GSMA Intelligence

Mobile Connectivity Index: Methodology

July 2017

© GSMA Intelligence

gsmaintelligence.com • info@gsmaintelligence.com • @GSMAi

Introduction

About the Mobile Connectivity Index

The Mobile Connectivity Index measures and tracks enablers of mobile internet connectivity. The Index has been constructed according to the steps set out in the guidelines developed by the OECD and the Joint Research Centre (JRC).[1] This methodology for the Index presents the theoretical framework that underpins the Index; the process for selecting the indicators, along with how they are structured; the approach used to normalise the data; the weights used in the Index; and the approach to aggregation.

Theoretical framework

What is measured?

The Index measures the enablers of mobile internet connectivity. It is therefore an input index. An input index measures a number of indicators that lead to an important outcome, in this case mobile internet connectivity. An input index is therefore distinct from an output index. In the context of mobile connectivity, an output index might seek to measure the intensity and diversity of mobile internet usage. It would seek to measure and understand how (or how much) people are using mobile internet services. By contrast, an input index seeks to measure and understand why people are not using mobile internet services.

Why is an index necessary?

There is no single barrier or enabler to mobile connectivity; rather, a number of prerequisites are necessary for a country's population to use mobile internet services. An index is required because it measures multiple enablers and can summarise complex and multi-dimensional realities.

A number of indices exist in the ICT sector, including:

- Affordability Index (Alliance for Affordable Internet)
- Networked Society City Index (Ericsson)
- Digital Economy & Society Index (European Union)
- Global Connectivity Index (Huawei)
- Broadband Development Index (IDBA)
- ICT Development Index (ITU)
- Barriers to Internet Adoption (McKinsey)

[1] Handbook on constructing composite indicators: methodology and user guide, OECD and JRC, 2008

- Networked Readiness Index (WEF)
- The Web Index (World Wide Web Foundation).

The Mobile Connectivity Index has been designed to ensure it does not replicate any of these or other related indices. In this respect, the index has four key characteristics that together distinguish it from other indices:

- It focuses specifically on mobile connectivity rather than internet connectivity in general (including fixed). Given that the digital inclusion gap in the developing world is expected to be addressed to a significant extent by mobile, it is important to understand the enablers of mobile connectivity specifically.
- As the index is focused on mobile connectivity, the majority of the underlying indicators are unique to the Mobile Connectivity Index and are either not available or not used in other indices (for example, spectrum availability and taxation on mobile services).
- It is an input index that seeks to measure the performance of countries against a set of key enabling characteristics, rather than an output index that measures internet take-up and usage.
- It is a global index, encompassing 150 countries that account for more than 98% of the world's population.

How are the enablers measured?

The enablers of mobile internet connectivity that inform the indicators selected for the Index are:

- 1) Infrastructure – the availability of high-performance mobile internet network coverage.
- 2) Affordability – the availability of mobile services and devices at price points that reflect the level of income across a national population.
- 3) Consumer readiness – citizens with the awareness and skills needed to value and use the internet, and a cultural environment that promotes gender equality.
- 4) Content – the availability of online content and services accessible and relevant to the local population.

Data selection

As the Mobile Connectivity Index is an input index, it is important that each indicator is an 'input' for mobile connectivity rather than an output or outcome (e.g. measuring the level of take-up). It is also important to develop a set of criteria against which each indicator can be considered for inclusion in the Index. The following criteria have therefore been used, based on guidance from the JRC and OECD.

- Relevance: the indicator should measure a barrier or an enabler in the take-up of mobile internet services.
- Accuracy: the indicator should correctly estimate or describe the quantities or characteristics they are designed to measure.
- Coverage: the data should cover as many countries as possible, as the Index is intended to be a global index. An indicator is not included if there is missing data on more than 25% of countries in the Index.
- Timeliness: the data should be collected consistently over time.

A key consideration in the assessment of accuracy is to include, to the greatest extent possible, 'hard' indicators that are objective and can be quantified. These are distinct from 'soft' indicators that are usually based on qualitative data from surveys or case studies. Such indicators are typically used to measure things that are difficult to quantify such as the quality of governance and corruption. Although soft indicators are very useful for some indices, particularly those where hard indicators are difficult to develop, they are not used in the Mobile Connectivity Index. This is to ensure that countries have objective benchmarks on which to target improved performance.

