New cellular networks in Malawi: Correlates of service rollout and effects on employment

Dmitris Batzilis, Taryn Dinkelman, Emily Oster, Rebecca Thornton, Deric Zanera[1]

Draft: June 30, 2010

[PRELIMINARY, DO NOT CITE]

Cellular technologies have become increasingly important in the developing world; mobile networks have expanded dramatically over the past two decades giving access to remote areas without previous phone access. Despite this expansion, relatively little is known about the correlates of cellular phone coverage across areas within a country or the impacts of this new coverage. In this paper we examine the rollout of cellular phones and address the effects of phone coverage on labor market outcomes using new data from Malawi. First, we compile a new data set combining administrative, geographic and Census data at a local level and use this to describe the cellular phone network rollout in rural areas of Malawi. We highlight the relative importance of market demand and cost factors at different points in the rollout history. Second, we merge individual-level data from the Demographic Health Survey data to the rollout data and show that areas gaining new access to the cellular network also experience significant changes in labor market outcomes. Using a difference-in-differences analysis, we show that female labor market participation increases by between 8.3 and 11.5 percentage points in areas gaining access to the network, relative to areas in which coverage is unchanging. This network expansion is correlated with large increases in women working away from home.

[1] Batzilis: Batzilis@gmail.com; Dinkelman: tdinkelm@princeton.edu; Oster: Emily.Oster@chicagobooth.edu; Thornton: rebeccal@umich.edu. Research for this paper was funded by the NBER Program on African Successes. We are grateful for assistance from Eirin Forsund and Ajay Shenoy. 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.

I. Introduction

Cellular phone technologies have become increasingly important throughout the developing world and in Africa in particular. Figure 1 shows growth in the physical area 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, Dasgupta, Thomas, and Wheeler 2008). 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 areas with low population density. As a result, cellular network technology may be the best way to bring telecommunications to most of Africa.

Popular media, and a growing body of research within economics, suggests that cellular technologies may be enormously important for improving productive efficiency, for providing a cheap method of transferring information (thereby generating efficient markets) 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 which often have trouble providing even very basic public goods. Moreover, although there is some very good evidence on the impact of cellular networks on the price of traded goods (i.e. Jensen, 2007; Aker, 2009), we have little rigorous evidence on how cellular phone technologies might affect other economic activities.[3] One particular activity of interest – especially in developing countries – is labor force participation.

In this paper we take a first step towards addressing each of these issues using data from Malawi. Although the two topics are broadly related, we approach them fairly independently, beginning with estimating the correlates of cellular phone coverage and then moving to the impacts on employment. One of the aims of our paper is to highlight the research potential in combining familiar household survey data with less familiar sources of information, obtained from administrative and engineering records. We use this unique linked dataset to describe some of the important market demand and supply-side correlates of

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

[3] New work by Hausman (2010) uses an expenditure function approach to estimate large, significant consumer welfare gains associated with the spread of mobile technology (a “new good”) in India and a set of less developed countries. Klonner and Nolen (2008) look at the impact of cellular phone rollout on employment in South Africa using an instrumental variables approach. Aker and Mbiti (2010), recognizing the potential that telecommunications technology has in poorly resourced countries, actively call for more rigorous studies of how cellular phone networks affect important social and economic outcomes.

rollout of the cellular network in Malawi and, using a difference-in-differences strategy, analyze the reduced form effect of new access to the network on employment outcomes between 2000 and 2004.

In the first part of the paper, we describe the determinants of cellular phone coverage across areas within Malawi using a newly constructed dataset. Cellular phone access in a particular area is determined by proximity to cellular phone towers. We collected detailed data from both Malawian cellular phone companies on tower locations (GIS) and date of construction; our data cover every tower in the country. Using this information, we construct coverage maps for the country over time. We focus on cellular phone access within two types of geographic areas: a larger unit called a Traditional Authority (TAs, of which there are 240 in Malawi) and smaller Enumeration Areas (EAs. of which there are about 9200).

The first cellular towers in Malawi were built in the mid-1990s, and coverage is initially focused around the major cities (i.e. Lilongwe and Blantyre) and tourist areas (Salima and Mangochi). However, increase in coverage is rapid: by 2004, 57% of the land area of the country has access, and by 2008 this is up to 86%. We use demographic data from the 1998 Malawian census to proxy for potential market demand; it is likely that the companies themselves used similar demographics in deciding where to locate towers. In our analysis there are a number of proxies for area income (which we do not observe directly) that correlate with the timing of access. Areas with higher population density receive coverage earlier. Higher levels of employment in agriculture correlate with delayed access, and higher levels of education in 1998 correlate with faster access in the subsequent 10 years. These correlates correspond with demand-side factors that engineers and other cellular phone industry representatives report as important for cellular phone rollout.

In addition to these variables capturing market potential, we estimate the correlation between 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 altitude are much less likely to get access in the later part of our sample period, whereas those with higher slope tend to have access retarded in the early periods. Areas that are far from a road are significantly less likely to get access to the network throughout the period.

In the second part of the paper we turn to examining the relationship between cellular phone coverage and labor force participation. We use our coverage data along with two waves of a nationally representative household survey to estimate the impact of the introduction of cellular phones into an area on the labor force participation of men and women. A lack of information about opportunities for work and about returns to education has been highlighted as an important constraint to the efficient operation of labor markets and to optimal investment in education in developing countries (e.g. Rosenzweig 1988; Jensen 2010). Since new access to the cellular phone network presents new access to information, we expect that the rollout could have effects on the level and price of market traded services, i.e. labor. In addition, we hypothesize that access to cellular phones makes it easier for individuals to learn about job opportunities (job search) and to work away from home while still remaining in touch with the family.

To investigate this, we link our data on cellular phone coverage to the Demographic and Health Survey (DHS) data, which we aggregate to the TA level. We take advantage of the fact that we observe two rounds of the DHS in Malawi in 2000 and 2004. Using these data, we estimate changes in labor force participation over time in for three groups: areas which have cellular coverage during both periods; those that get coverage between the two periods; and those which never have it. In this standard difference-in-difference analysis we measure changes in labor market outcomes that occur in areas with changes in phone coverage, and compare these changes to the changes in areas with consistent phone coverage (either always covered or never covered).

We find evidence that cellular phone coverage is correlated with large increases in female employment as well as changes in the type of employment for women and men. When we focus on women's work overall, we find a large (8.7 to 11.5 percentage point) and significant increase in labor force participation in areas with increased access to cellular phone coverage. There is also a significant 9.1 percentage point increase in the fraction of women working away from home. Some of this increase in female employment appears to be occurring in the agricultural sector, although our results are weaker here. These increases in employment on the extensive margin are not at all evident for men; although we do see a shift in men’s work towards skilled manufacturing, in TAs where cellular network access expanded.

Although our data are limited, we are able to do some preliminary tests for whether the relationship between changes in cellular phone coverage and employment is due to other positive economic trends. The big concern here is that the cellular companies are expanding their networks in to areas with positive market potential. To the extent this is the case, we would expect to see employment increasing in areas that get cellular coverage in the near future. We can test this by estimating the impact of the area getting cellular phone coverage soon after the 2004 survey.

For women, we do not see evidence that future coverage increases employment. If anything, there seems to be a marginally significant decrease in employment for women in areas which get coverage in the near future. However, for men we do see some evidence that employment, particularly when we measure working at all in the last year, increases in areas with future cell coverage. This suggests we should take the employment results with some caution, since it points to evidence that labor markets are changing in advance of cellular phones. A cleaner research design, perhaps with more years of data, would be helpful in more fully establishing these results.

We briefly examine how the employment changes for women differ by demographics. The effects on employment are fairly consistent across groups, although we do see some evidence that the impacts on working overall and working away from home are larger for women with children still at home. It may be that these women were most constrained prior to cellular phones by a need to stay close to home to be in contact with their families; or that children at home can substitute for time in home production, thereby allowing these women to take up new opportunities learned about with new access to cellular networks; however, we certainly do not have enough evidence to say this conclusively.

The rest of the paper is organized as follows. Section 2 presents some background information on cellular networks in Malawi. Section 3 describes our data. Section 4 shows our first set of results on patterns and correlates of cellular phone coverage. Section 5 presents our evidence on outcomes from the DHS, and Section 6 concludes.

II. Background

II.A. 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 $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 infant mortality rates are 83 per 1,000 (CIA, 2009)).

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 800km 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% percent of households use electricity for lighting.

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.

II.B. Cellular Telephones in Malawi

Overview of Providers In 1998, fewer than 6,000 individuals in Malawi (out of 4.2 million adults) had a cellular phone subscription, and less than 10% of households had access to a landline. Over the next 10 years, the development of a cellular network across the country has been extensive. The first cellular operator, TNM (Telekom Networks Malawi Limited), was majority-owned by the government of Malawi in early years. This operator was licensed by the government agency MACRA (Malawi Communications Regulation Authority) in 1995, but initial growth in the network 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 that gain new access to a cellular network in a particular year from 1995 to 2009 (we define "access" in more detail below). The figure shows clearly the initial slow coverage of EAs in early years; many more 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 operator, Celtel, was awarded a government license. Celtel was purchased in 2008 by Zain, headquartered in Kuwait and has a strong regional presence in over 20 other African countries. By late 1999 both operators were active in establishing their networks across the country, and by 2007, the number of cellular phone subscribers had risen to over 600,000 people, or 5% of the population. Many more than these subscribers have access to this telecommunications service, as most consumers use a cellphone on a pre-paid (non-subscription) basis. Currently, over 70% of the country has access to the cellular phone network and 70% of the current Malawian market is served by Zain.

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

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. Zain and TM were both granted licenses to provide cellular telephony services by MACRA, a government-appointed regulatory board. 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.

One important aspect of the license agreements was that each company was required to build cellular phone towers in certain target areas before specific deadlines. While it is not entirely clear how these targets were set (they were set after licenses were granted), 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 programme covering rural areas" (p2)[5] In discussions with engineers and managers of both companies in Malawi, it is clear that many of the target areas in rural parts of the country would not have been 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. cellular

Cellular Phone Pricing 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. Recent data from the International Telecommunications Union records cellular phone penetration in Malawi at 12 handsets per 100 individuals, compared to a world average penetration rate of 22 per 100 people (cited in Hausman 2010). Malawi's Integrated Household Survey of 2004 indicates that 17% of urban households and only 1% of rural households owned a cellular phone. 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 also high relative to other developing countries, although there is some indication that they have fallen over time. Prices of the two providers are a difficult to keep track of and standardize over time, since price setting responsibility is opaque, tariff changes sometimes occur

[5] 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.

