawi is still somewhat limited to wealthier individuals. Malawi's Integrated Household Survey of 2004 indicates that 17% of urban households and only 1% of rural households owned a cellular phone (although the rates of access to a cell phone are likely to have increased substantially since then).[8] The initial cost of buying a handset still represents an important barrier to using the network: in 2009, the cost of the cheapest handset offered by Zain was MK 2,500 = US$18, or about 50 days of day-laborer work.

In addition to the cost of a handset, tariffs are high relative to other developing countries, although there is some indication that they have fallen over time. Service prices for the two providers are a difficult to keep track of and standardize over time, because price setting responsibility is opaque, tariff changes sometimes occur through temporary promotions and changes are not required to be reported to MACRA (the regulatory authority has no power to set prices). At various points during the period we study, there

[7] How and to whom licenses are awarded may in itself be of interest related to public good and infrastructure provision in Africa. We do not address this specifically in this paper.

[8] In other surveys that have recently been conducted in Malawi in subsequent years, 72.5% of urban men had their own cell phone with 75.5% of those without a cell phone having access to one; 23.2% of rural men had their own cell phone (Godlonton and Thornton, 2010).

have been changes in both pre- and post-paid rates, as well as changes in how calls are billed (per minute versus per second) and changes in prices for within and outside network calling. Given this difficulty in standardizing the appropriate "unit of telecommunications" to use to compare prices over time (a difficulty present in all research related to cell phone services, not just in the context of Malawi), no single statistic will capture the true cost of using the cellular phone network.

However, to get an overview of how prices have changed on each network over time, we created an average consumption basket (making assumptions about average call durations, call destinations (in or out of network) and average call timing (peak/off-peak) using information from cellular phone traffic data from the later 2008-2010 period) and calculated the price per second of this basket at the relevant tariffs between 1999 and 2010. On average, the price per second of talk time across both companies was about 1.2 US cents in the late 1990s; by 2009, this average price had fallen by 50%, to 0.6 US cents per second. Despite this substantial reduction in the price of airtime, cell phone rates in Malawi are still about 5 times higher than the per-second cost of using a cellular phone in the US (OECD, 2009). As a result of these generally high tariffs in Malawi, individuals often communicate using cheaper options such as text messaging (10 US cents per message), or "flashing"/"pinging" (ringing and hanging up to signal a wealthier party to call back and pay for the call). It is also worthwhile to note that while it costs money to place a call or to send a text message, receiving a call or a text message is free. Thus, even though prices may be high enough to prohibitive many people from initiating communication, having access to network coverage may be beneficial, if only to receive calls.

Access to a cell tower is clearly necessary to use the new technology, but the network also needs to provide a reliable service in order to be useful for individuals. We know remarkably little about how well cell networks function in Africa. Given the dismal quality of many other types of infrastructure across the continent (roads and railways in particular), it is important to know whether cell infrastructure performs any better and to understand more about the correlates of better performance. In Malawi, network performance varies substantially across the country. Using the monthly data we have access to, we find individual towers report a dropped call rate of up to 30% in some months. Undoubtedly, poor network performance based on this particular measure limits some of the potential effects of cell phones on society and the economy.[9]

Lastly, individuals must have access to some means of charging their cellular phone handset. Given the low prevalence of electricity in Malawian homes (according to the NSO (2004), only 6% of households

[9] It is difficult to find accurate statistics on dropped call rates in the US for comparison. A recent article in the popular press (Manjoo, 2010) reports statistics collected directly from several cell providers that indicate monthly call drop rates of below 2%.

(32% urban and 2% rural) had access to electricity), individuals in more remote areas must resort to charging cellular phones using car batteries, or must pay for additional charging services in order to use their phones.

Our analysis in this paper primarily focuses on the *availability* of the cellular phone network from cellular towers, rather than *use*. For the analysis, we assume that an area has coverage if it is reached by cellular phone towers, ignoring the fact that some people may have differing ability to take advantage of the network. We augment the coverage analysis with an examination of how well each company provides a reliable service to an area, where we define one particular measure of service reliability below.

III. Data

This paper uses a number of new datasets, which we discuss in turn below. We have collated four types of data for the description of infrastructure rollout and network performance: (i) administrative data on the placement and timing of new cellular phone towers from both companies, (ii) engineering data on cell phone traffic at the tower level for several months in 2008 and 2009; (iii) geographic data describing Malawi's physical terrain and (iv) demographic data from the Census. We link these datasets at the Enumeration Area (EA) level and at times will also refer to the larger Traditional Authority (TA) in which EAs are nested.

IIIA. Data on Cellular Phone Network Coverage

We begin with our data on cellular phone network coverage. As mentioned above, there are currently two cellular networks in Malawi: TNM and Zain. To generate information on overall coverage over time, we collected data from both providers. These data include the precise GIS location and construction date for each cellular tower for the period from 1995 to 2009.

Using these data on latitude and longitude of the cellular towers, as well as data on elevation, we determined what areas of the country became part of the cellular phone network and when. We used the *Viewshed* tool of the ArcGIS software, which identifies the points in a map that can be seen from a set of observation points. For a full description of the tool, the reader can consult the help website of ESRI.[10]

With the *Viewshed* analysis, we assigned to each point in the map a value that is equal to the number of cellular phone towers (of either network) that are (a) visible from that location and (b) within a distance of

[10] http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How%20Viewshed%20works

30 km of the tower. The idea is straightforward: there can be no coverage if the tower signal is blocked by physical obstructions, and hence the tower is not "visible", or if the source of the signal is too far away. Discussions with engineers in Malawi suggested that 30km is an appropriate range for the towers, although this distance may not be exact in all cases.

Each EA consists of many points on the map of Malawi. The cellular coverage value assigned to an EA is the expected (average) number of towers that cover a randomly chosen point that lies within the area of the EA. In mathematical terms, if x and y are respectively the longitude and latitude of a point in the map, $f_j(x,y)$ equals the number of cellular sites that cover the point (x,y) in year j, and A_i is the area of the EA, then the coverage value for the EA can be expressed as:

$$Coverage_{i,j} = \iint_{A_i} f_j(x,y)\,dxdy$$

It is important to note that having a coverage value of, for example, 0.50 does not necessarily imply that 50% of the area is covered. For example, if 25% of an EA's area is covered by two towers simultaneously, then the value of the coverage variable is 0.5 as well. In 2008, the average coverage measure ranges from 0 to 47. Perhaps the best way to think of this measure is as an indicator of cellular network intensity in the EA, where EAs with the highest coverage value of 47 have access to the densest network.

We use this coverage measure in several ways. First, we run OLS regressions on a binary indicator of coverage in early (2000), middle (2004) and late (2008) periods of our data. For these regressions, we define an entire EA as "covered" in the years when the above coverage value was larger than 0.5. Second, we run OLS regressions on the continuous coverage variable in each period. Finally, when we analyze hazard models, we experiment with various definitions of "covered": whether the area is covered at all (coverage>0), whether the coverage is over 0.5 and whether the coverage is over 1.0.

Panel A of Table 1 provides basic information about the fraction of all, urban and rural, EAs that are covered by at least one network in each year, using the binary definition of covered as coverage greater than 0.5.[11] As early as 1999, most of the urban EAs were covered by the networks, but only 20% of rural areas had access. Coverage in rural areas started to ramp up from 2003 onwards. Because of this quick initial saturation of network coverage in the main urban areas and the greater variation in coverage in rural areas, the analysis in this paper excludes the urban areas and focuses on rural Malawi, where the

[11] Although there are about 9,200 EAs in all of Malawi, we have data for only 8,924 so our analysis focuses on those areas.

majority of the population resides. In Figure 2, as described above, we show the fraction of rural EAs with new cellular network coverage for each year. Maps in Figures 3 to 5 provide a visual representation of the cellular coverage data by year in 1997 (when only TNM was active), 2004 and 2008.

IIIB. Cell phone traffic data and network performance measures

We collected traffic data from each cell company for several months in 2008 and 2009. We limit our analysis here to the month of August 2008, taking this as a representative month during the period. For each cell tower, we calculate a measure of average percent of calls dropped during the month. The measures differ slightly across the two network providers since the level of aggregation differs. One provider's data was available at the daily level, while the other's was available at the monthly level. Because of this difference, we do not aggregate the data across providers, and rather present statistics separately; because of the potential sensitivity of this information, we anonymize each provider. In this case, Company 1's drop call rate is computed using monthly data on the number of dropped calls and the number of total attempted calls, while for Company 2, we use the reported fraction of dropped calls by day and average these up to the monthly level.