Although the indicators included in the Mobile Connectivity Index have all been carefully chosen based on the above criteria, there are some cases where data constraints require the use of proxy indicators if it is not possible to perfectly measure a certain enabler. For example:

- Indicators such as international bandwidth per user, number of secure servers and number of IXPs are included as proxies for the quality of a country's core network. This is because end-to-end mobile services require a resilient and high-capacity backhaul and core network.
- There is currently no data comparing a large number of countries in the area of digital skills or awareness. More traditional skills indicators are therefore used to measure consumers' ability to effectively use and engage with digital technology (for example, literacy and years of schooling).
- Indicators measuring the availability of mobile-specific content primarily focus on smartphone applications due to the lack of comparable data across countries on other types of mobile content.

Although the vast majority of the indicators are highly correlated with mobile internet penetration, suggesting that on average they are associated with higher take-up, there may be specific countries where they work less well as proxy indicators. These indicators will therefore be reassessed going forward and, where they can be improved, incorporated into future versions of the Index.

Table 1 presents the indicators that make up the Index. The Index comprises four key enablers, which in turn comprise a number of dimensions. These dimensions are constructed by aggregating one or more indicators.

Table 1: Mobile Connectivity Index Indicators
Source: GSMA Intelligence

Enabler	Dimension	Indicator	Original unit of measurement	Source
Infrastructure	Mobile infrastructure	2G network coverage	% Population covered	ITU
		3G network coverage	% Population covered	GSMA Intelligence
		4G network coverage	% Population covered	GSMA Intelligence
		Years since 3G network launch	Years	GSMA Intelligence
	Network performance	Mobile download speeds	Mbps	Ookla's Speedtest Intelligence
		Mobile upoad speeds	Mbps	Ookla's Speedtest Intelligence
		Mobile latencies	Milliseconds	Ookla's Speedtest Intelligence
	Other enabling infrastructure	International bandwidth per user	Bits per second	ITU
		Number of secure servers	Secure servers per 1 million people	World Bank
		Access to electricity	% of population with access	IEA/WEO (a)
		Number of Internet exchange points (IXPs)	IXPs per 10 million people	Packet Clearing House
	Spectrum	Digital dividend spectrum (b)	MHz per operator (c)	GSMA Intelligence
		Other spectrum below 1 GHz	MHz per operator	GSMA Intelligence
		Spectrum in bands 1–3 GHz	MHz per operator	GSMA Intelligence

Affordability	Mobile tariffs	Cost of entry usage basket (100 MB)	% of GNI per capita	Tarifica
		Cost of medium usage basket (500 MB)	% of GNI per capita	Tarifica and ITU
		Cost of high usage basket (1 GB)	% of GNI per capita	Tarifica
	Handset price	Cost of entry-level handset	% of GNI per capita	Tarifica, Strategy Analytics
	Income	GNI per capita	US dollars (PPP)	World Bank
	Inequality	Atkinson Measure of Inequality in Income	Index value (0–100)	UNDP
	Taxation	Cost of taxation	Cost of tax as a % of TCMO (d)	GSMA
		Cost of mobile-specific taxation	Cost of mobile-specific taxes as a % of TCMO	GSMA
Consumer	Basic skills	Adult literacy rate	% of literate adult population (above 15 years old)	UN
		School life expectancy (e)	Years	UN
		Mean years of schooling (f)	Years	UN
		Tertiary enrolment rate	%	UN
	Gender equality (g)	Gender literacy ratio	Female/male ratio	UN
		Gender years of schooling ratio	Female/male ratio	UN
		Gender bank account ratio	Female/male ratio	World Bank Global Findex
		Gender labour participation ratio	Female/male ratio	ILO
		Gender GNI per capita ratio	Female/male ratio	UN

Content	Local relevance	TLDs per capita (h)	Number of domains per person	ZookNIC
		E-government services (i)	Index value (0=worst, 1=best)	UN
		Mobile social media penetration	% of population	We are Social
		Mobile application development	Number of mobile apps developed per 1,000 mobile internet users	Appfigures and GSMA Intelligence
	Availability	App accessibility in first language	Proportion of population with mobile apps available in their first language	Appfigures, GSMA Intelligence and Ethnologue
		Number of apps accessible in first language	Average number of apps available to the population in their first language	Appfigures, GSMA Intelligence and Ethnologue
		Accessibility of top 400 ranked Google Play apps in any language	Average of the % of population that can use each app in the top 400 for that country	Appfigures, GSMA Intelligence and Ethnologue
		Accessibility of top 400 ranked Apple Store apps in any language	Average of the % of population that can use each app in the top 400 for that country	Appfigures, GSMA Intelligence and Ethnologue

(a) International Energy Agency/World Energy Outlook.