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 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 compare prices over time, 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. Figure 3 plots the average price per second of this basket over time and shows that in the early part of the period, per second costs were 1.2 US cents for both companies and at the end of the period, this marginal cost falls by over 50%, to below 0.6 US cents per second (see notes to that figure for a description of where tariff data are from). Significant reductions in prices occurred for both companies between 2002 and 2004. This is still about 5 times higher than the per-second cost of using a cellular phone in the USA (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" (ringing and hanging up to signal a wealthier party to call back and pay for the call).

Lastly, individuals must have access to some means of charging their cellular phone handset. The majority of cellular phone users use electricity from Escom (government provided) services. However access to this infrastructure is limited and often not reliable which is why individuals in more remote areas may resort to charging cellular phones using car batteries.

Our analysis in this paper primarily focuses on the *availability* of the cellular phone network, 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 more ability to take advantage of the network. The results we find when examining the impact of the new network on economic outcomes will be similar to results in an *Intention to Treat* analysis and may therefore underestimate the effect of the treatment on the treated – that is, the effect of having a working cellular phone and availability of the network.

III. Data

This paper uses a number of new datasets, which we discuss in turn below. We have collated three types of data for the description of infrastructure rollout and the analysis of employment outcomes: (i) administrative data on the placement and timing of new cellular phone towers from both companies, (ii) geographic data describing Malawi's physical terrain and (iii) socioeconomic and demographic data from census and household surveys. We link these datasets at two different localities: Traditional Authorities (TA: larger, more aggregated) and Enumeration Areas (EA: smaller). EAs are nested in TAs and both are defined in more detail below.

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 1995-2009.

Using data on latitude and longitude, 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.[6]

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 30km of the point. The idea is straightforward: there can be no coverage if the tower signal is blocked by physical objects, and hence the tower is not "visible", or if the source of the signal is too far away. Discussions with engineers in Malawi led us to believe that 30km is an appropriate range for the towers, although obviously not 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:

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

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

We define an entire EA as "covered" in the years when the above coverage value was larger than 0.5. Note that having a coverage value larger than 0.5 does not necessarily imply that 50% of its area is covered. For example, if 25% percent 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.

Table 1, Panel A provides basic information about the fraction of all, urban and rural EAs that are covered by at least one network in each year. 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 presence of more variation in coverage in rural areas, the analysis in this paper excludes these areas and focuses on rural Malawi, where the 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 while maps in Figures 4 to 6 provide a visual representation of the cellular coverage data by year in 1997 (when only TNM was active), 2003 and 2008.

IIIB. Geographic Data

Since cellular network expansion involves choices about where to optimally situate towers for best network performance at lowest cost, 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.[7] 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.[8] 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 was provided from the National Statistics Office.

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

Panel B 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 next to a road to more than 50 km distant.

IIIC. 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 percent sample containing detailed information for 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. 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 follows 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 usually contain at least one larger peri-urban center (Boma) with a district hospital, police unit, and commissioner's office. Each district contains a number of Traditional Authorities in the rural areas, and administrative wards in the four urban centers; in 1998 there were a total of 250 Traditional Authorities and 110 wards. 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 all of our analyses.

Enumeration areas (EA) are defined by the National Statistics Office mainly for data collection, while Districts, traditional authorities, and wards are common political divisions used by the government and other institutions. There are approximately 9,200 EAs across the country; we use 8,118 rural EAs for our analyses. The lowest level of disaggregation consists of villages, governed by a village chief. EAs do not uniquely contain villages.

Panel C 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. Almost 90% of adults in rural areas work in agriculture and the average population density is high, at 313 persons per square kilometer, although this mean masks considerable spatial variation.

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 using smaller surveys (an income and expenditure survey from 2004 and Demographic Health Surveys we describe next) and published population projections based on the 1998 Census, companies would have had to collect their own market-level data to construct up-to-date measures of potential market demand.

IIIE. Outcome Data: Demographic and Health Surveys

The final data source we use is from the Demographic and Health Surveys, which provide information on the outcomes we consider here—labor market participation, characteristics of job and sector of work. We use two rounds of the DHS, from 2000 and 2004, which bracket the period of ramp-up of rural network expansion from 27% in 2000 to 57% in 2004. The DHS is a repeated cross-section of individuals sampled from locations across the country. The data contain GIS locations for survey clusters, but the same clusters are not surveyed in subsequent years. Because of the limited overlap in geographic coverage, we cannot compare outcomes over time at the more refined EA level. Instead, we aggregate our coverage data to the TA level, and match DHS data at the TA-level over time. There are 180 TAs sampled in both years of the DHS survey, 159 of which are in rural areas.

The Demographic Health Surveys ask a range of questions about work activities of men and women. Since women's work is often difficult to define in these rural contexts, the DHS asks several probing questions for whether women are doing any work (e.g. working on family farms, working informally) (DHS report, 2004). We use several measures of labor market participation for women and men from these surveys. First, we define a variable over all adult women sampled (ages 15 to 49 inclusive) that describes whether a woman has worked at all in the last 12 months (including whether she is currently working). Then, for the set of women who have worked in the past 12 months, we define several additional labor market variables: whether women report working "away from home", working for themselves, working seasonally or all the time, working for pay or no pay, and their reported sector of work. For men, we have a more limited set of outcomes and can only observe whether they are currently working or working in the past 12 months and, conditional on working, what sector they work in.

For work sector, we focus most of our results on the three major employment sectors of agriculture, sales and skilled labor. Agriculture includes farming for own accord and for a commercial farm, sales includes shopkeepers and shop assistants, while skilled labor is interpreted broadly, including people working as tailors, bricklayers, weavers and in the more formal manufacturing sector.

Table 2 shows summary statistics for these binary outcomes in the combined DHS 2000 and 2004 data as well as for age, education and marital status controls. The sample is restricted to individuals in rural TAs that overlap in both years of the DHS. Note that the sample of women (20,737) is substantially larger than the sample of men (5,238); this is because the DHS surveys every woman between age 15 and 49 in selected households, but only a subset of men (ages 15 to 54 inclusive). All statistics are weighted using DHS survey weights. In this table, we see that women and men have almost the same *current* labor force participation rates: 58% of women and 57% of men report that they are currently working. However, 62% of women report they have worked at all in the past year and 81% of men report doing the same. Conditional on working in the past year, the majority of female employment is self-employment (68%) and many women (56%) work away from home. Given that agriculture employs the largest fraction of female workers (about three quarters), it is not surprising that 60% of women report working seasonally and that only 25% of them report working for cash. Men, on the other hand, are somewhat less likely to be employed in agriculture (62%) and more likely in skilled work (14%); they are also more likely to report working for pay than women.[9]

To get a better sense of how these work characteristics are distributed across different occupations, we present the fraction of working women who are self-employed, working away from home, working without pay etc. for each of the three main occupations in Table 3. Women in agriculture are much more likely to work seasonally and to be unpaid than women in other sectors; while sales and skilled manual work are more likely to involve self-employment. Interestingly, both agricultural work and sales entail working away from home for women. Another striking feature of the table is that only 27% of men in agriculture report working for no pay, while 71% of women in agriculture are unpaid. This might have to do with differences in how men and women interpret the question on “are you working for pay”.

[9] Omitted occupation categories include technical/professional (2% for women, 4% for men), clerical (1% for men), domestic work (1% for women, 2% for men), services (2% for men) and unskilled manual work (1% for women, 2% for men).

The summary statistics for our sample reflect the agricultural basis of Malawi's economy and it will be important to bear this structure of the economy in mind in the second part of our analysis. While there are some indications that there is some diversification out of agriculture, at least for men, any relationship between cellular network expansion and labor market outcomes is likely to operate primarily through changes in the agricultural sector (primary production or in sales). We will be able to observe some, but not all, of these changes since the DHS data do not measure individual wages or income. [10]

IV. Results 1: Patterns of Rollout and Correlates of Coverage

Our work in this section is descriptive; we do not build a model of firm decision-making that tells us where to expect tower placement. However, we are able to learn something about the relative importance of variables proxying for market demand versus those capturing cost considerations in the timing of cellular network expansion in a reduced-form approach.

The maps in Figures 4, 5 and 6 show our basic descriptive evidence on rollout of cellular phone coverage over time. Figure 4 illustrates the areas which had cellular coverage in 1997, when only TNM was active. Coverage is concentrated around the large cities (Lilongwe the capital, and Blantyre the more industrial and market center), and a several of the main tourist destinations around Lake Malawi (Mangochi and Salima). By 2003 (Figure 5) coverage has expanded significantly, moving further out from the cities and specifically, along the main road network. Figure 6 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 7.1 to 7.3, we present estimated Kaplan-Meier survival functions for the time to first cellular coverage within our sample of EAs. Note first that the "time of coverage" on the x-axis measures the year in which 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. In each figure, we adjust the survival functions for variables capturing aspects of market demand and cost of building and maintaining the cellular tower infrastructure. 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

[10] Two leading example of how cellular phones have impacted economic outcomes focus on markets in the primary sector: fishing (Jensen 2006) and grains (Aker 2008). In these papers, the introduction of cellular technology had large impacts on prices of traded primary products, which we are unable to observe in the DHS.

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 7.1 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 7.2 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 between 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 are 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 on some of the important factors related to increased network access, many of these factors are correlated (i.e. altitude and population density).[11] We next turn to examining the determinant of access in a multivariate regression framework. We estimate regressions of the following form for three different years:

$$PHONE_c = \beta_0 + \beta_1 ALTITUDE_c + \beta_2 SLOPE_c + \beta_3 ALTITUDE_c * SLOPE_c + \sum_{k=1}^{K} \gamma_k X_{kc} + \omega_c \quad (1)$$

These regressions are run at the EA level (denoted by c) and each specification contains a full set of TA fixed effects. This means that we identify off of the EA-level variation in demographic variables and geographic factors. Coverage is measured as a binary variable, $PHONE_c = 1$ if the EA is covered by some cellular phone tower in a particular year, where coverage is defined as above. Once an EA has a value of $PHONE_c = 1$ in one year, it maintains that value for all other years in our sample. X_{kc} are a set of k demographic variables computed by aggregating 1998 Census data to the EA-level: we use log of population density, average age within the EA and average education 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 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 4 turns to estimating our first correlates of coverage: demographic variables that proxy for potential market size. Columns 1 through 4 estimate the impact of population density and the age and education structure of the population on coverage in 2000, 2004 and 2008, respectively (the first two years in columns 1 and 2 correspond to the years for which we have DHS data on outcomes). Column 4 reports coefficients from a Cox proportional hazard model for the timing of initial cellular coverage where 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 regressors G_c shift the hazard rate around multiplicatively (Cameron and Trivedi 2005).