Each tower t is placed in a unique EA. However, because there is a 30 km radius around each cell tower that receives signal from this tower, a typical tower will provide service to more than one EA. We address this by assigning the dropped call information from tower t to *each* EA which falls in its radius. For EAs that are covered by multiple towers, we take the average of the dropped call rate across all towers from which it receives coverage. EAs that do not receive any network coverage from one or the other company do not have any traffic data assigned to them. In Table 1, Panel B, we see the mean dropped call rate (fraction) for each cell company for the set of rural EAs with cell coverage. Company 1's dropped call rate is higher at 2.0% per month, with Company 2 being 0.9% per month in August 2008.

Note that the fraction of dropped calls is not a perfect measure of network congestion. Dropped calls may occur for reasons unrelated to the quality of the network infrastructure or excess demand for services on the existing network. For example, weather and geographic conditions can also affect the fraction of calls dropped. We use the dropped call rate as one indicator of network performance although we recognize that other factors can also contribute to variation in this outcome.

IIIC. Geographic Data

Since profit maximizing cell phone providers take the cost of building and maintaining cell towers into account when deciding how to optimally expand the cellular network, we collated important geographic data to be used in our rollout analysis. The source of the elevation data we use is the national map seamless server that contains high-analysis international geographic data.[12] As mentioned in section IIIA, the elevation data were used in the *Viewshed* analysis, in order to determine which EAs were covered by towers. Moreover, altitude data were used to calculate the slope of each point in the map. This was done using the slope calculating tool in ArcGIS.[13] The average altitude (measured in meters at the EA level) and slope (measured in degrees at the EA level) are used as controls in our regression analysis. The slope is essentially a measurement of the steepness of the terrain. EAs with precipitous cliffs or mountains have steeper slope than EAs that are located in plains or plateaus.

We created a measure of distance to the nearest road for each EA (using the centroid of the EA as the starting point) where the national roads data from 1998 was provided from the National Statistics Office.

Panel C of Table 1 shows some simple summary statistics on geographic data at the EA level – the average altitude and slope observed in the data. The average height of an EA is almost one kilometer above sea level although there is a great deal of variation across the country. Malawi's terrain varies from the high plateaus in the northern and central parts of the country (west of Lake Malawi) to the relatively flatter area around the Shire River in the south. We also see in Table 1 that rural areas range from being immediately next to a road to more than 50 km distant.

IIID. Demographic Data at Baseline

Baseline demographic data comes from the 1998 National Census, provided by the National Statistics Office in Malawi. We obtained a 100% sample containing detailed information on each household's membership, levels of education of adults, occupation, religion and other variables. One important aspect of the data which is worth mentioning is the limited information on income and labor force participation. These questions were not asked in 1998, possibly due to the fact that the majority of Malawians are subsistence agricultural workers and collecting accurate income and labor data would lengthen the questionnaire. To link the National Census to the geographic data, we use spatial files also from the National Statistics Office which were collected in 1998 during the Census. These files contain geographic information on the boundaries of the administrative divisions and the Census enumeration areas.

The Census data follow the administrative structure of the country – Regions, Districts, and either Traditional Authorities (TAs) or administrative wards. Malawi consists of three regions – Northern, Central and Southern. Each region is divided into Districts, with a total of 27 in the country. Districts

[12] http://seamless.usgs.gov/index.php
[13] Following the process described at http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Calculating_slope

usually contain at least one larger peri-urban center with a district hospital, police unit and commissioner's office. Each District contains a number of Traditional Authorities in the rural areas; in 1998 there were a total of 250 Traditional Authorities. Traditional Authorities are governed by a "Traditional Authority", which is a non-elected office and determined by the tribal politics of the area. The four urban centers of Malawi consist of Blantyre, Lilongwe, Mzuzu, and Zomba; we exclude these urban areas from our empirical analyses.

Enumeration areas (EA) are defined by the National Statistics Office mainly for data collection, while Districts, Traditional Authorities, and administrative wards are common political divisions used by the government and other institutions. There are approximately 9,200 EAs across the country, and we observe a total of 8,924 of them. We restrict to the 8,118 rural EAs for the analysis. The lowest level of disaggregation consists of villages, governed by a village chief. EAs do not uniquely contain villages.

Panel D of Table 1 shows summary statistics for the Census demographic measures at the EA level. As noted above, Malawi has a relatively youthful population with an average age of 22 in rural areas. Education levels are low with an average of under 3 years of completed schooling for females (males) ages 15 to 49 (54). Almost 90% of adults in rural areas work in agriculture, and the average population density is high, at 314 persons per square kilometer, although this average masks considerable spatial variation in density.

An important aspect of the 100% Census data is that they provide the most complete, albeit crude, set of proxies for market demand that cellular phone network engineers would have had available to them at the beginning of the period. The data also represent the most complete picture of population density and socioeconomic characteristics for most years between 1998 and 2010. A more recent Census was conducted in 2008, but the data are still unavailable to the public. As a result, apart from smaller surveys (an income and expenditure survey from 2004 and Demographic Health Surveys) , companies would have had to collect their own market-level data to supplement the national coverage of the 1998 Census or use published population projections in order to construct up-to-date measures of potential market demand. What we will see is that the 1998 Census data variables are strong predictors of cell coverage expansion, particularly in the earlier and middle periods of this expansion; and that these demographic variables from 1998 are also important predictors of network congestion ten years later.

IV. Results: Patterns of Rollout and Correlates of Coverage

Graphical Evidence We begin by illustrating our basic results in graphical form, using maps and estimating survival functions.

The maps in Figures 3, 4 and 5 show our basic descriptive evidence on rollout of cellular phone coverage over time. Figure 3 illustrates the areas that had cellular coverage in 1997, when only TNM was active. Coverage is concentrated around the large cities (Lilongwe the capital, and Blantyre the industrial and market center), and a several of the main tourist destinations around Lake Malawi (i.e., Salima and Mangochi). By 2004 (Figure 4) coverage has expanded significantly, moving further out from the cities and specifically, along the main road network. Figure 5 shows coverage in 2008. At this point, coverage has expanded significantly beyond the main road network and the majority of the country has some cellular phone access. We should note that, even in this later period when coverage extends more widely, it still appears to be concentrated around major population centers and tourist centers.

In Figures 6.1 to 6.3, we present estimated Kaplan-Meier survival functions for the time to first cellular coverage within our sample of EAs. These graphs serve as the first illustration of our evidence on correlates of rollout. In these figures, "time of coverage" (on the x-axis) denotes when an EA first received cellular network coverage and ranges from 1995 (year 1) to 2009 (year 15). Each line shows the fraction of EAs that have not yet received cellular phone coverage; as coverage spreads across the country, these "uncovered" lines step downwards. We plot survival functions separated by one measure of demand and two measures of cost.

We treat population density (measured in 1998) as a proxy for potential market demand, while slope and altitude within the EA affect the cost and feasibility of building cellular towers. Although the three factors are clearly related (e.g., population density is higher in flatter EAs), we examine the relationship between each one individually and the time to first coverage for each EA. We show survivor functions for each of the groups ordered from low to high values of slope, altitude or population density as well as the 95% confidence intervals for each group.

Population density is clearly correlated with time-to-coverage, as Figure 6.1 shows. Places with the lowest population density get coverage much later in the period while about 25% of EAs in the two highest density groups get coverage by 1999. Although all density groups see more EAs getting coverage after 2003 (year 9), the gap between the survivor functions increases after 2004. This suggests that market demand factors are still important in prioritizing network expansion, even after the backbone structure of this network is in place.

Figure 6.2 provides evidence that in the early part of the period (1995 to 2003), areas of different slope appeared equally likely to make the transition from no coverage to coverage (the lines and confidence intervals of the survivor functions overlap). After 2003 (year 9), areas with lower values of the slope variable (i.e., flatter areas) are significantly more likely to make this transition, compared to places in the steepest gradient group. Combining this with the prior graph that indicates the important role for population density in early years, the figures suggest that cost considerations appear more important in discriminating among areas in the second half of the period, once the basic network structure was already established.