(b) Digital dividend spectrum refers to spectrum in 700 MHz and 800 MHz bands that are particularly well suited to achieving wider coverage.

(c) When constructing the metric on spectrum per operator, we exclude operators with very small spectrum holdings and market shares (e.g. operators only active in specific regions or that provide niche services).

(d) Total cost of mobile ownership.

(e) This is the total number of years of schooling (primary to tertiary) that a child can expect to receive given current enrolment rates. It is therefore a forward-looking indicator.

(f) This measures the average number of years of education received by people aged 25 and older, based on current attainment levels. It is different from school life expectancy because the latter is calculated using enrolment rates.

(g) Each of the indicators in this dimension is calculated by dividing he relevant female indicator (e.g. female literacy) by the relevant male indicator (e.g. male literacy).

(h) This includes the number of generic top-level domains (gTLDs) registered in a country and the number of registered country-code top-level domains (ccTLDs).

(i) This indicator uses the Online Service Index score in the UN's E-Government Survey.

Mobile tariffs and handset prices

Pricing data was provided by Tarifica, with retail prices captured as of the first quarter of 2017, including all relevant taxes.

Mobile tariffs

In order to produce comparable price metrics across countries, three baskets were defined based on usage allowance, contract and technology. The baskets were designed to capture entry or basic usage as well as more intense users. In order to construct the baskets, the following information was taken into account:

- Historic trends in average data consumption across countries, sourced from GSMA Intelligence, Ofcom, Tefficient and Opera. Data requirements going forward (which are likely to increase) were taken into account. We also gave due consideration to the fact that average values are often distorted by particularly intensive users of mobile services.
- A selection of allowances currently offered by operators in developed and emerging markets, provided by Tarifica.
- Baskets used in existing mobile pricing benchmark studies from OECD, Tarifica, Ofcom, EC and the ITU. These represent basket designs that are often used in economics literature that analyses pricing in the mobile industry.

The baskets resulting in from this analysis are described in Table 2.

Table 2: Usage basket profiles
Source: GSMA Intelligence

Usage allowance	100 MB data	500 MB data	250 minutes, 100 SMS, 1 GB data
Tariff	Prepaid	Any	Any
Technology	2G, 3G or 4G	3G or 4G	3G or 4G

Having defined these baskets, Tarifica researched all tariffs offered by mobile network operators in each country and selected the cheapest available plan under which the basket requirements could be met. In addition, the following guidelines were applied to ensure prices were representative of regular usage and consumption patterns:

- Prepaid plans lasting less than one month were included – in such cases, the usage allowance and price were scaled up to one month to ensure comparability across tariffs (e.g. the usage and price of a five-day plan were multiplied by 6 to derive a monthly usage and price).
- Short-term promotional offers were excluded.
- Plans targeted or restricted to certain profiles (e.g. youth, student, senior) were not included.
- Where a tariff included an initial one-off fee (e.g. activation, SIM card), this was amortised over a period of 24 months.

Prices were captured in local currencies for most countries. These were then converted into US dollars using exchange rates as of Q1 2017. This approach was used to obtain pricing data for 2016. To derive pricing data for 2014 and 2015, we incorporated data from the ITU, which has historically collected data on the 500 MB basket.[2] This gave a complete historic dataset for the medium usage basket. In order to estimate prices for the entry and high usage baskets, for each country we applied the same historic growth rates observed for the medium usage basket. This approach assumes that changes in tariff prices were the same across different levels of usage during the past three years. When the index is updated in 2018, we will have an additional year of data for each basket and so can adjust this assumption if necessary.