We find that a number of proxies for income and potential market demand seem to be important in driving coverage timing. Areas with higher education levels and a lower share of individuals working in

[11] In our sample of rural EAs, the correlation between population density and altitude is -0.1, the correlation between density and slope is -0.17 and the correlation between population density and distance to a road is -0.30.

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 get 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 get early and any cellular network coverage.

Since each cellular tower serves a range of about 30 kilometers, it is the density of individuals within the EA that is most likely to affect the timing of cellular network coverage, rather than total population of the EA. For each year, and for the hazard model covering all years, we control for (log) population density. Our results show that higher population density is indeed associated with significantly earlier coverage in every year. For example, a 1% higher population density (about 30 additional people per square kilometer) is associated with a 2% increase in probability of being covered by the network in 2000.[12]

Adding geographic factors In Table 5 we explore the effect of geographic factors on coverage, controlling for demographics, again at the EA level. The structure of this Table mimics Table 4: the first three columns estimate the effect of demographics and geographic measures on binary measures of coverage in each year and the fourth column estimates a hazard model.

A number of interesting points emerge from this table. First, even when controlling for demographic variables, EAs with lower slope (that is to say, areas with flatter terrain) have a significantly higher probability of getting cellular coverage in early years while areas of higher altitude are significantly less likely to get cellular coverage in the later years, and this coefficient is also significant in the hazard model. 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 (average age, average education and the fraction of the workforce in agriculture) and cellular coverage remains about the same. 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

[12] 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.

substantially weakened and in the last year, is no longer statistically different from zero. This finding is related to 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. Or, in other words, it appears difficult to identify a set of variables that only affect the cost and technology considerations of building cellular towers and others that only affect the market potential.[13]

Together, the geographic and demographic results suggest that both supply-side and demand factors drive cellular phone coverage. We see higher coverage in areas where it seems likely there is higher potential market demand – i.e. those areas with richer people and with more potential users. This concords 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, 2008; Gentzkow and Shapiro, 2008; Jensen and Oster, 2009).

In the next section, it will be useful to bear in mind what our descriptive analysis of the rollout showed about the importance of market potential: population density actually increases in importance as a predictor of new access to the network between 2000 and 2004. Although we do not have good measures of market growth potential in this period, the fact that 1998 population density can predict new coverage more strongly during the period alerts us to the fact that areas gaining new towers are likely different on important baseline conditions compared to areas that do not gain new towers.

V. Results 2: Effects of Cellular Phones on Labor Market Outcomes

The previous section describes some important factors which drive cellular phone coverage introduction at the EA level across Malawi. We found fairly rapid increase in network coverage, correlated with factors related to demand and to supply-side cost factors. In this section, we turn to some preliminary evidence on the effect of cellular phones on labor market participation measured at the TA level.

There are obvious difficulties in pinning down the causal effects of expanded access to communications infrastructure. Cellular coverage and rollout, as we have described above, is not random: various correlates of income (i.e. schooling or health), as well as geographic factors, are correlated with earlier coverage. Thus, simply estimating the correlation between coverage and outcomes will not allow us to

[13] The inclusion of mean distance to road entirely knocks out the relationship between population density and cellular coverage in 2008.

estimate results with a causal interpretation. There are at least two solutions to this. One would be to use an instrument for coverage. Of the coverage determinants above, the most promising instrument is slope. Slope seems to drive coverage for supply-side, cost, reasons, which makes it not as obviously problematic as the demand-side correlates of coverage. However, using slope as an instrument would require us to argue that slope is unrelated to other outcomes for any reasons other than cellular coverage. This seems unrealistic in Malawi, where much economic activity centers on agriculture. Variations in slope are likely to drive variations in what type of crops flourish over time; they are likely correlated with distance to Lake Malawi, and so on. In the language used by Deaton (2009), slope is "external", but is unlikely to be exogenous.

As an alternative, we take advantage of variations in cellular coverage over time *within* an area.[14] This is similar to the approach used by both Jensen (2008) and Aker (2009) as well as, in another context, that used by Jensen and Oster (2009) and La Ferrara et al (2009). We observe 180 TAs (denoted by γ_c) in the DHS data over the two years: 2000 ($YEAR_t=0$) and 2004 ($YEAR_t=1$). Our approach is to estimate whether there are changes in the individual level outcomes we consider (y_{ict}) which correlate with changes in cellular phone coverage at the TA level ($CELL_{ct}$). This is a difference-in-difference approach. We use areas without changes in coverage (either those that already have coverage in 2000, or those who do not get coverage until 2004) as the "control" group, and compare to areas which get coverage between the two survey rounds (the "treatment" group). This corresponds to the following empirical specification:

$$y_{ict} = \alpha_0 + \alpha_1 YEAR_t + \alpha_2 CELL_{ct} + \alpha_3 X_{ict} + \sum_{c=1}^{C} \gamma_c + \varepsilon_{ict} \qquad (3)$$

where standard errors are clustered at the TA-year level. Because we focus on the relationship between new cellular network *access* and labor market outcomes rather than between cellular phone *use* and outcomes, α_2 estimates an *Intent to Treat* effect, which is likely to be an underestimate of the *Treatment on the Treated* effect, or the effect of giving one individual a phone and network access.

The obvious concern with this type of approach is the possibility that some third, omitted, factor drives both increases in coverage and changes in the outcomes of interest, so that $COV(CELL_{ct}, \varepsilon_{ct}) \neq 0$. Although we will not be able to fully rule out this possibility, we use information on areas which get coverage right after the 2004 survey to evaluate whether there are pre-trends in outcomes which anticipate increases in cellular coverage.

[14] In order to create a matched sample of areas in the DHS 2000 and 2004, we must use variation in cellular coverage at the more aggregate TA level.

Labor Market Participation, Type and Sector of Work We begin by looking at whether labor market participation is correlated with cellular coverage. In our sample, 58 percent of women and 57 percent of men report that they are currently working (see Table 2). If having access to cellular phones increases productivity (as observed in Jensen, 2008) then it seems plausible that it might increase labor market participation, either through job creation or through more efficient job search.[15] In addition, as noted in the data section, we also observe for women whether they are engaged in work away from home, whether they work for themselves and whether this work is constant, seasonal or occasional. It seems plausible that having cellular phone coverage could increase one's ability to work further away, since contacting home is easier. It potentially also reduces the uncertainty associated with working in distant markets, as information about these markets and employment opportunities can be collected ahead of time (Jensen, 2008; Aker, 2009).

Figures 8, 9 and 10 show the relationship between new cellular coverage and labor market participation. In all three figures we divide the sample into three types of areas: areas that already have cellular coverage in 2000, those who do not have coverage even by 2004, and those who get coverage between the surveys. Figure 8 presents overall male labor force participation. There is no evidence here that changes in cellular coverage influences the labor market for men: if anything, there is a *decline* in working in areas that change coverage.

Figure 9 reports effects on overall female labor market participation. For the areas that do not change coverage, both those which never have coverage and those which always have it, there is a decline in working between the two years. In contrast, for areas where coverage changes between surveys, there is a 5 percentage point increase in the chance of working. Considering the average decline in the other two groups as the counterfactual, this suggests about a 10 percentage point impact of cell phones. Of course, since men and women work in the same labor market (largely in agriculture), a reduction in male employment in areas gaining access to cellular phones could be connected to the increase in female employment in these areas.

Perhaps most interesting is Figure 10 which reports effects on women working “away from home”. There seem to be some secular increases in working away from home in the areas without changes in coverage, but they are dwarfed by the 16 percentage point increase in this variable in the areas that change cellular phone coverage between the surveys. Since we saw female labor force participation increase by 7 percentage points (Figure 9), this larger impact on working away from home is likely driven by both an

[15] Aker and Mbiti (2010) point out that better job search and effective job creation are plausible channels through which new cellular networks could change the way labor markets operate in Africa.

extensive and intensive margin effect: some women newly working are doing so away from home, while others who were working before now choose to work away from home.

Table 6 evaluates these claims statistically for current work and for work within the past year. We also present results for women working away from home, for self-employment, for work in agriculture and for unpaid work, conditional on women working at all. In each column we control for TA fixed effects and year as well as the individual's age, years of education and marital status, and estimate the impact of having cellular coverage on the outcome as in equation (3) above. Because we include the TA fixed effects, the coefficient on cellular phone coverage is identified off of TAs where coverage changes; this is the statistical equivalent to Figures 8-10. The unit of observation in these regressions is the individual and the sample is restricted to rural TAs. In our sample of 159 matched rural TAs from the DHS, 75 TAs gain coverage between 2000 and 2004; 33 TAs never have coverage and 51 TAs always have coverage. About half of the individuals in the DHS sample live in TAs that gain new access to the cellular phone network.

Panel A shows the basic estimates of the effect of cellular coverage on employment for women and for men. For each outcome, results in the first column correspond to estimates of equation (2) and we present only the coefficient on $CELL_{ct}$. Consistent with Figures 8 and 9, we see some evidence that women are working more in areas gaining new access to network, relative to areas that are not gaining coverage. Current labor force participation increases by 8.3 percentage points while work in the past year increases by a larger, 11.5 percentage points. Given what we saw from Tables 2 and 3 that women work mainly in seasonal agricultural jobs, the gap between these two estimates is not surprising; it also suggests that some of this increase in work is operating through more jobs in agriculture. There is also no evidence that cellular coverage has an impact on men's labor force participation: the coefficients on cellular coverage for men's work (current and recent) are negative, but not significant.

As noted above, there is a concern that these effects are driven by other trends in areas which get cell phones. It seems plausible that cellular phone networks are expanding into areas that are already growing, especially if companies are trying to anticipate future market demand. If this is the case, the coefficient on cellular coverage does not reveal the impact of new telecommunications networks, but is rather driven by positive economic trends. To get some sense of whether this is an issue, the second column for each outcome variable in Panel A presents estimate from the same regressions as in (3), and adds a variable measuring whether the area gets cellular coverage in the four years *after* the 2004 survey.[16] In doing so, we aim to test for the existence of pre-trends in the outcomes which anticipate increases in cellular coverage.

In all of the cases in which we see an effect of current cellular coverage on female employment outcomes, we do not see that future coverage is associated with improvements between 2000 and 2004. The measure of the effects of future coverage on female labor force participation is actually negative (and significant at the 10% level for the outcome “working in the past year”) which would go against the story that cellular network expansion is “following” positive trends in economic activity. For current and recent work and working away from home and for women working away from home, we can reject that the coefficients on cellular coverage and future cellular coverage are equal.