The message from the survival functions for EAs of different altitude in Figure 6.3 is more mixed: in some periods, EAs in different altitude groups have very similar chances of receiving cellular coverage by a certain year. For many of the later years, places with *lower* altitudes actually have a higher likelihood of transitioning to a covered state. While we might think that locations with the highest altitude would be those places most likely to receive network towers since they offer uninterrupted line of sight, these locations may not be the lowest cost options (e.g., maintenance costs may rise as towers are built on taller and taller mountains) nor are they likely to represent areas with large market potential (EAs at lower altitudes have higher population density). Hence, the advantage of building towers in lower altitude EAs seems to emerge only later on in the sample period.

Demographic Correlates While the figures discussed above are illustrative of some of the important factors related to increased network access, many of these factors are correlated (i.e., altitude and population density).[14] We next turn to examining the determinants of access in a multivariate regression framework. We begin by using an OLS framework to estimate the correlates of cell phone access in three different years (2000, 2004 and 2008). We estimate a regression of the following form:

$$Phone_{kc} = \beta_0 + \beta_1 PopDens_{kc} + \beta_2 PctAg_{kc} + \beta_3 Educ_{kc} + \beta_4 Age_{kc} + \gamma_k + \eta_{kc} \qquad (1)$$

In this regression, $Phone_{kc}$ is an indicator for phone coverage (binary or continuous) or for quality of network (dropped calls) in TA k and EA c. The regressions are estimated at the EA level (denoted by c) with a full set of TA fixed effects (γ_k). This means that we identify off of the EA-level variation. Demographic variables (population density, percent of workforce in agriculture, average education and average age) are computed by aggregating 1998 Census data to the EA-level. These variables are proxies for potential market demand in two ways. First, they represent the potential customer base for originating calls, especially in areas with more educated adults earning higher incomes. Second, they capture

[14] In our sample of rural EAs, the correlation between population density and altitude is -0.10, the correlation between density and slope is -0.17 and the correlation between population density and distance to a road is -0.30.

potential customer base for receiving calls from other parts of the country – therefore, even poor areas with high population density could be more attractive markets than areas with lower population densities.

Table 2 turns to estimating our first correlates of coverage: demographic variables that proxy for potential market size. Panel A estimates the impact of these demographic variables on a binary for coverage in 2000, 2004 and 2008. Panel B estimates similar regressions using the continuous measure of coverage.

We find that most of the demographic variables seem to be important in driving coverage timing. Areas with higher education levels and a lower share of individuals working in agriculture, both typically correlated with higher income, tend to get coverage earlier and are still more likely to have coverage in 2008. The relationship between education and cellular network expansion seems to grow stronger over time. Note that our estimation sample excludes the four main urban areas: so even outside of urban areas, it is the relatively richer rural areas that are more likely to receive early and any cellular network coverage. Higher population density is also associated with greater coverage, and this remains true even through 2008.[15]

Adding Geographic Factors In Table 3 we explore the effect of geographic factors on coverage, controlling for demographics, again at the EA level. The structure of Table 3 mimics Table 2: in Panel A the dependent variable is a binary for coverage, in Panel B it is the continuous coverage measure. We are interested in the coefficients on these cost-side variables as well as how the coefficients for the demand-side variables change between Tables 2 and 3.

A number of interesting points emerge from this table. First, even when controlling for demographic variables, EAs with smaller values of slope (that is to say, areas with flatter terrain) have a significantly higher probability of receiving cellular coverage in early years while areas of higher altitude are significantly less likely to receive cellular coverage in the later years. Looking at the interaction term, EAs with steeper slope *and* at higher altitudes also have a higher probability of getting coverage in the early years. This result has a fairly natural interpretation that it may only make sense to locate cellular towers in EAs with steep slopes when there is a high point to locate the tower. If population density is very low in areas with high values of slope, then having a tower at a higher point with completely uninterrupted line of sight may be required to reach more individuals, conditional on this low density.

Second, conditional on geographic factors, the relationship between most of the demographic factors (percent of the workforce in agriculture, average education and average age) and cellular coverage

[15] When we estimate these regressions including total population as well as population density, the coefficient on log population density does not move much at all; the coefficient on total population is negative and significant.

remains about the same, or becomes even stronger in the early and middle periods. However, once we control for slope, altitude, the interaction of the two and mean distance to the nearest road, the relationship between population density and cellular coverage is weakened and in the last year, with the binary measure of coverage, is no longer statistically different from zero. This possibly reflects the fact that a large part of the relationship between geographic factors and cellular coverage indirectly captures the relationship between population density and cellular coverage.

We noted at the start of the paper that one of the motivations for estimating the relationship between demand and cost side factors and cell network expansion was to see how large a role these factors play in accounting for the pattern of expansion. Our regressions suggest that, with an EA, these demand and cost factors account for about 50% of the variation in coverage. This suggests that although there is some possible role for other variables (e.g., political factors or the availability or lack of complementary infrastructure), likely they do not account for the majority of the variation.

Hazard Models In Table 4 we move to estimating Cox proportional hazard models where the outcome is coverage. The model we estimate is:

$$\lambda_c(t|G_c,\beta) = \lambda_0(t)\exp\,(G_c{}'\beta) \qquad (2)$$

where, $\lambda_c(t|G_c,\beta)$ is the EA specific hazard rate, $\lambda_0(t)$ is the baseline hazard rate, t is the year in which the EA obtains cellular coverage and G_c is the set of EA-specific demographic and geographic variables. The Cox proportional hazards model allows us to estimate the relationships between geographic and demographic variables semi-parametrically. This model does not make assumptions about the form of $\lambda_0(t)$ (which is unidentified here) but does assume that time-invariant repressors G_c shift the hazard rate around multiplicatively (Cameron and Trivedi, 2005). The advantage of these models over the OLS models is they provide a more complete picture of which areas get coverage faster, without relying on inference from multiple regressions as in Tables 2 and 3. The hazard models use all of the time variation in rollout between 1998 and 2008, not just the three years which we analyze in the OLS analysis.

The evidence in Table 4 largely echoes what we have found already. In this case, the three columns represent three different cutoff values for coverage. In Column 1, the area is defined as covered in the first year of any cell phone coverage. In Column 2, it is defined as covered in the first year that our coverage measure exceeds 0.5 (recall this does not necessarily mean that 50% of the area is covered; it could be that 25% of the area is covered by 2 towers. In Column 3, it is defined as covered in the first year our coverage measure exceeds 1.0.

Regardless of which coverage definition we use, we find that the demand-side factors matter as expected (Panel A). More dense areas have faster coverage, as do areas with a higher fraction of educated people; more agricultural areas have slower coverage. When we add in geographic factors (Panel B), we find some evidence they matter, but it is fairly weak. Higher altitude areas seem to have slower coverage, and the interaction between altitude and slope matters in some specifications. Distance to a road is more consistently important. In these models, the demographic factors continue to be strongly influential even when we include geographic variables.

Together, the demographic and geographic results suggest that both demand-side and cost-side factors drive cellular phone coverage. Even though Malawi is one of the poorest countries in the world, we see coverage expanding into areas where it seems likely there is higher potential market demand – i.e., those areas with more income and more potential users. This is in line with what each company's marketing unit highlighted as one of the key factors guiding rollout: market potential. In addition, controlling for these demand-side factors we see evidence of more coverage in areas that appear to be easier to reach and build on – i.e., those with a less severe slope and areas that are less remote. We see similar patterns in rollout of other types of infrastructure (for example, television and electricity) in both the developing and developed world (Dinkelman, 2010; Gentzkow and Shapiro, 2008; Jensen and Oster, 2009).

V. Results: Correlates of Network Performance: Dropped Call Rates

In our final piece of analysis, we describe variation in traffic congestion rates across EAs, within a TA, for each of the two network providers. Table 5 presents estimates from OLS regressions of the form in equation (1) above where the outcome variable is the percent of dropped calls for each company (defined above) in the month of August 2008. In addition to the prior set of controls, we also control cell phone network coverage within the EA in 2007. For the samples of interest, mean network coverage for Company 1 in 2007 is 0.72 while mean network coverage for Company 2 in 2007 is 0.70.