Handset price

As the Mobile Connectivity Index is focused on connecting the unconnected, we are primarily interested in measuring prices of entry-level handsets that allow users access to the internet rather than high-end devices that are sold at premium prices. In order to obtain this data, Tarifica researched the cheapest handset available in each market with internet-browsing capability – it could therefore be a smartphone[3] or a feature phone[4]. Given that the performance for basic internet mobile applications (such as basic video or social networking) is only functional with 3G and 4G, this analysis excluded devices with 2G and WAP connectivity. The fieldwork of collection of device prices was carried out by inspecting the devices available on the websites of all mobile network operators in each country; other retailers' websites were analysed for the countries where MNOs did not offer handsets. This approach was taken due to the resources required to inspect all non-MNO retailers in each country. It therefore means that in some markets there may be cheaper devices available – but by keeping the approach consistent across countries, the relative differences across countries should remain unbiased.

[2] See Measuring Information Society Reports. The ITU's approach to collecting pricing data differs slightly to the approach used by Tarifica, particularly because the ITU only collects data from the largest mobile operator whereas Tarifica considered all operators. However, we carried out an analysis of the changes in pricing from 2015 to 2016 and in the vast majority of cases, the changes were reasonable.

[3] A smartphone is a device that has an open operating platform (where new applications can be developed and installed by the user).

[4] A feature phone is a device with a closed platform.

As with mobile tariffs, prices were captured in local currencies for most countries. These were then converted into US dollars using exchange rates as of Q1 2017. This approach was used to obtain pricing data for 2016. To derive pricing data for 2014 and 2015, we used data from Strategy Analytics on the average selling price (ASP) of smartphones for each country and estimated historic prices by applying ASP growth rates to 2016 data. This approach assumes that changes in entry-level internet-enabled handset prices are similar to the average change in all smartphone prices. When the index is updated in 2018, we will have an additional year of data and so can adjust this assumption if necessary.

Taxation

The taxation indicators are developed by estimating the proportion of the total cost of mobile ownership (TCMO) that are: (i) accounted for by all taxes and (ii) accounted for by mobile-specific taxes.

The TCMO is calculated in monthly terms on the basis of three building blocks:

- Handset price. This represents a one-off cost that can be spread over the lifecycle of the device (after which it is assumed to be replaced). Handset prices were converted to a monthly price based on a handset lifecycle of three years for developing markets and two years for developed markets, in order to take into account differences in usage patterns, disposable income and willingness to pay. [5]
- The activation and connection price or any other charges incurred to connect to the MNO's network. For prepaid customers, this usually consists of an initial charge for activating the SIM card. For postpaid customers, there may be additional upfront costs, such as an initial charge for activating the number. Activation and connection prices were converted into monthly prices assuming they follow the lifetime of the device.
- The price related to use and comprised voice, SMS and data charges. This is already expressed in monthly terms.

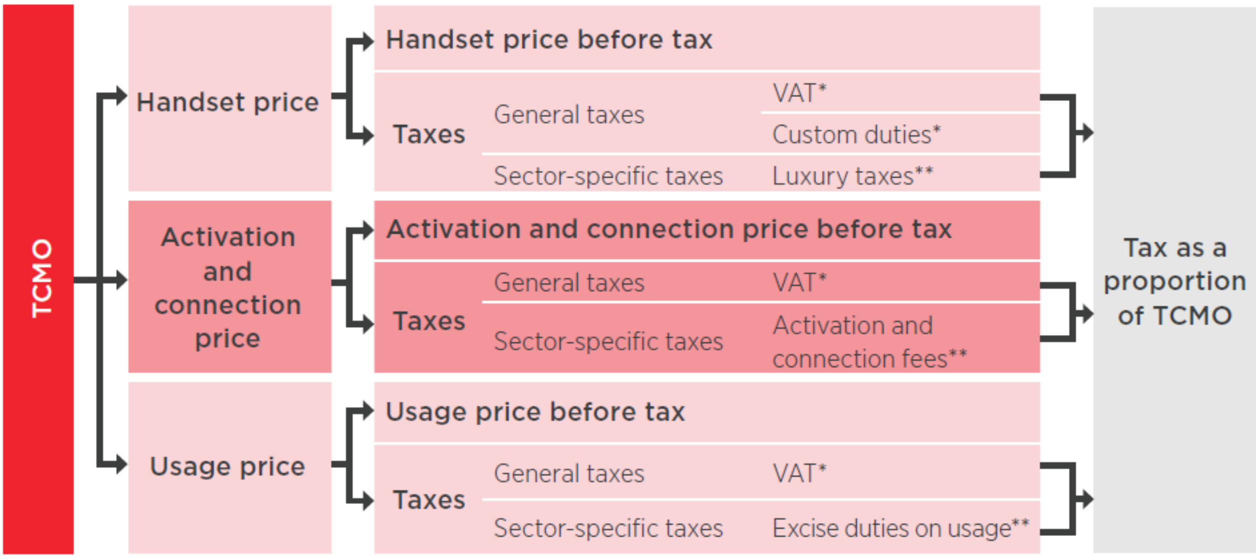
The TCMO was calculated for each basket b of country i as follows:

$$\text{TCMO}_{bi} = \frac{\text{Handset price}_i}{\text{Handset lifecycle}_i} + \frac{\text{Activation and connection price}_{bi}}{\text{Handset lifecycle}_i} + \text{Usage price}_{bi}$$

[5] Global Mobile Tax Review, GSMA and Deloitte, 2011

In order to calculate tax as a proportion of TCMO, taxes in Figure 1 were considered.

Figure 1: Calculation of proportion of tax in TCMO



* Ad valorem tax rates **Tax rates can either be ad valorem or fixed fees

Taxes in the TCMO were calculated by applying tax rates to the appropriate tax base.

- In the case of ad valorem tax rates (VAT and excise duties), the relevant tax base is the retail price of the relevant TCMO building block that was used.
- In the case of customs duties, the selected tax base was the retail price of the device building block in the TCMO.
- In the case of fixed amount taxes, activation and connection fees were applied on the value of the SIM card. For general fixed fees, tax payments were converted to a monthly level.

Estimates of the proportion of TCMO accounted for by all taxes and mobile-specific taxes were derived for all baskets (entry, medium and high). Subsequent analysis showed a very high correlation (above 0.95) between baskets and so only the entry basket has been used in the Index for the taxation indicators.

Availability of content

The Content enabler comprises two dimensions – local relevance and availability. The former measures the amount of content produced in a given country, including e-government services, web domains, social media[6] and mobile applications. These are included because content that is created or developed within a country is likely to be relevant to many of the people living there.

[6] This provides a platform to generate content that local populations need or are interested in.

However, many people consume content that is produced outside of their own country, so it is important to measure the extent to which this more widely available content is accessible and relevant to users. We therefore developed four indicators to assess whether a country's population has content they can understand and engage with. We did this by using language as a measure for accessibility and relevance – if an individual has content available in a language they speak then they are more likely to find the internet useful.

Each of the indicators required data on the languages spoken in each country, which is sourced from Ethnologue. Specifically, a dataset was provided that had a list of languages spoken in each country as well as data on the proportion of population speaking each language, both as a first language and in total.

The second data source is a list of all mobile applications (almost 10 million) – this includes applications available on Google Play, Apple Store, Windows, Amazon and other smaller platforms. This dataset was provided by Appfigures. For each application, information is provided on the languages and countries it is available in as well as the app category (gaming, education, health etc) and the year of release. Data was also provided on the top 400 ranked apps (based on downloads) for Google Play and Apple Store at the end of each year since 2014, giving an indication of content most popular in each country.

Using these data sources, we developed four indicators that assess the extent to which a country's population has content they can understand and engage with by mapping the languages spoken in each country against the language of mobile apps.

App accessibility in first language

This indicator measures the proportion of population that have apps available in their first language. For example, if 50% of a population speaks English as a first language and 50% speak Pashto, a country would score 50 on this indicator as there are currently very few mobile applications available in Pashto.

Number of apps accessible in first language

This indicator measures the average number of apps available to a population in their first language. If there were 1,000 apps available in English and 100 available in Pashto in the country, the weighted average is 0.5*1000 + 0.5*100=550. This complements the first indicator by allowing countries to score higher as the amount of accessible content increases. We also apply a logarithmic transformation so that scores increase more at the lower end of the distribution (i.e. increasing app availability from 1,000 to 2,000 results in a much larger score improvement that increasing from 1 million to 1.001 million).

The first two indicators measuring content availability focus on people's first language as this is likely to be their preferred choice of access and will incentivise greater take-up of mobile internet. The first indicator measures the availability of at least some content while the second measures the amount of content.