Even though this second set of regressions in Table 6 showed no indication that cellular phone changing areas were on an upward trajectory in terms of increased employment for women, results for men do indicate that these places may be experiencing different trends for male employment. In areas that are soon to get access to the network, men are significantly more likely to report recent employment. It is possible that this pattern is also responsible for the negative coefficient for women in Column 2 if male and female labor supply are substitutes. Although the pre-trend analysis for women was encouraging, this analysis of male employment cautions against too strong a causal interpretation. More data, perhaps including more years, would be necessary to fully rule out the possibility that our results are driven by other changes in the areas which adopt cellular technology.

Table 6 also allows us to examine how the type and sector of male and female work is changing in areas getting new access to the cellular phone network. Panel B and C presents estimates from regressions for women working for self, working away from home, working in agriculture, sales and skilled manufacturing. Panel C presents estimates for male employment in each of the three major sectors of work.

Cellular coverage is associated with a 9.1 percentage point increase in working away from home, and a 4.9 percentage point increase in women working in agriculture, although this is not statistically different from zero. Women are also about 4.9 percentage points *less* likely to be working in sales. Since we have defined these outcomes to be for the set of employed women, this suggests that the new jobs that we see in areas into which the cellular network expands are skewed towards agricultural work, at least for women. The results for men, on the other hand, suggest that in these same areas, men are not more likely to be employed in agriculture, but rather have a higher likelihood of working in skilled manufacturing

[16] Controlling for pre-trends in economic outcomes would be an alternative option if we had pre-2000 economic data.

(Panel C). We should caution that, again, we see significant pre-trends for men: areas which get cellular access in the near future see reductions in man involved in sales between 2000 and 2004.

Labor market participation by demographic group: Our data are not sufficient to get a strong sense of *why* cellular phones matter for labor market participation for women in particular-- i.e. what mechanisms drive these effects. However, we may be able to make some progress on this by exploring how the effects vary across demographic groups. This is done in Table 7. In this table, each row is a separate regression with the same specification as in Table 6. For each of the outcomes of interest, we run the regression separately by age groups, education groups, marital status and whether or not the individual has children at home. In all regressions, we include the same set of controls as in Panel A, Table 6.

The change in employment in response to change in cell phone access is similar across age, education and marital status. We find the most interesting variation in whether the woman has children at home or not. Women who have children at home see a much larger increase in employment upon the introduction of mobile access.

In Panel B, when we consider working away from home, we see a similar pattern. Women who have children at home, conditional on working, are 19.1 percentage points more likely to report working away from home in areas experiencing an increase in the telecommunications network.

Our ability to more fully understand these patterns is limited by the set of questions available in the DHS data. The larger impacts on women with children could point to the additional value of good communication technology for women in this category. However, at this point the data are not rich enough to make this claim with certainty.

VI. Conclusion

This paper makes use of a new, very detailed, dataset on the history and location of each cellular phone tower construction in Malawi. The goals of this paper are three-fold. First, we provide detailed evidence on cellular phone rollout in a very low-income context. We show evidence that, despite the fact that Malawi is extremely poor, cellular rollout occurs quite quickly and that by 2008, over half of the country has access to at least one network. Further, both supply-side cost factors and demand-side market potential variables are important for the timing of initial coverage. We showed that population densityis an important predictor of rollout in most years. Our descriptive analysis also highlighted that it is often difficult to separate demand and supply variables, since factors that increase cost of tower building (e.g.

altitude) are correlated with factors that proxy for market demand (i.e. population density). This evidence cautions against using pure cost variables as instruments for infrastructure placement in this particular setting.

The second part of the paper presents new results suggesting the influence of that new cellular coverage may have on some important economic outcomes which have not often been connected to this type of infrastructure expansion.[17] We find that female labor force participation seems to rise by between 8 and 11.5 percentage points in the wake of new cellular coverage and that these effects operate through more women working away from home. These effects are preliminary, and subject to the usual criticisms about clean identification in a difference-in-differences analysis. However, the new patterns that we highlight here – that female (and not male) employment overall is increasing in areas gaining new access to the network, that the structure of work for men and women is shifting in areas newly connected and that these changes are different depending on the individual's marital status and the presence of dependents – are interesting in their own right. We believe they suggest some new directions for research into the impact of cellular coverage in the developing world; in particular, how cheaper communications technologies in rural labor markets may improve the flow of information, provide new types of work opportunities and therefore affect labor market outcomes.

The final aim of our paper has been to highlight research potential of combining different data sources from an African country to learn more about how large companies operate and impact the economy. Although obtaining administrative data on network expansion was not straightforward, we believe that with careful thought about research design, there is much to be learned from using existing data on various program expansions (not limited to infrastructure) in developing countries, in combination with more familiar census data and household surveys.

[17] Klonner and Nolen (2008) is an exception.

References

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

Aker, J. and I. Mbiti (2010) "Mobile phones and economic development in Africa" *Journal of Economic Perspectives*, Summer

Alesina, A., R. Baqir, and W. Easterly (1999): "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114(4), 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

Alesina, A., and E. L. Ferrara (2000): "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115, 847–904.

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

Benson. T., J. Kaphuka, S. Kanyanda, R. Chinula (2002) "Malawi: An atlas of social statistics". IFPRI , Washington D.C. and National Statistics Office, Zomba

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)

Cadot, Oliver, Lars-Hendrick Roller and Andreas Stephan. (1999) "A Political Economy Model of Infrastructure Allocation: an Empirical Assessment." *CIG Working Papers*

Cameron, A. C. and P. K. Trivedi (2005) Microeconometrics: Methods and applications, Cambridge University Press.

Castells, Antoni and Albert Sole-Olle. (2005) "The regional allocation of infrastructure investment: The role of equity, efficiency and political factors." *European Economic Review*, 49(5)

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

Collier, P. (2000). "Doing Well Out of War," in Greed and Grievance: Economic Agendas in Civil Wars, ed. by M. Berdal, and D. Malone

Deaton, A. (2010). "Instruments, randomization and learning about development", *Journal of Economic Literature*, Vol. 48, No. 2, June

Dinkelman, Taryn. (2008) "The effects of rural electrification on employment: New evidence from South Africa." University of Michigan, Mimeo

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

Fink, C, Mattoo, A. and R. Rathindran (2002) "An assessment of telecommunications reform in developing countries", *World Bank Policy Research Working Paper* No. 2909

Gasmi, F. and Virto, L. R. (2010) “The determinants and impact of telecommunications reforms in developing countries”, *Journal of Development Economics*, In press

Gentzkow, Matthew and Jesse Shapiro. "Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study," *Quarterly Journal of Economics*. Vol. 123, Issue 1. February 2008.

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

Jack, W. and T. Suri (2009) “Mobile money: The economics of M-Pesa”, MIT Sloan, Mimeo

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

Jensen, Robert. (2010). “Economic opportunities and gender differences in human capital: Experimental evidence from India. *NBER Working paper 16021*

Jensen, Robert. (2007) "The Digital Provide: Information (Technology), Market Performance and Welfare in the South Indian Fisheries Sector." *Quarterly Journal of Economics.* Vol 122, Issue 3. August

Jensen, Robert and Emily Oster (2009) "The Power of TV: Cable Television and Women's Status in India," *Quarterly Journal of Economics.* Vol 124, Issue 3. August

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

La Ferrara, Eliana, Alberto Chong and Suzanne Duryea (2009) ”Soap Operas and Fertility: Evidence from Brazil“ *CEPR Discussion Paper* 6785

Luijs, L. (2000): “The Impact of Ethnic Heterogeneity on the Quantity and Quality of Public Spending,” *IMF Working Paper* 00-49

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

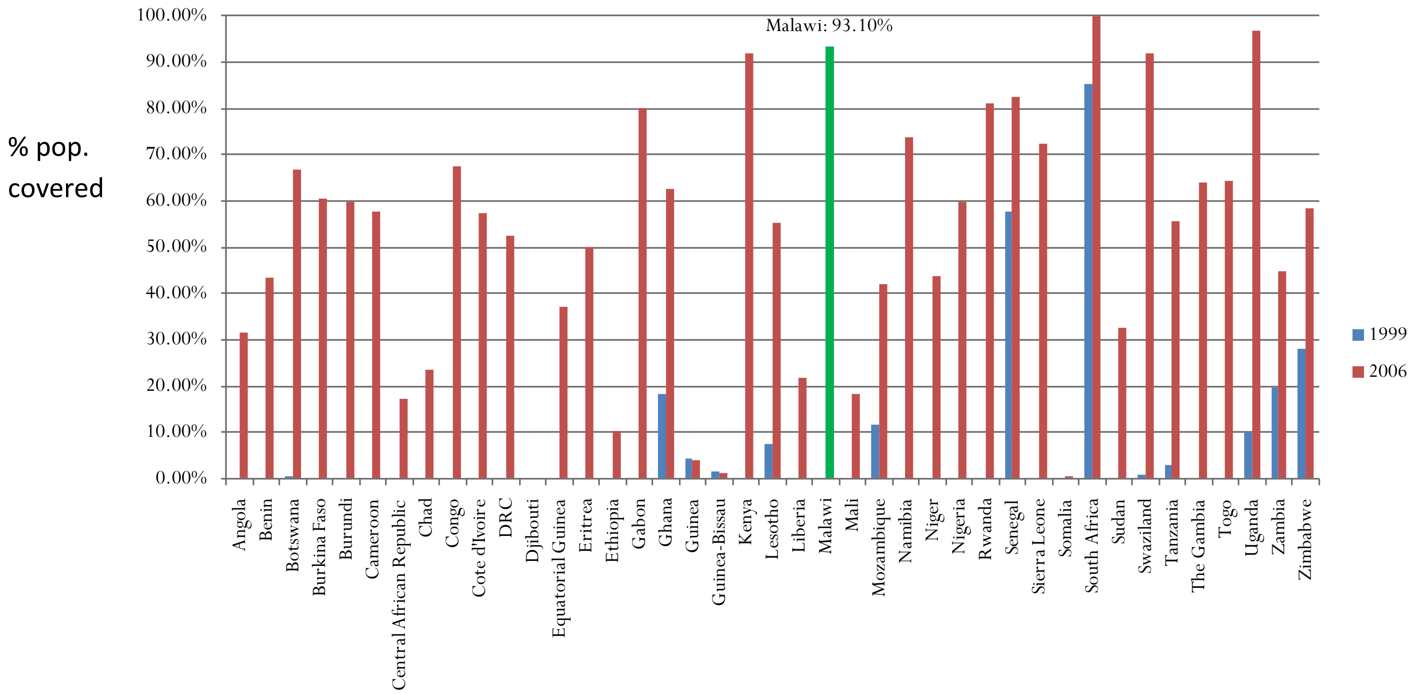
McGreal, Chris. (2009) “Front: From Congo to Kathmandu, how mobiles have transformed the world”, *The Guardian (London)*, March 3.