From this analysis, we learn that population density in 1998 is an important predictor of the percent dropped calls for both companies, and these relationships are of about the same magnitude. This negative relationship is not obvious *a priori:* with more people using a given network, higher demand pressure could result in a higher dropped call rate. What we are seeing in Malawi appears to be the reverse: that in places with higher actual demand (as proxied by higher population density in 1998), the performance of the network is better. This suggests that the cell phone providers are responding to high demand areas by increasing their coverage quality. In fact, they are *more than* responding to the increased demand such that coverage is actually better, not just the same, in high demand areas. This is similar to what we see in the developed world: cell phone coverage is typically at its best in large urban areas, and worse in more isolated rural areas, despite the lower demands on the network.

It is interesting to note that, in contrast to the analysis of rollout, network congestion as measured by the dropped call rate is not sensitive to most of the geographic variables that we use to proxy for cost of actually building the network.

Finally, prior network coverage in 2007 in the EA is also a strong predictor of the dropped call rate. Better, more dense coverage by Company 1's network reduces the dropped call rate for Company 1 users, and more dense coverage by Company 2's network reduces the dropped call rate for Company 2 users.

VI. Conclusion

This paper makes use of new, very detailed, data on the history and location of each cellular phone tower constructed in Malawi matched with Census data, geographic data and information on network traffic. The goals of this paper are two-fold. First, we provide detailed evidence on cellular phone rollout in a very low-income context. We show that despite the fact that Malawi is extremely poor, cellular rollout occurs rapidly and within 10 years, over half of the country has access to at least one network. This stands in stark contrast to the backlog in access to roads and electricity in this largely rural country. Part of the success of the cell phone rollout may be due to licensing requirement that cell companies provide access to specific areas by predetermined target dates. However, our evidence shows that both cost-side factors and demand-side market potential variables are important for the timing of initial coverage; cell companies are not simply providing coverage in response to MACRA-mandated target areas. This suggests that even in a setting in which the government may struggle to fairly provide public goods, the timing of infrastructure provision and the quality of the services provided for cell phone networks seem to be driven by market-oriented factors.

References

Aker, Jenny C. 2009. "Does Digital Divide or Provide? The Impact of Cell Phones on Grain Markets in Niger." *BREAD Working Paper*.

Aker, Jenny C. and Isaac M. Mbiti. 2010. "Africa calling: Can mobiles make a miracle?" *Boston Review*, March/April.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics*, 114 1243-1284.

Arnquist, Sarah. 2009. "In rural Africa, a fertile market for cellphones", *The New York Times*, October 5 http://www.nytimes.com/2009/10/06/science/06uganda.html

Balakrishnan, Angela. 2008. "Mobile phone technology vital to growth in the developing world." *The Guardian (London)*, February 7, No. 32.

Banerjee, Abhijit, and Rohini Somanathan. 2007. "The Political Economy of Public Goods: Some Evidence from India." *Journal of Development Economics*, 82(2), 287-314.

Benson, T., J. Kaphuka, S. Kayanda and R. Chinula. 2002. Malawi: An atlas of social statistics. National Statistics Office Malawi.

Besley, Timothy, Rohini Pande, Lupin Rahman, and Vijayendra Rao. 2004. "The Politics of Public Good Provision: Evidence from Indian Local Governments." *Journal of the European Economic Association*, 2, 416-426.

Buys, Piet, Susmita Dasgupta, Timothy S. Thomas, and David Wheeler. 2009. "Determinants of a Digital Divide in Sub-Saharan Africa: A Spatial Econometric Analysis of Cellular Phone Coverage." *World Development*, 37(9), 1494-1505.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. Microeconometrics: Methods and applications, Cambridge University Press.

CIA, 2009. The World Factbook. CIA Publication. Accessed at: https://www.cia.gov/library/publications/the-world-factbook/index.html.

Dinkelman, Taryn. 2010. "The effects of rural electrification on employment: New evidence from South Africa." *RPDS Working Paper*.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics*, 112(4), 1203-1250.

Gentzkow, Matthew, and Jesse Shapiro. 2008. "Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study." *Quarterly Journal of Economics*. 123(1), 279-323.

Godlonton, Susan and Rebecca Thornton. 2010. "Peer Effects in Learning HIV Results." Working Paper, University of Michigan

Hausman, Jerry. 2010. “Mobile phones in developing countries.” Department of Economics, MIT, Mimeo.

International Telecommunications Union. 2009. “World Telecommunications Indicators/ICT database” http://www.itu.int/ITU-D/ict/publications/world/world.html

Jack, William, and Tavneet Suri. 2009. “Mobile money: The economics of M-Pesa.” MIT Sloan, Mimeo.

Jensen, Robert. 2007. “The Digital Provide: Information (Technology), Market Performance and Welfare in the South Indian Fisheries Sector.” *Quarterly Journal of Economics*, 122(3), 879-924.

Jensen, Robert, and Emily Oster. 2009. “The Power of TV: Cable Television and Women's Status in India.” *Quarterly Journal of Economics*. 124(3), 1057-1094.

Khwaja, Asim Ijaz. 2009. “Can Good Projects Succeed in Bad Communities?” *Journal of Public Economics* 93.7-8 (August): 899-916.

Kimenyi, Mwangi S. 2006. “Ethnicity, Governance and the Provision of Public Goods.” *Journal of African Economies*, 15, 62-99.

Klonner, Stefan, and Patrick Nolen. 2008. “Does ICT benefit the poor? Evidence from South Africa.” Department of Essex, UK, Working Paper.

Krudy, John P. 2009. “Africa Rings Up Trade Benefits.” *The Washington Times*, June 22, B5.

Manjoo, Farhad. 2010. “How often does your phone drop calls?” *Slate*, July 22.

Mauro, Paolo. 1995. “Corruption and Growth.” *Quarterly Journal of Economics*, 110(3), 681-712.

McGreal, Chris. 2009. “Front: From Congo to Kathmandu, how mobiles have transformed the world.” *The Guardian (London)*, March 3.

National Statistics Office. 2004. “Malawi Integrated Household Survey 2004”

Ngowi, Rodrique. 2005. “Africa’s cellphone explosion changes economics, society.” *USA Today*, October 16.

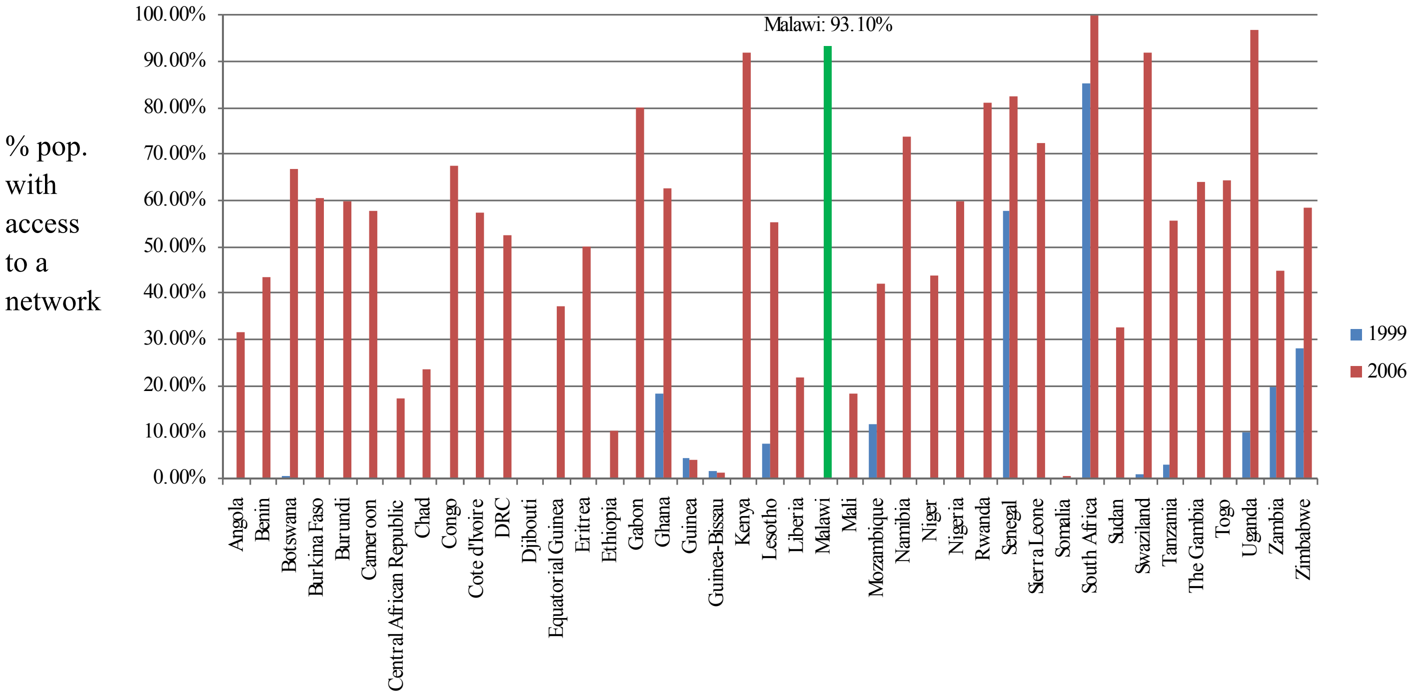
OECD. 2009. “OECD Communications Outlook 2009.” *OECD Publication*, August 11.