Accessibility of top 400 ranked apps on Google Play and Apple Store

For each app, we estimate the proportion of the population that are able to use it based on the languages it is available in. If an app is available in English, French and Hindi and 80% of a country's population speaks one of these languages (either as a first or second language), we assume that the app is accessible to 80% of the population. We then take the average of the top 400 ranked apps in each store.

These indicators focus on the most popular apps available in a country and measure the proportion of the population that can use them, whether in their first or second languages. This complements the first two indicators because although content in someone's first language is likely to be preferred, in many countries a significant proportion of the population speak second languages in which content is more readily available (e.g. English and French in many African countries).

Data treatment

Having obtained data and carried out the necessary calculations for the above indicators, we check to ensure that each country has data on at least 75% of indicators overall and at least half the indicators within each enabler. This ensures that a significant proportion of data for a country is not imputed. Similarly, we also ensure there is data for at least 75% of countries for each indicator.

The next step is to then treat the data, dealing with outliers and imputing missing data. If data is skewed by certain outliers, this could impact the overall index scores (for example, a country with exceptionally low 2G network coverage compared to all other countries will score very low but will also cause all other countries to score relatively highly with little variation). In order to identify outliers, indicators are assessed to see if they have an absolute skewness above 2 and kurtosis above 3.5[7]. Where these thresholds are met, one of two treatment approaches is adopted:

- Winsorisation – outlier variables are trimmed to the nearest value until the indicator is brought within the specified ranges for skewness and kurtosis. For example, if a country has an outlier value of 1,000 and the next highest value is 90, the former is trimmed to 90. If this gives acceptable skewness and kurtosis scores, the process stops there. If not, the two values are trimmed to the next highest value (which might be 80 in the above example). This process is continued until the indicator falls within the specified skewness and kurtosis ranges. In order to ensure that a large number of observations are not adjusted, a maximum of six observations are trimmed. If this still isn't sufficient to reduce skewness and kurtosis, the second approach is implemented.
- Transformation – as the majority of the indicators with high skewness and kurtosis are skewed to the right, a logarithmic transformation is used to bring the indicator within the specified ranges.

[7] These thresholds are generally used in identifying outliers for composite indices.

There are a few indicators where a logarithmic transformation is applied even though Winsorisation would suffice. This is because a logarithmic transformation has a conceptual benefit in that it results in improvements in the lower end of the indicator distribution being more 'beneficial' to a country than improvements at the high end of the distribution. An example of this is in relation to GNI per capita. Increasing average incomes from $1,000 to $2,000 per year is likely to have a bigger impact on mobile affordability than increasing from $100,000 to $101,000, so – from the perspective of the Index – should be rewarded with a higher increase. Logarithmic transformation achieves this.

A logarithmic transformation has been applied to the following indicators, for either data treatment or conceptual reasons:

- international bandwidth per user
- number of secure servers
- cost of entry usage basket
- cost of medium usage basket
- cost of high usage basket
- GNI per capita
- TLDs per capita
- mobile application development
- number of apps accessible in first language.

The next step in the data treatment process requires the imputation of missing data. Where data is missing, historic information is used before implementing a modelling-based approach. For data that is generally updated annually, we use the previous year's value where the latter is available and the current year is missing. This is used for indicators such as GNI per capita and number of servers. This is likely to result in a more accurate estimate for a specific country than using a modelled or imputed value based on data for other countries. However, if there is no data for the current or previous year, then historic values are not used because indicators such as GNI per capita and number of servers can vary significantly over two years and so using data that is older than one year will be subject to greater inaccuracy.

For some of the indicators, data is only updated every few years (or sometimes longer) if it is not expected to significantly vary year-to-year and/or if collecting the data is particularly complex. This applies to the following indicators in the Mobile Connectivity Index:

- Access to electricity
- inequality in income
- Education indicators and their gender ratio counterparts.

For these variables, if data exists in the period since 2011 then the most recent value is used. Otherwise, it is imputed using the methods described below.

The remaining missing data is imputed with a multivariate normal (MVN) data augmentation approach that uses multiple imputation. The MVN method generates imputed values assuming an underlying joint multivariate normal model.[8] In order to account for variation caused by missing data, the model is run 20 times – the average of these 20 imputations is then used to impute the missing value. To ensure the Index rankings are robust to the imputation method, missing values were also imputed using a hot deck imputation approach[9] and a multiple imputation method based on predictive mean matching[10] (PMM). We found that no countries moved more than 10 places in the 2016 rankings when using the PMM and hot deck imputation approaches. This shows that the Index is not particularly sensitive to the imputation methodology used for missing data.