Miguel, E., and M. K. Gugerty (2005): “Ethnic Diversity, Social Sanctions and Public Goods in Kenya,” *Journal of Public Economics*, 89, 2325–2368

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

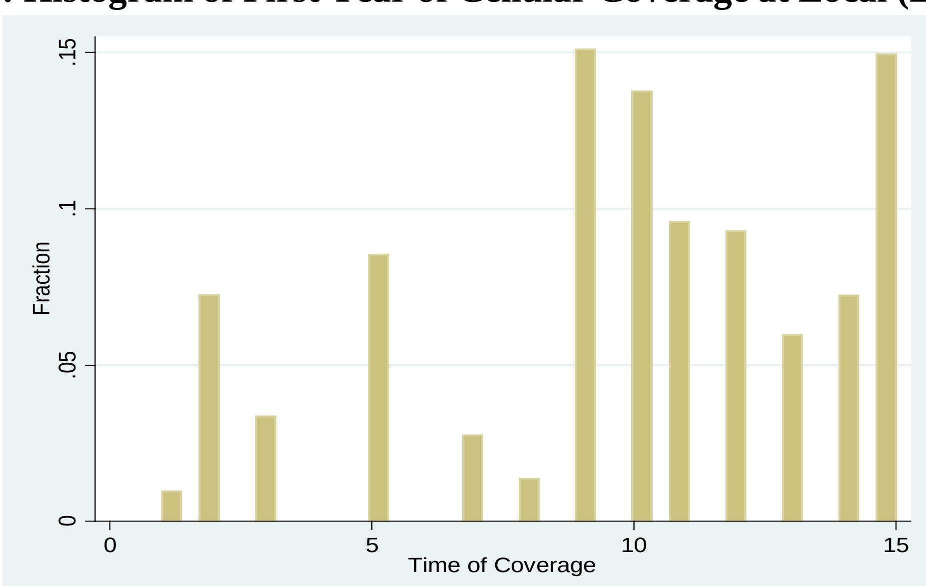
OECD (2009). "OECD Communications Outlook 2009". *OECD Publication*, August 11.

Figure 1: Cellular phone coverage in Africa



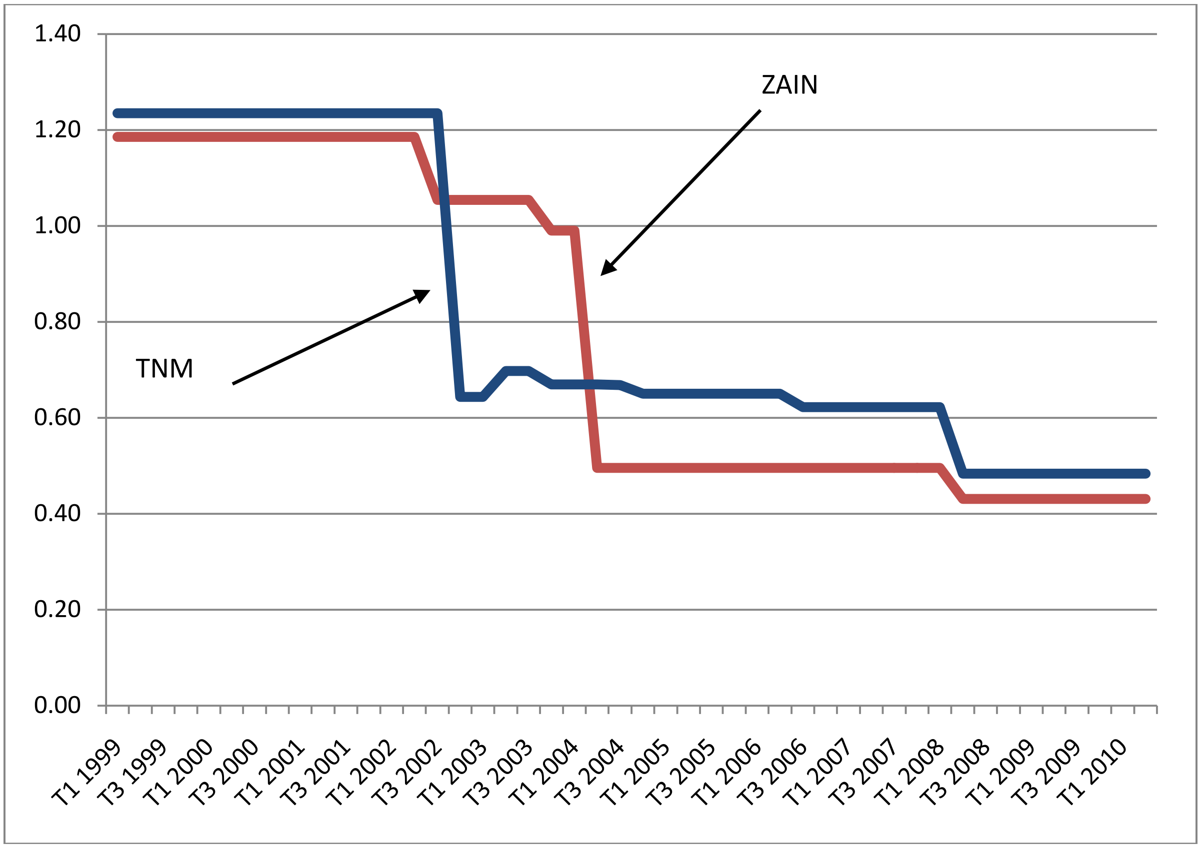
Source: Buys, Dasgupta, Thomas, Wheeler (2008)

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



Unit of observation is the Enumeration Area. Histogram shows distribution of years in which EAS are covered, as defined in the text. Range of years is from 1995 (year 1) to 2009 (year 15). Observations without coverage by 2009 are given a value of 15. Sample is restricted to rural areas with positive population in 1998 Census.

Figure 3: Average price per second (US cents) for a representative consumption bundle, pre-paid consumer in Malawi



Notes: Price per second series is constructed using an average consumption basket for a pre-paid consumer. The basket assumes a distribution of call durations, call destinations (in or out of network) and call timing (peak/off-peak) where the assumptions are based on information we have on cellular phone traffic from the later 2008-2010 period. Tariff information is gathered from a comprehensive survey of all the major newspapers in Malawi from 1999 to 2010.

Figure 4: Map of Cell Phone Coverage in 1997 (TNM Only)

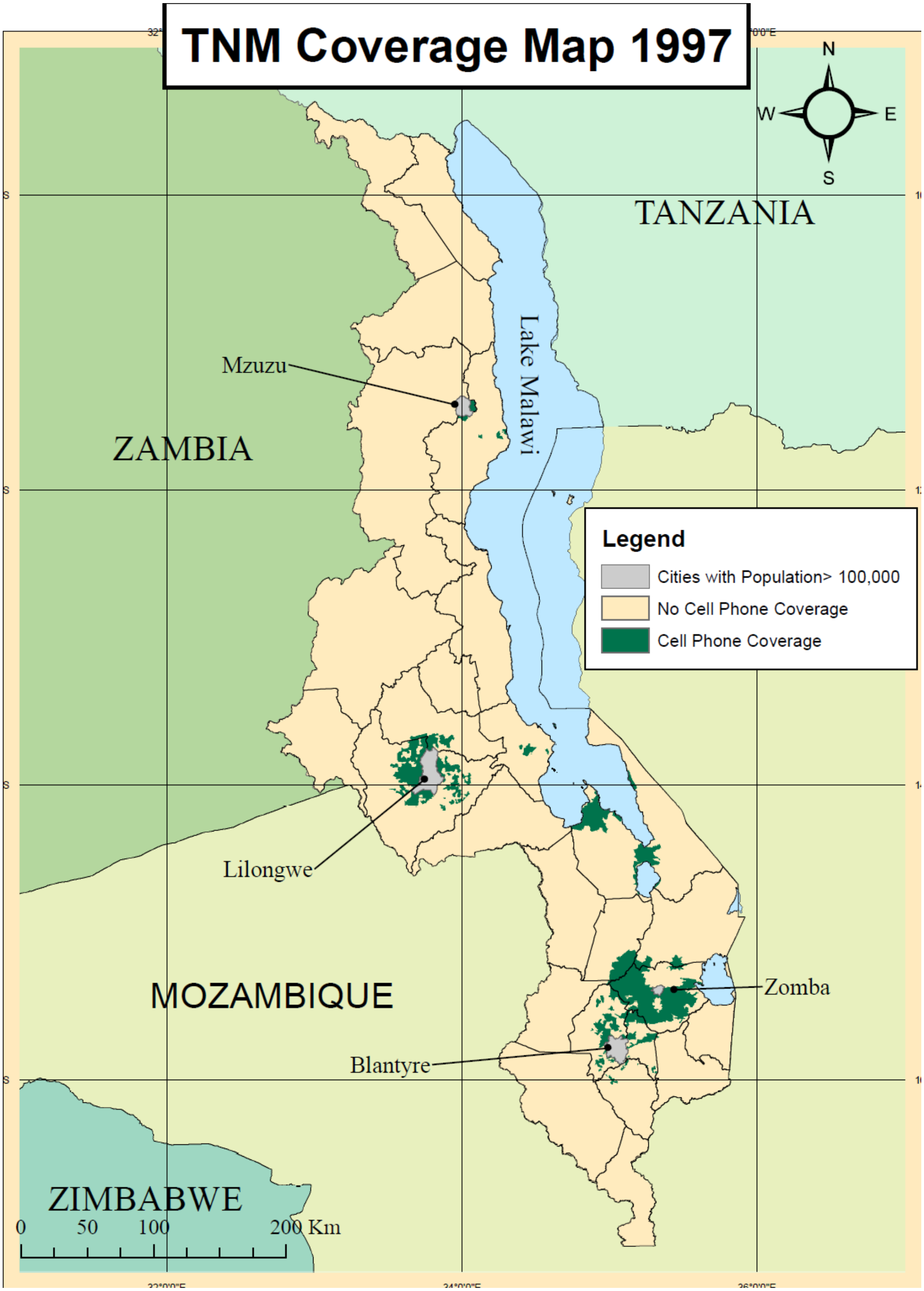


Figure 5: Map of Cell Phone Coverage in 2003 (Both Networks Active)