Paul Budde Communication Pty Ltd. 2009. “Global Mobile Communications - Statistics, Trends & Forecasts”, http://www.budde.com.au/Research/2009-Global-Mobile-Communications-Statistics-Trends-Forecasts.html

Tanzi, Vito, and Hamid Davoodi. 1998. “Roads to Nowhere: How Corruption in Public Investment Hurts Growth.”*Economic Issues* 12. International Monetary Fund, Washington, D.C.

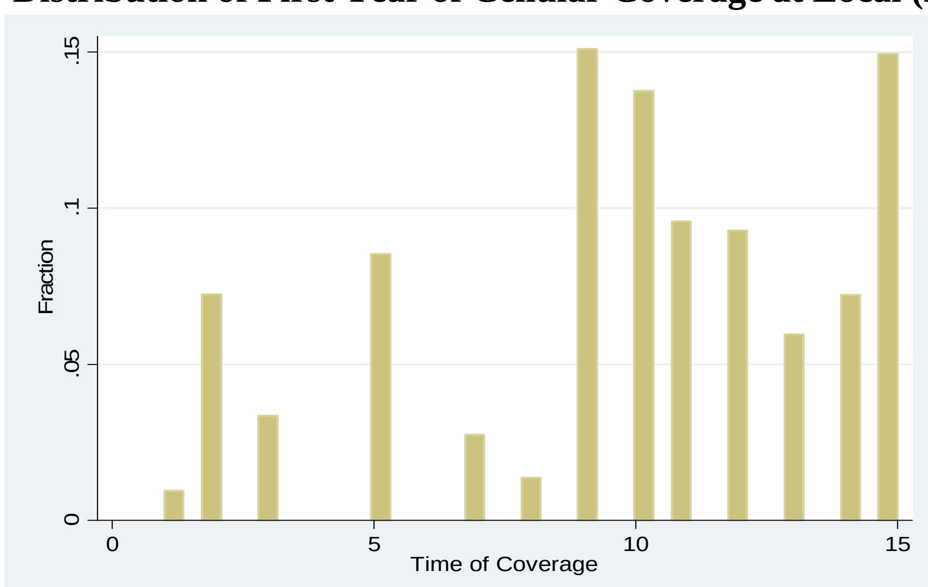
Vail, Leroy, editor. 1989. The Creation of Tribalism in Southern Africa. London Berkeley: Currey University of California Press. http://ark.cdlib.org/ark:/13030/ft158004rs.

Figure 1: Cellular Phone Coverage in Africa



Source: Buys et al. (2009)

Figure 2: Distribution of First Year of Cellular Coverage at Local (EA) Level



Range of years is from 1995 (year 1) to 2009 (year 15).

Figure 3: Map of Cellular Phone Coverage in 1997 (TNM Only)

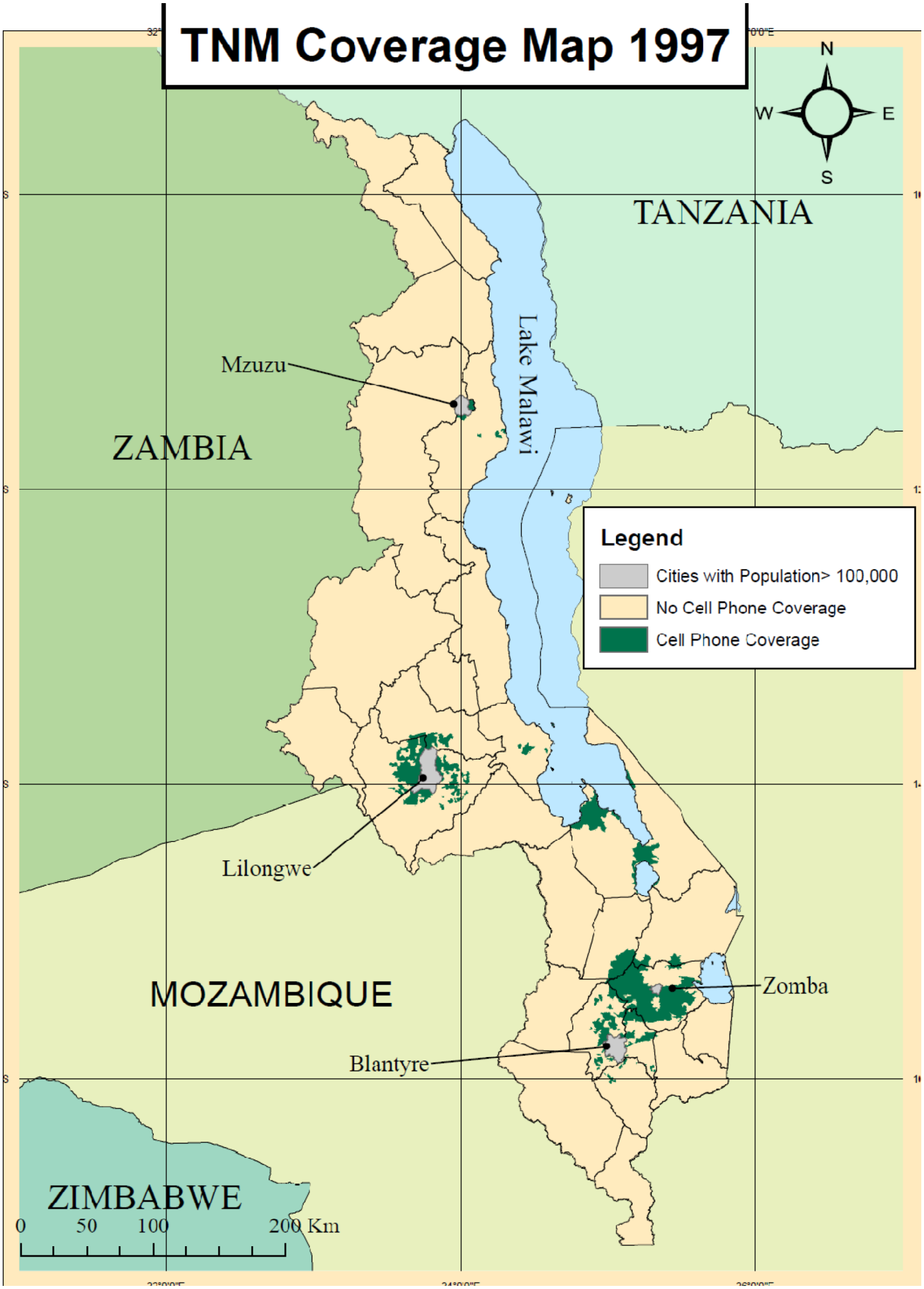


Figure 4: Map of Cellular Phone Coverage in 2004 (Both Networks Active)