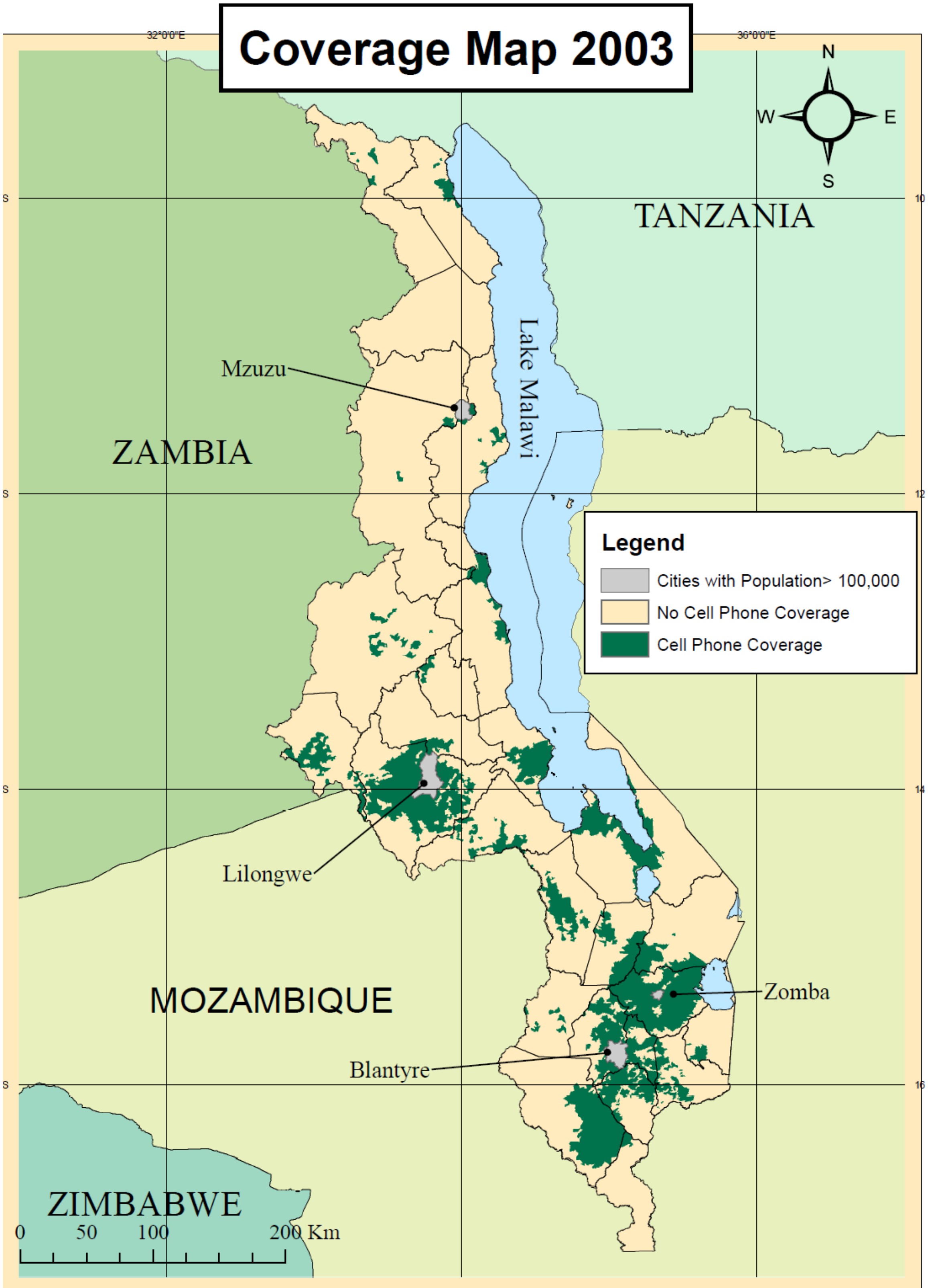


Figure 6: Map of Cell Phone Coverage in 2008 (Both Networks Active)

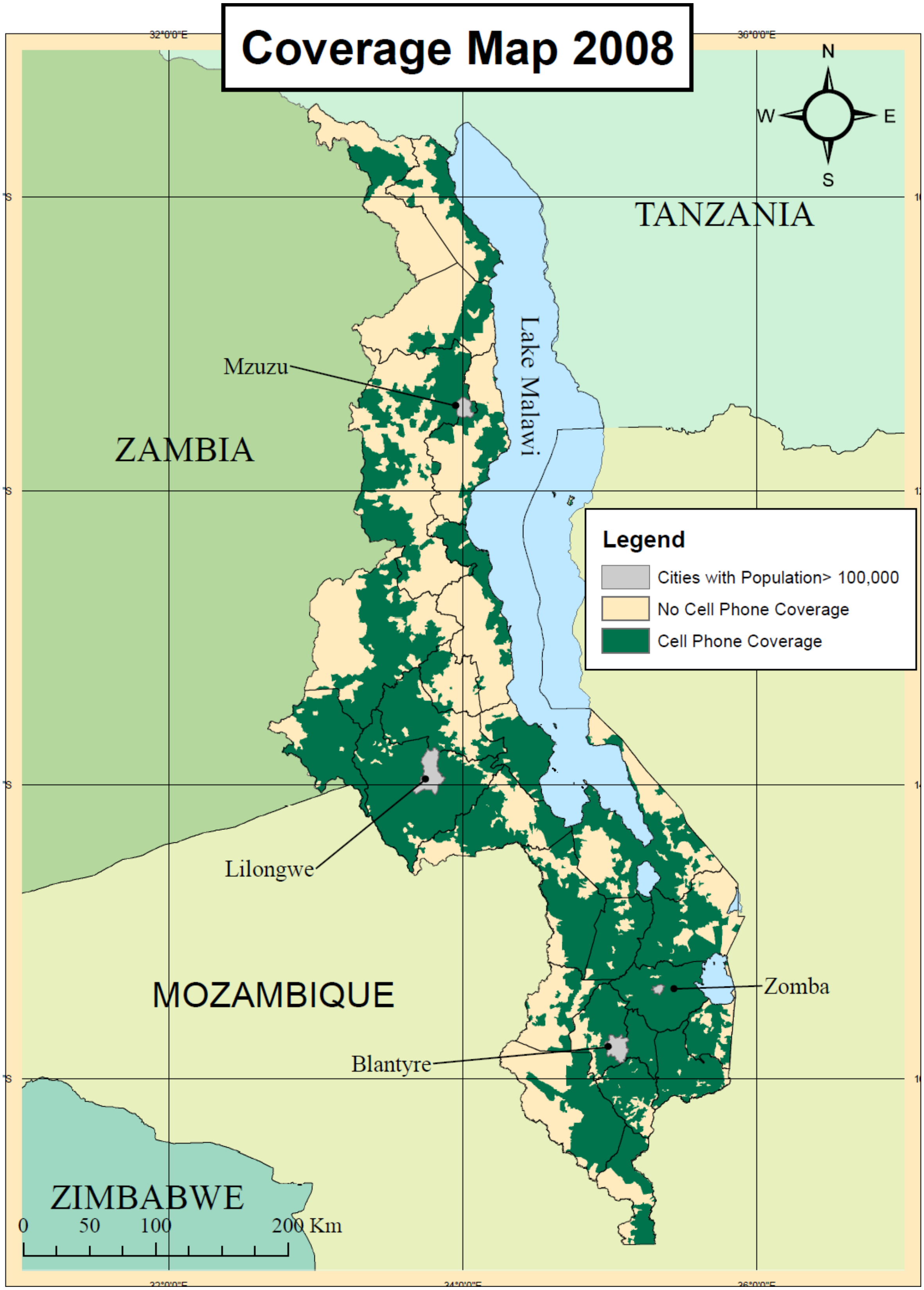


Figure 7: Arrival of cellular coverage at the local level

Figure 7.1 By 1998 Population Density (1=sparse, 4=dense)

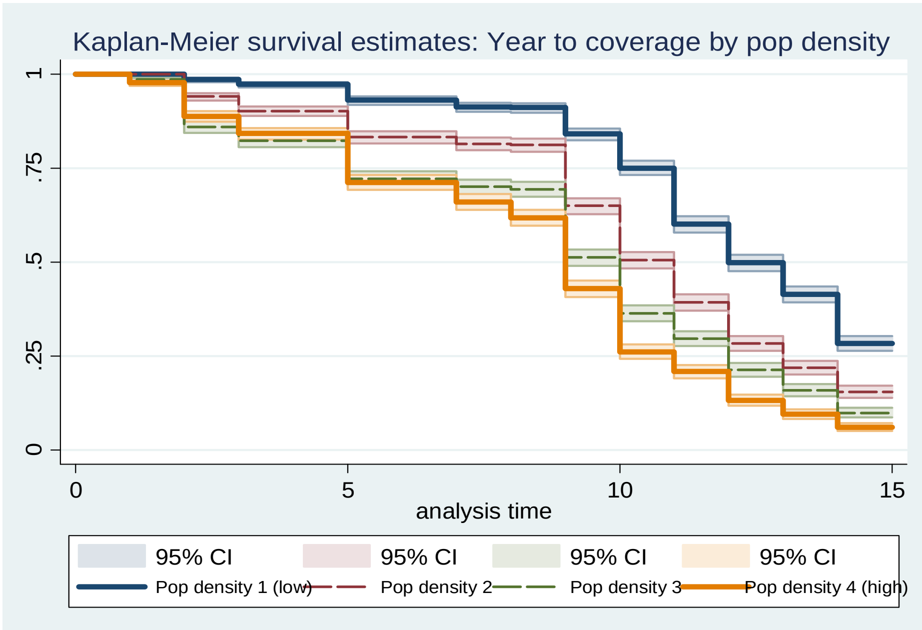


Figure 7.2 By slope group (1=flat, 4=steep)

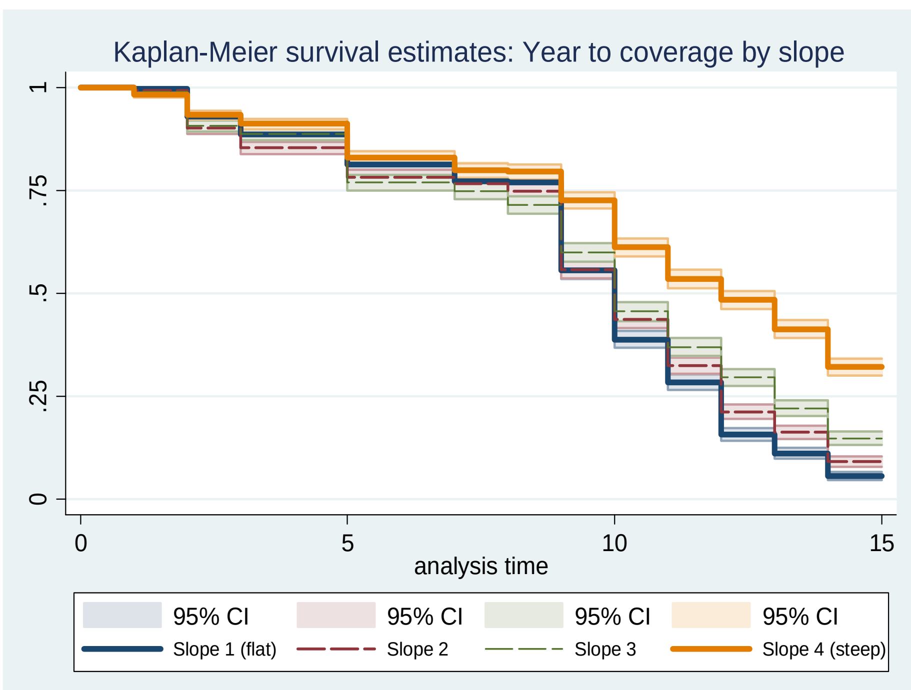
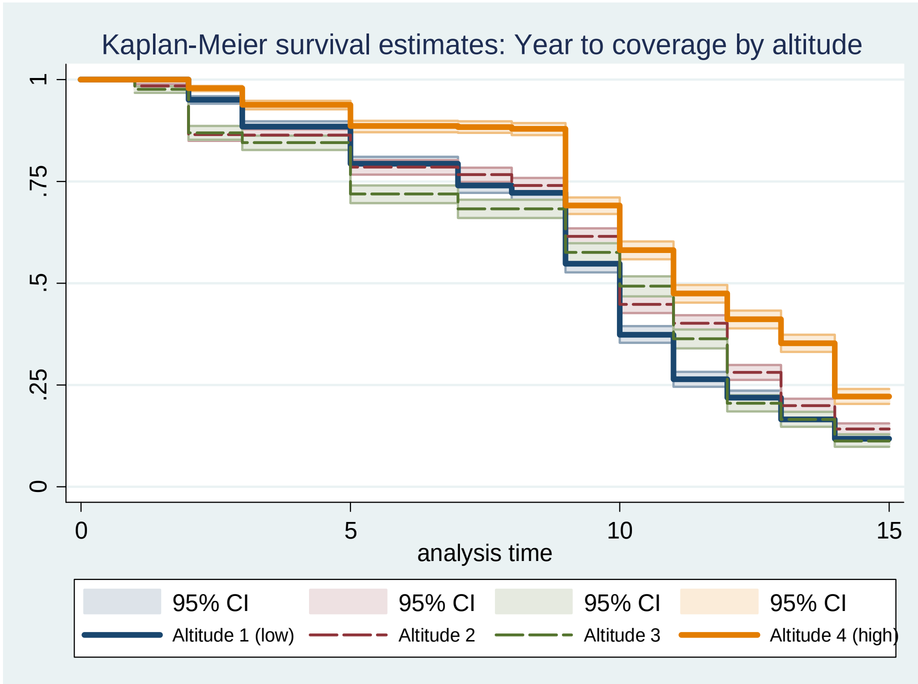
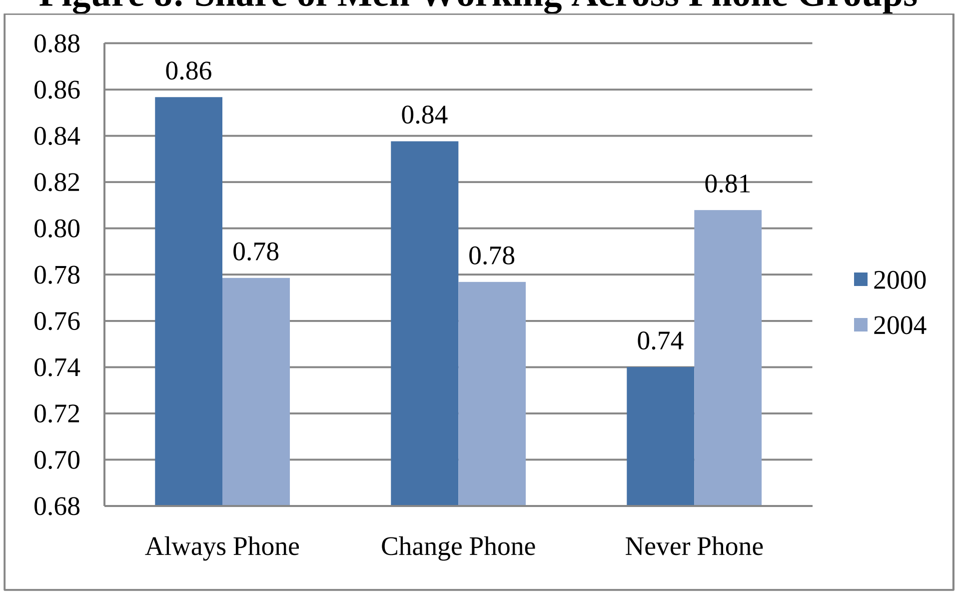


Figure 7.3 By Altitude group (1=low, 4=high)



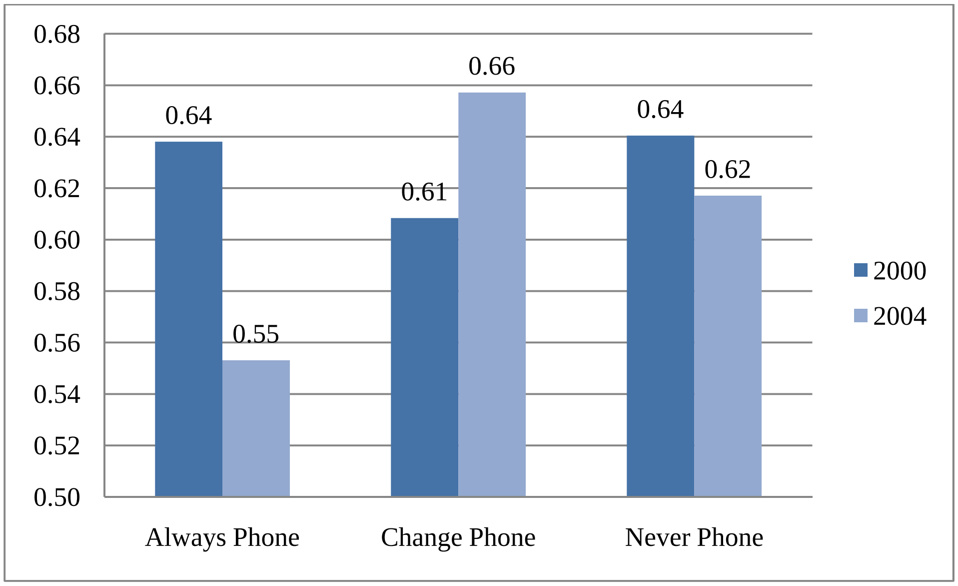
Unit of observation is the Enumeration Area. Graphs show Kaplan-Meier survival functions for EAs with different values of slope, altitude and population density, where exit is into the "covered" state. Range of years is from 1995 (year 1) to 2009 (year 15. Observations without coverage by 2009 are assigned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

Figure 8: Share of Men Working Across Phone Groups



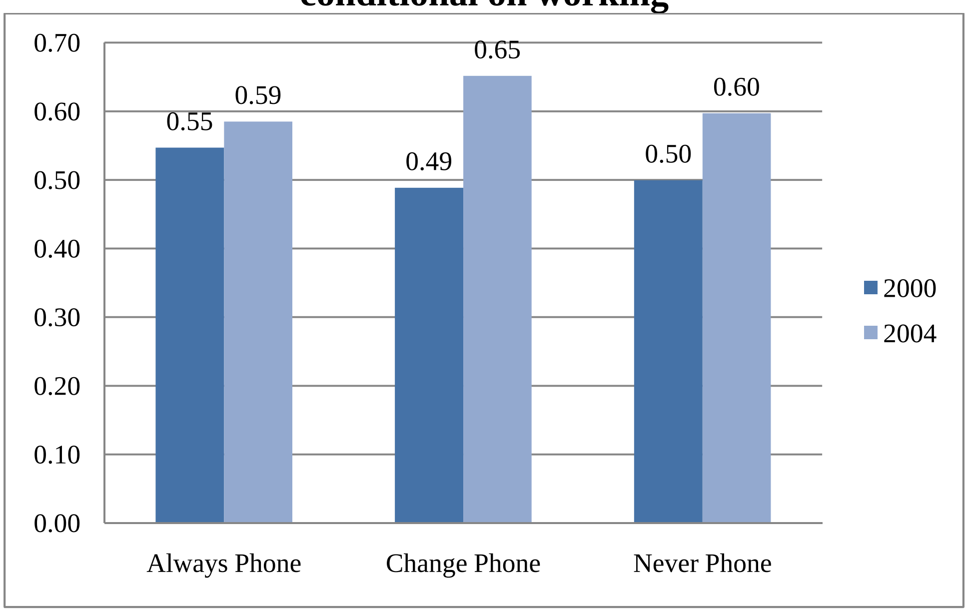
Notes: Based on data from the DHS 2000 and 2004. Areas are defined as changing phone status if they do not have cell coverage in 2000 and do have it in 2004. Unit of observation is the individual; cell phone coverage by year is defined at the TA level. Means are weighted using DHS sample weights.

Figure 9: Share of Women Working Across Phone Groups



Notes: Based on data from the DHS 2000 and 2004. Areas are defined as changing phone status if they do not have cell coverage in 2000 and do have it in 2004. Unit of observation is the individual; cell phone coverage by year is defined at the TA level. Means are weighted using DHS sample weights.

Figure 10: Share of Women Working Away from Home By Phone Groups, conditional on working



Notes: Based on data from the DHS 2000 and 2004. Areas are defined as changing phone status if they do not have cell coverage in 2000 and do have it in 2004. Unit of observation is the individual; cell phone coverage by year is defined at the TA level. Means are weighted using DHS sample weights.

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 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 C: Summary statistics of demographic variables from Census 1998: 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 and C 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: Summary statistics for DHS 2000 and 2004, Individual level

	Women					Men				
	N	Mean	s.d.	min	max	N	Mean	s.d.	min	max
<u>Demographics</u>										
Age	20,737	27.98	9.35	15	49	5,238	29.43	10.61	15	54
Education (years)	20,737	3.89	3.42	0	19	5,226	5.23	3.36	0	19
Married	20,737	0.69	0.46	0	1	5,238	0.62	0.49	0	1
<u>Fraction of adults:</u>										
Currently working	20,735	0.58	0.49	0	1	5,237	0.57	0.50	0	1
Working in the last year	20,735	0.62	0.49	0	1	5,238	0.81	0.40	0	1
<u>Conditional on working in past yr, fraction of adults:</u>										
Self-employed	12,853	0.68	0.47	0	1	--	--	--	--	--
Work away from home	12,794	0.56	0.50	0	1	--	--	--	--	--
Work always	12,856	0.25	0.44	0	1	--	--	--	--	--
Work seasonally	12,856	0.60	0.49	0	1	--	--	--	--	--
Work occasionally	12,856	0.07	0.26	0	1	--	--	--	--	--
Paid in cash	12,867	0.25	0.43	0	1	4,205	0.44	0.50	0	1
Paid at all	12,873	0.27	0.44	0	1	4,206	0.54	0.50	0	1
<u>Fraction of working adults in (collapsed categories)</u>										
Agriculture	12,873	0.74	0.44	0	1	4,206	0.62	0.48	0	1
Sales	12,873	0.18	0.39	0	1	4,206	0.11	0.31	0	1
Skilled work (manufacturing)	12,873	0.03	0.18	0	1	4,206	0.14	0.35	0	1

Notes: Unit of observation is the individual in DHS 2000 or DHS 2004. Sample includes all individuals in matched TAs and excludes observations in the four urban areas. The sample of adults intereviewed includes women age 15-49 inclusive and men age 15-54 inclusive. All statistics are weighted using DHS weights. Not all employment questions were asked of men in the DHS 2004. Residual occupation categories are: professional/technical, clerical, domestic services, services, unskilled manual work.