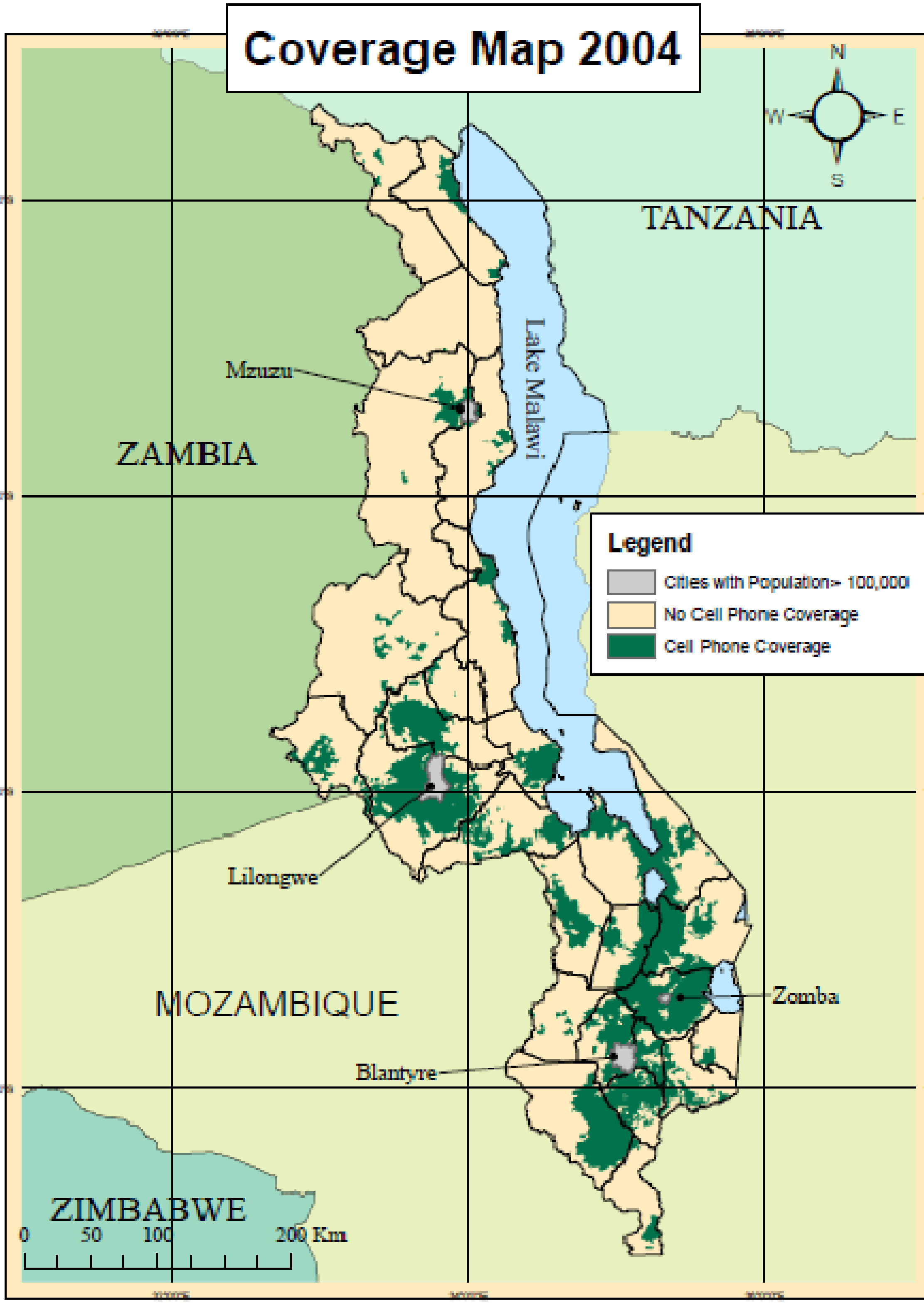


Figure 5: Map of Cellular Phone Coverage in 2008 (Both Networks Active)

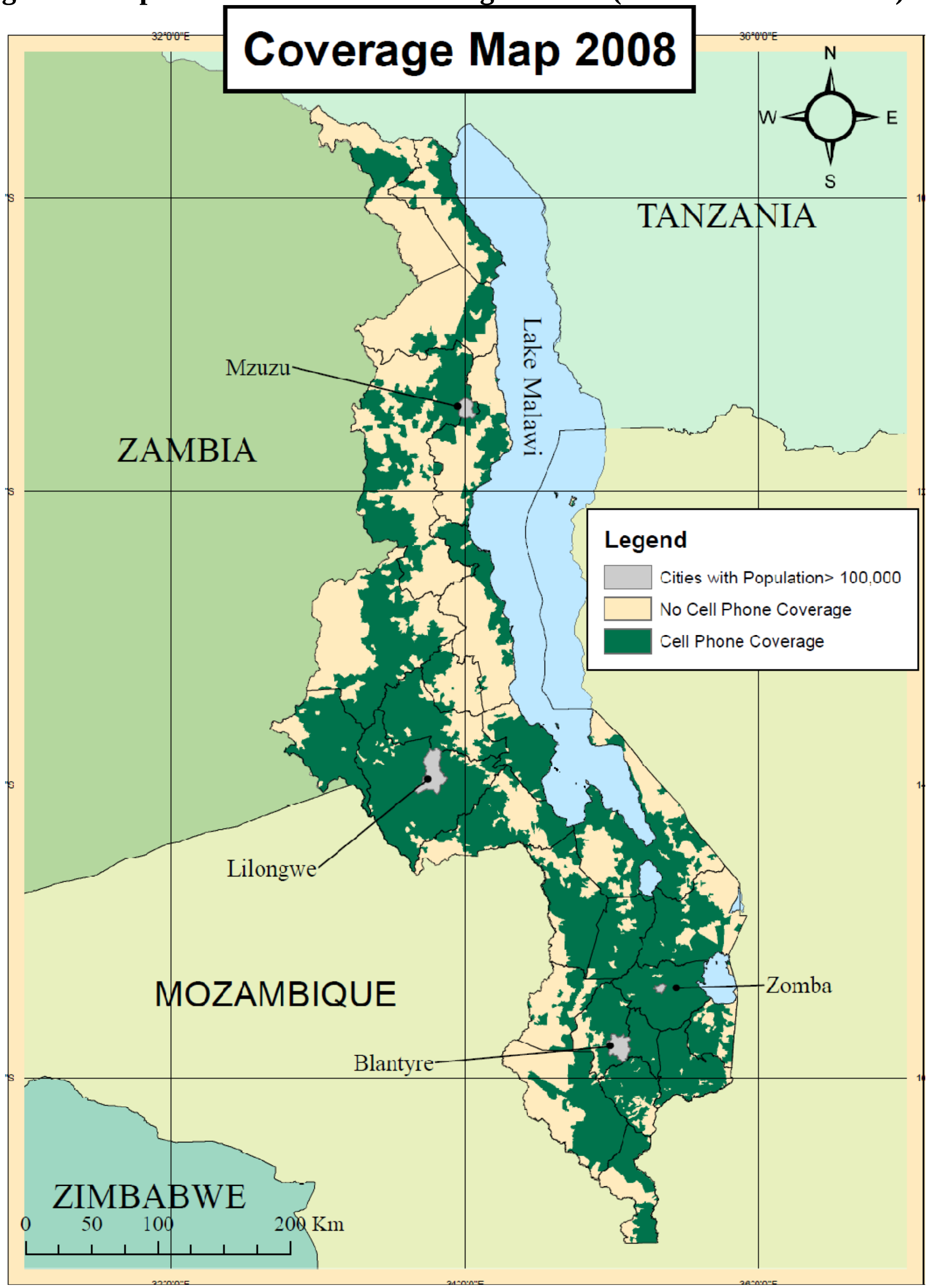


Figure 6: Arrival of Cellular Coverage at the Local Level

Figure 6.1 By 1998 Population Density (1=sparse, 4=dense)

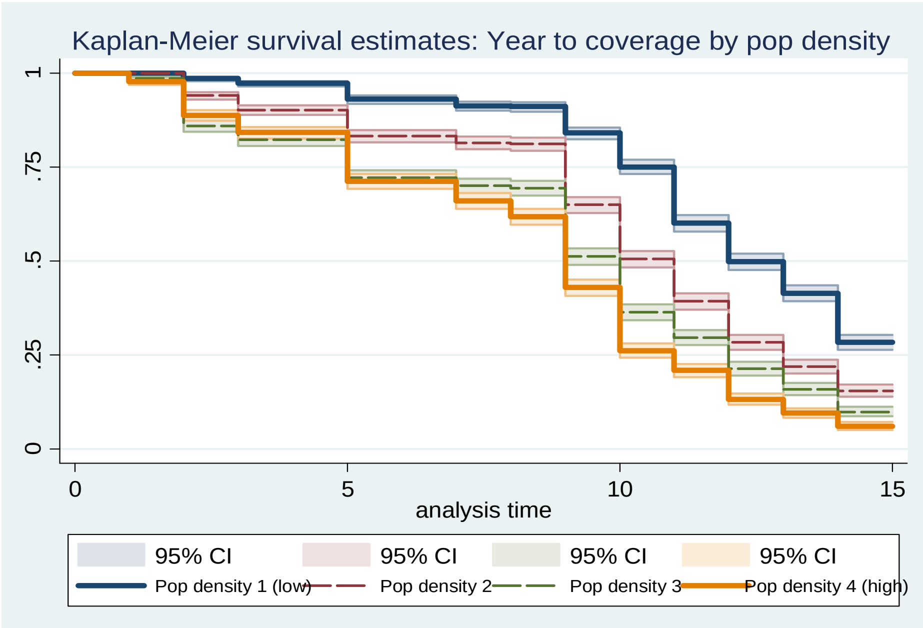


Figure 6.2 By Slope Group (1=flat, 4=steep)