Table 3: Characteristics of the three major occupations, DHS 2000 and 2004

	Fraction women working in:			Fraction men working in:		
Who are:	Agriculure	Sales	Skilled manual	Agriculture	Sales	Skilled manual
Self-employed	0.68	0.81	0.76			
Work away from home	0.57	0.54	0.23			
Work always	0.19	0.39	0.37			
Work seasonally	0.71	0.34	0.25			
Work occasionally	0.02	0.23	0.30			
Paid in cash	0.12	0.55	0.69	0.28	0.59	0.70
Unpaid	0.71	0.35	0.19	0.27	0.14	0.06
Paid in cash/kind	0.10	0.05	0.05	0.10	0.04	0.05

Notes: This table shows, for each occupation, what fraction of individuals report working for self, working away from home etc. Means are weighted using DHS weights, data are from the DHS 2000 and 2004. Unit of observation is the individual who reports working in the last 12 months, sample includes individuals in TA-matched data and excludes individuals in the four urban areas. Women age 15-49 inclusive and men age 15-54 inclusive make up the sample.

Table 4: Cellular phone coverage and demographic proxies for market demand

	OLS 2000	OLS 2004	OLS 2008	Cox
	(1)	(2)	(3)	(4)
Log EA Population Density	0.020***	0.048***	0.027***	0.280***
	(0.00)	(0.01)	(0.00)	(0.01)
% Empl in agriculture	-0.033	-0.136**	-0.020	-0.492***
	(0.03)	(0.04)	(0.03)	(0.10)
Av. Education (years)	0.018***	0.029***	0.036***	0.066***
	(0.00)	(0.01)	(0.00)	(0.01)
Av. Age	-0.001	0.016***	0.012***	0.099***
	(0.00)	(0.00)	(0.00)	(0.01)
R-squared	0.558	0.475	0.367	
N	8,118	8,118	8,118	8,118
Mean of dependent var.	0.20	0.53	0.85	9.9

Notes: This table shows the relationship between cellular phone coverage in different years and market demand variables. The unit of observation is the EA, sample is restricted to rural EAs with positive population according to 1998 Census data. All regressions and Cox models contain TA fixed effects and a constant. Standard errors in parentheses are clustered at the TA level, *p<0.1, **p<0.05, ***p<0.01

Table 5: Cellular phone coverage, demographic proxies for market demand, and geographic variables

	(1)	(2)	(3)	(4)
	OLS 2000	OLS 2004	OLS 2008	Cox
Log EA Population Density	0.010*	0.027***	0.000	0.210***
	(0.00)	(0.01)	(0.00)	(0.01)
% Empl in agriculture	-0.044	-0.161***	-0.040	-0.578***
	(0.03)	(0.04)	(0.03)	(0.10)
Av. Education (years)	0.013**	0.025***	0.031***	0.086***
	(0.00)	(0.01)	(0.00)	(0.01)
Av. Age	-0.002	0.012***	0.006*	0.093***
	(0.00)	(0.00)	(0.00)	(0.01)
Log Altitude	0.020	-0.004	-0.101***	-0.054**
	(0.02)	(0.03)	(0.03)	(0.02)
Log Slope	-0.138***	-0.216**	-0.086	-0.052
	(0.04)	(0.07)	(0.06)	(0.17)
Int of Log Alt & Log Slope	0.018**	0.021*	0.004	-0.032
	(0.01)	(0.01)	(0.01)	(0.03)
Mean Distance to Road (kilometers)	-0.014***	-0.018***	-0.025***	-0.058***
	(0.00)	(0.00)	(0.00)	(0.01)
R-squared	0.560	0.484	0.390	
N	8,115	8,115	8,115	8,115
Mean of dependent variable	0.20	0.53	0.85	9.90

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 Cox models 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

Table 6: Cellular Phone Coverage and Employment Outcomes

Panel A: Cellular phone coverage and Labor Market Participation

	Women working currently		Women working in the past year		Men working currently		Men working in the past year			
Cellular Coverage	0.083***	0.082***	0.115***	0.113***	-0.037	-0.037	-0.016	-0.012		
	(0.030)	(0.030)	(0.030)	(0.030)	(0.043)	(0.043)	(0.025)	(0.025)		
Future Cellular Coverage		-0.044		-0.056*		-0.008		0.070***		
		(0.034)		(0.032)		(0.050)		(0.026)		
N	20,735	20,735	20,735	20,735	5,225	5,225	5,226	5,226		
R2	0.11	0.11	0.11	0.11	0.24	0.24	0.27	0.27		
Fstat: difference		9.38		19.77		0.20		7.20		

Panel B: Women, Type of Work, Conditional on Working

	Self employment		Work away from home		Occupation: Agricultural		Occupation: Sales		Occupation: Skilled manuf.	
Cellular Coverage	-0.013	-0.012	0.091***	0.091***	0.049	0.049	-0.049*	-0.048*	-0.011	-0.012
	(0.026)	(0.026)	(0.025)	(0.025)	(0.031)	(0.030)	(0.026)	(0.025)	(0.009)	(0.009)
Future Cellular Coverage		0.034		-0.007		-0.017		0.029		-0.006
		(0.027)		(0.024)		(0.044)		(0.036)		(0.013)
N	12,853	12,853	12,794	12,794	12,873	12,873	12,873	12,873	12,873	12,873
R2	0.13	0.13	0.10	0.10	0.17	0.17	0.10	0.10	0.04	0.04
Fstat: difference		1.57		7.18		1.01		2.14		0.15

Panel C: Men, Type of Work, Conditional on Working

					Occupation: Agricultural		Occupation: Sales		Occupation: Skilled manuf.	
Cellular Coverage					-0.029	-0.027	-0.008	-0.011	0.041**	0.041**
					(0.031)	(0.030)	(0.020)	(0.019)	(0.020)	(0.020)
Future Cellular Coverage						0.038		-0.066***		0.014
						(0.033)		(0.021)		(0.019)
N					4,198	4,198	4,198	4,198	4,198	4,198
R2					0.20	0.20	0.10	0.10	0.10	0.85
Fstat: difference						1.80		3.92		0.08

Notes: This table shows the relationship between cellular coverage and labor market outcomes, using data from the 2000 and 2004 DHS surveys. Future Cell Coverage indicates that the TA gets coverage within 3 years after the 2004 survey. *** p<0.01, ** p<0.05, * p<0.1Robust standard errors in parentheses Regressions are weighted using DHS weights, controls for DHS year, individual age, years of education and marital status are also included.

Table 7: Cellular phone coverage and women's employment by demographic variables: OLS regressions

	Coeff. on PHONE	std. err.	N	R2
Panel A: Working at all in past year				
Women is under 30	0.113***	(0.041)	20,735	0.11
Women is at least 30	0.102**	(0.045)	12,857	0.11
No education	0.119***	(0.043)	7,878	0.10
Education: 1 to 8 years	0.122**	(0.049)	5,651	0.11
Education: more than 8 years	0.100**	(0.045)	13,108	0.12
Married	0.123*	(0.068)	1,976	0.27
Unmarried	0.121***	(0.043)	14,302	0.10
No children at home	0.08	(0.050)	6,433	0.16
Children at home	0.143***	(0.051)	4,524	0.13
Panel B: Conditional on working: Working away from home				
	Coeff. on PHONE	**std. err.**	**N**	**R2**
Women is less than 30	0.091***	(0.034)	12,794	0.10
Women is at least 30	0.111***	(0.041)	6,736	0.12
No education	0.131***	(0.046)	5,206	0.12
Education: 1 to 8 years	0.077	(0.049)	3,484	0.17
Education: more than 8 years	0.147***	(0.044)	7,445	0.11
Married	0.094	(0.111)	1,013	0.28
Unmarried	0.119***	(0.042)	8,706	0.12
No children at home	0.115**	(0.055)	3,236	0.14
Children at home	0.191***	(0.059)	2,577	0.19
Panel C: Conditional on working: Working for self				
	Coeff. on PHONE	**std. err.**	**N**	**R2**
Women is less than 30	-0.015	(0.035)	12,853	0.10
Women is at least 30	-0.029	(0.036)	7,305	0.12
No education	0.01	(0.045)	5,548	0.12
Education: 1 to 8 years	0.02	(0.046)	3,722	0.15
Education: more than 8 years	-0.032	(0.040)	8,056	0.10
Married	0.031	(0.092)	1,075	0.27
Unmarried	0.018	(0.042)	9,347	0.11
No children at home	-0.120**	(0.047)	3,506	0.20
Children at home	-0.094*	(0.052)	2,806	0.18
Panel D: Conditional on working: Working in agriculture				
	Coeff. on PHONE	**std. err.**	**N**	**R2**
Women is less than 30	0.048	(0.042)	12,873	0.16
Women is at least 30	0.033	(0.047)	6,778	0.18
No education	0.06	(0.052)	5,243	0.19
Education: 1 to 8 years	0.061	(0.040)	3,517	0.14
Education: more than 8 years	0.027	(0.053)	7,489	0.14
Married	0.152	(0.116)	1,015	0.46
Unmarried	0.047	(0.047)	8,771	0.17
No children at home	0.041	(0.059)	3,250	0.23
Children at home	0.032	(0.060)	2,589	0.25

Notes: Table shows relationship between cellular phone coverage and labor market outcomes, using data from the 2000 and 20004 DHS surveys, for different sub-groups of the data. Panels B-D outcomes are conditional on working at all.Unit of observation is the individual, regressions are weighted using DHS sample weights, other controls include age, years of education, TA level fixed effects and a constant.Robust s.e. in parentheses, clustered at the TA-year level.***p<0.01,**p<0.05,*p<0.1