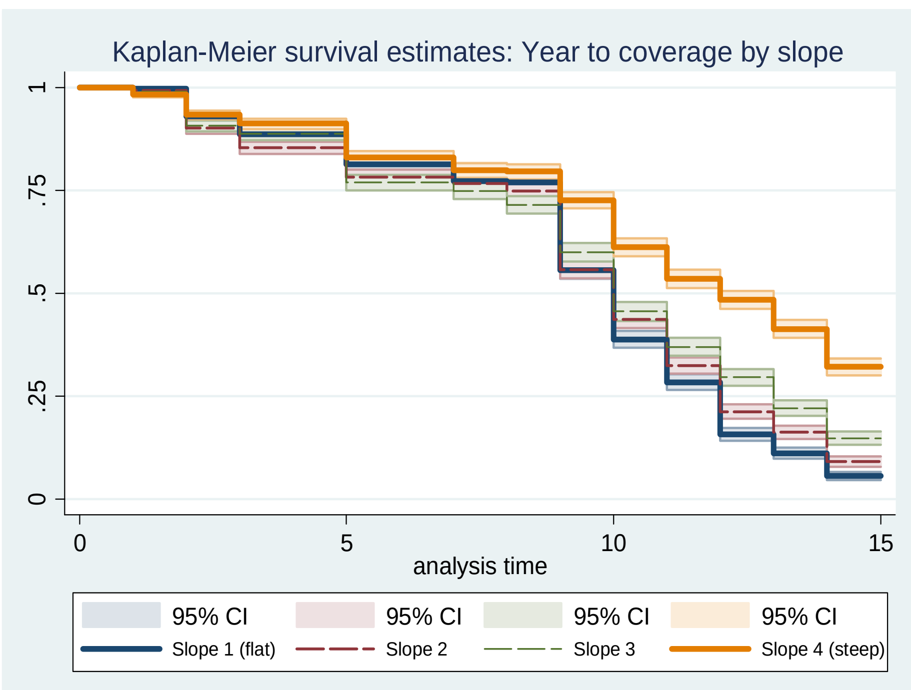
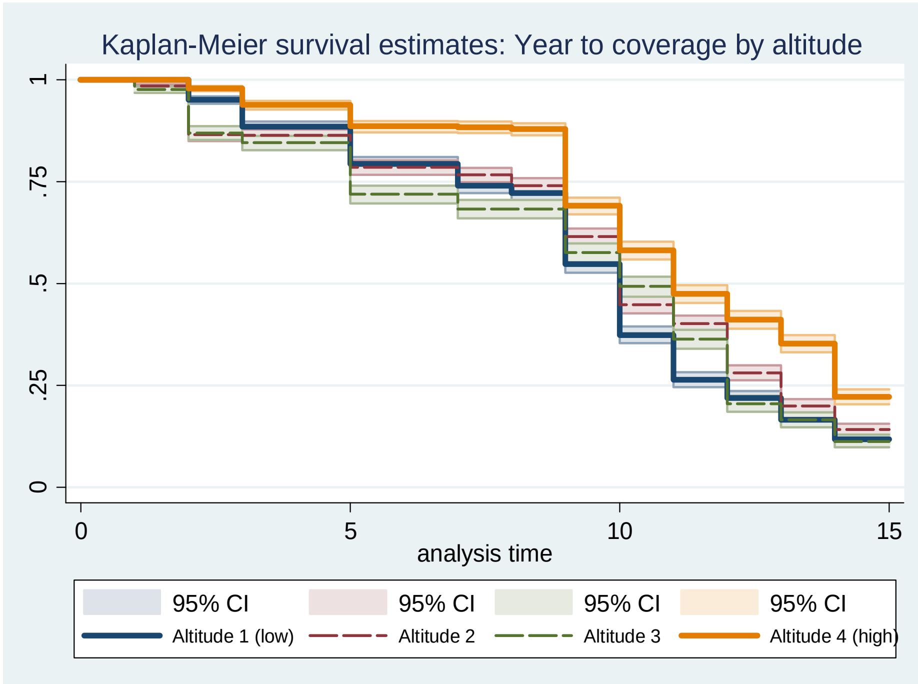


Figure 6.3 By Altitude Group (1=low, 4=high)



Unit of observation is the Enumeration Area (EA). Graphs show Kaplan-Meier survival functions for EAs with different values of slope, altitude and population density, where exit is into the "covered" state. Range of years is from 1995 (year 1) to 2009 (year 15). Observations without coverage by 2009 are assigned a year value of 15. Sample is restricted to rural areas with positive population in 1998 Census.

Table 1: Summary Statistics for Enumeration Areas (EA)

Panel A: Cellular Phone Coverage Over Time

	Fraction of EA covered		
	All EAs	Rural EAs	Urban EAs
1999	0.27	0.20	0.97
2000	0.27	0.20	0.97
2001	0.30	0.23	0.97
2002	0.31	0.24	0.97
2003	0.45	0.39	0.99
2004	0.57	0.53	0.99
2005	0.66	0.63	1.00
2006	0.74	0.72	1.00
2007	0.80	0.78	1.00
2008	0.86	0.85	1.00
N	8,924	8,118	806

Panel B: Summary statistics for traffic congestion data: Rural EAs with cell coverage, by network

	Mean	S. d.	Num. obs.	Min	Max
Fraction dropped calls: Company 1, Aug 2008	0.020	0.010	5,971	0.002	0.050
Fraction dropped calls: Company 2, Aug 2008	0.009	0.003	8,046	0.003	0.023

Panel C: Summary statistics for geographic data: Rural EAs only

	Mean	S. d.	Num. obs.	Min	Max
Altitude (meters above sea level)	848.52	342.23	8,118	34	1,988
Slope (% rise)	3.23	3.14	8,118	0	24
Distance to road (km)	1.45	2.17	8,118	0	62

Panel D: Summary statistics of demographic variables from 1998 Census: Rural EAs only

	Mean	S. d.	Num. obs.	Min	Max
Age	22.16	1.59	8,118	16	40
Education (yrs)	2.97	1.22	8,118	0	8
Percent adults married	0.55	0.07	8,118	0	1
Percent adults in agriculture	0.89	0.17	8,118	0	1
Population	1,054.01	372.46	8,118	3	3,661
Population density (pop/km^2)	313.84	770.43	8,118	0.03	21,142

Notes: This table presents summary statistics for the primary variables used in the rollout analysis. In all panels, the unit of observation is the enumeration area (EA). Summary measures are calculated over EAs with positive population density. Panel B, C and D summary statistics are restricted to rural area EAs. Cellular coverage is a combined TNM and Zain coverage measure, constructed as described in the text.

Table 2: Cellular Phone Coverage and Demographic Proxies for Market Demand

	OLS 2000	OLS 2004	OLS 2008
Panel A: Binary Coverage			
	(1)	(2)	(3)
Log EA population density	0.020***	0.048***	0.027***
	(0.00)	(0.01)	(0.00)
Percent adults in agriculture	-0.033	-0.136**	-0.020
	(0.03)	(0.04)	(0.03)
Education (yrs)	0.018***	0.029***	0.036***
	(0.00)	(0.01)	(0.00)
Age	-0.001	0.016***	0.012***
	(0.00)	(0.00)	(0.00)
R-squared	0.558	0.475	0.367
N	8,118	8,118	8,118
Mean of dependent variable	0.20	0.53	0.85
Panel B: Continuous Measure of Coverage			
	OLS 2000	OLS 2004	OLS 2008
	(1)	(2)	(3)
Log EA population density	0.037***	0.104***	0.259***
	(0.01)	(0.01)	(0.03)
Percent adults in agriculture	-0.303***	-0.688***	-1.575***
	(0.08)	(0.13)	(0.36)
Education (yrs)	0.038***	0.087***	0.165***
	(0.01)	(0.02)	(0.04)
Age	0.001	0.016**	0.034*
	(0.00)	(0.01)	(0.02)
R-squared	0.527	0.556	0.518
N	8118	8118	8118
Mean of dependent variable	0.20	0.53	0.85

Notes: This table shows the relationship between cellular phone coverage in different years and market demand variables measured in the 1998 Census. The unit of observation is the EA, sample is restricted to rural EAs with positive population according to 1998 Census data. All regressions contain TA fixed effects and a constant. Standard errors in parentheses are clustered at the TA level. Binary coverage is defined as an indicator of whether more than 50% of the area is covered. Continuous measure of coverage indicates the number of towers covering each area. *p<0.1, **p<0.05, ***p<0.01

Table 3: Cellular Phone Coverage, Demographic Proxies for Market Demand, and Geographic Variables

Panel A: Binary Coverage			
	OLS 2000	OLS 2004	OLS 2008
	(1)	(2)	(3)
Log EA population density	0.010*	0.027***	0.000
	(0.00)	(0.01)	(0.00)
Percent adults in agriculture	-0.044	-0.161***	-0.040
	(0.03)	(0.04)	(0.03)
Education (yrs)	0.013**	0.025***	0.031***
	(0.00)	(0.01)	(0.00)
Age	-0.002	0.012***	0.006*
	(0.00)	(0.00)	(0.00)
Log Altitude	0.020	-0.004	-0.101***
	(0.02)	(0.03)	(0.03)
Log Slope	-0.138***	-0.216**	-0.086
	(0.04)	(0.07)	(0.06)
Log Altitude * Log Slope	0.018**	0.021*	0.004
	(0.01)	(0.01)	(0.01)
Distance to road (km)	-0.014***	-0.018***	-0.025***
	(0.00)	(0.00)	(0.00)
R-squared	0.560	0.484	0.390
N	8,115	8,115	8,115
Mean of dependent variable	0.20	0.53	0.85
Panel B: Continuous Measure of Coverage			
	OLS 2000	OLS 2004	OLS 2008
	(1)	(2)	(3)
Log EA population density	0.024***	0.064***	0.158***
	(0.01)	(0.01)	(0.04)
Percent adults in agriculture	-0.329***	-0.749***	-1.717***
	(0.08)	(0.13)	(0.36)
Education (yrs)	0.029***	0.073***	0.126***
	(0.01)	(0.02)	(0.04)
Age	0.00	0.008	0.016
	(0.00)	(0.01)	(0.02)
Log Altitude	0.123***	0.059	0.183***
	(0.03)	(0.06)	(0.12)
Log Slope	-0.355***	-1.151***	-2.197***
	(0.07)	(0.16)	(0.32)
Log Altitude * Log Slope	0.047***	0.152***	0.284***
	(0.01)	(0.02)	(0.05)
Distance to road (km)	-0.019***	-0.037***	-0.108*
	(0.00)	(0.01)	(0.02)
R-squared	0.514	0.549	0.507
N	8,115	8,115	8,115
Mean of dependent variable	0.20	0.53	0.85

Notes: This table shows the relationship between cellular phone coverage in different years, market demand variables and geographic features. The unit of observation is the EA, sample is restricted to rural EAs with positive population according to 1998 Census data. All regressions and contain TA fixed effects and a constant term. Standard errors in parentheses are clustered at the TA level. Binary coverage is defined as an indicator of whether more than 50% of the area is covered. Continuous measure of coverage indicates the number of towers covering each area. *p<0.1, **p<0.05, ***p<0.01

Table 4: Cox Regressions: Hazard Out of No Cellular Coverage

Panel A: Demand Covariates Only

	(1)	(2)	(3)
Coverage cutoffs	0	0.5	1
Log EA population density	0.132***	0.280***	0.355***
	(0.01)	(0.01)	(0.01)
Percent adults in agriculture	-0.541***	-0.491***	-0.358***
	(0.09)	(0.10)	(0.10)
Education (yrs)	0.002	0.066***	0.070***
	(0.01)	(0.01)	(0.01)
Age	0.079***	0.099***	0.107***
	(0.01)	(0.01)	(0.01)
N	8,118	8,118	8,118

Panel B: Demand and Geographic Covariates

	(1)	(2)	(3)
Coverage cutoffs	0	0.5	1
Log EA population density	0.114***	0.210***	0.272***
	(0.01)	(0.01)	(0.02)
Percent adults in agriculture	-0.574***	-0.578***	-0.457***
	(0.09)	(0.10)	(0.10)
Education (yrs)	0.001	0.086***	0.094***
	(0.01)	(0.01)	(0.01)
Age	0.077***	0.093***	0.106***
	(0.01)	(0.01)	(0.01)
Log Altitude	-0.015	-0.054***	-0.031
	(0.02)	(0.02)	(0.02)
Log Slope	0.267*	-0.051	0.058
	(0.14)	(0.17)	(0.19)
Log Altitude * Log Slope	-0.039*	-0.032	-0.057**
	(0.02)	(0.03)	(0.03)
Distance to road (km)	-0.027***	-0.058***	-0.088***
	(0.01)	(0.01)	(0.01)
N	8,115	8,115	8,115

Notes: This table shows the relationship between cellular phone coverage and market demand variables measured in 1998 Census and geographic features. The unit of observation is the EA, sample is restricted to rural EAs with positive population according to 1998 Census data. In Column 1, the area is defined as covered in the first year of any cell phone coverage. In Column 2, it is defined as covered in the first year that our coverage measure exceeds 0.5. In Column 3, it is defined as covered in the first year our coverage measure exceeds 1.0. *p<0.1, **p<0.05, ***p<0.01

Table 5: Network Performance Measured as Percent Dropped Calls in August 2008

	OLS regressions: Company 1			OLS regressions: Company 2		
	(1)	(2)	(3)	(4)	(5)	(6)
Log EA population density	-0.0120*	-0.0132*	-0.0119*	-0.0128***	-0.0131***	-0.0125***
	(0.006)	(0.007)	(0.007)	(0.002)	(0.002)	(0.002)
Percent adults in agriculture	-0.192***	-0.200***	-0.216***	0.021	0.017	0.011
	(0.055)	(0.054)	(0.055)	(0.014)	(0.014)	(0.014)
Education (yrs)	-0.0265***	-0.0231***	-0.0218***	-0.003	-0.003	-0.002
	(0.007)	(0.007)	(0.007)	(0.002)	(0.002)	(0.002)
Age	0.006	0.005	0.005	-0.00467***	-0.00476***	-0.00472***
	(0.004)	(0.005)	(0.005)	(0.001)	(0.001)	(0.001)
Log Altitude		0.000	0.000		0.0207**	0.0206**
		(0.053)	(0.053)		(0.008)	(0.008)
Log Slope		-0.0534	-0.103		-0.012	-0.0172
		(0.143)	(0.143)		(0.022)	(0.022)
Log Altitude * Log Slope		0.003	0.010		0.000	0.001
		(0.022)	(0.022)		(0.003)	(0.003)
Mean Distance to road (km)		0.009	0.008		0.002	0.002
		(0.006)	(0.006)		(0.002)	(0.002)
Company 1 coverage in 2007			-0.0211***			-0.003
			(0.006)			(0.002)
Company 2 coverage in 2007			0.006			-0.00692*
			(0.012)			(0.004)
N	5,971	5,968	5,968	8,046	8,043	8,043
Mean of outcome	2.03	2.03	2.03	0.93	0.93	0.93
R-squared	0.850	0.850	0.851	0.827	0.827	0.828
F stat for demand side variables	5.25	4.43	4.52	16.61	14.23	12.03
F stat for geographic variables		3.04	3.60		3.48	3.69
F stat for prior coverage variables			22.85			19.32

Notes: This table shows the relationship between network performance for both of the network providers in August 2008, and a set of variables proxying for market demand (from 1998 Census data), cost factors and prior network coverage. The unit of observation is the EA, sample is restricted to rural EAs with positive population according to Census data. All regressions contain TA fixed effects and a constant term. Standard errors in parentheses are clustered at the TA level, *p<0.1, **p<0.05, ***p<0.01