Normalisation

Normalisation is required in an index to adjust for different units of measurement and different ranges of variation across the indicators. For the Mobile Connectivity Index the min-max method is used, which transforms all indicators so they lie within a range between 0 and 100 using the following formula:

$$I_{q,c} = \frac{x_{q,c} - min_c(x_q)}{max_c(x_q) - min_c(x_q)}$$

Where 'I' is the normalised min-max value, 'x' represents the actual value and the subscripts 'q' and 'c' represent the indicator and country respectively.

This method has been chosen over alternatives such as rankings and categorical scales because it retains interval-level information. For example, in the case of ranking 3G coverage, Country A might have 100%, Country B might have 99% and Country C might have 90%. These would be ranked in order as 1, 2 and 3 respectively (or they may all be categorised as having the highest score on an ordinal scale). However, this doesn't take into account the differences between the two – specifically the fact that B is much closer to A than it is to C. Furthermore, as the Mobile Connectivity Index is updated over time, a ranking-

[8] Even if the normality assumption is violated the MVN approach has still been shown to lead to reliable estimates given a sufficient sample size.
[9] This estimates missing values by using the value of the country that is mathematically closest to it. It is implemented by identifying indicators with high correlation with the indicator with missing data. These are then used to calculate the Mahalanobis distance to all other countries. The country with the smallest distance is identified as the nearest neighbour and data is imputed using that country.
[10] This generates estimates of missing values using a regression model; the independent explanatory variables are selected if they have a high correlation with the variable being imputed. However, for a number of the indicators, imputing a value by regression produces results that are not valid – for example, negative download speeds, coverage figures greater than 100% and negative prices. A predictive mean matching (PMM) approach is therefore applied. This generates an estimated value using the regression for a country that is missing data and then matches it with the country with the closest regression output. The actual value of that country is then taken. As with the MVN approach, in order to account for variation caused by missing data, the regression is run 20 times with slightly different coefficients. The average of these 20 estimates is then used to impute the missing value.

based approach may not track a country's progress as well as min-max or standardisation because a country might improve its coverage without increasing its rank.

For most indicators, the minimum and maximum used for normalisation are based on the actual minimum and maximum for that indicator, although in some cases they have been amended. For example the gender indicators, which represent female/male ratios, have a maximum threshold of 1 as this represents gender equality. Any country with a value greater than this is therefore not rewarded with a higher score.

To allow for comparisons of index scores over time, the minimum and maximum for each indicator are fixed. Some of the indicator maxima have therefore been adjusted where there are likely to be increases during the next few years in order to give all countries room to improve. These adjustments are based on an analysis of historic data and statistical analysis (ensuring that the maxima do not significantly exceed a threshold of being two standard deviations above the mean).

As part of the normalisation process, all indicators are also transformed such that they have the same orientation – i.e. a higher score always represents a 'better' score. This is necessary for indicators that are negatively correlated with mobile internet penetration – for example, mobile tariffs, income inequality and latency.

To ensure the Index is robust to the normalisation methodology, Index scores were also calculated by normalising indicators using 'z-scores'. This transforms all indicators such that they have a mean of 0 and a standard deviation of 1. The Index rankings are robust to the normalisation method, with only one country moving more than 10 places if z-scores are used.

Weightings

To construct the weights at the dimension, enabler and overall index level, a number of considerations have been taken into account, including the following:

- Statistical relationship between indicators and dimensions with mobile internet penetration – this includes both correlation and regression analysis.
- Analysis of consumer survey responses regarding perceived barriers to mobile internet access.
- Principal component analysis – this identifies weights that correct for the overlapping information implied by grouping indicators that are correlated (rather than representing a measure of importance).
- Research carried out by the GSMA and other organisations on digital inclusion and barriers to mobile connectivity.
- Qualitative evidence and expert opinion within the GSMA.

Based on this, the following weights have been used for the dimensions (Table 3) and enablers (Table 4).

Table 3: Indicator weights for dimensions
Source: GSMA Intelligence

Dimension	Indicator	Indicator weights
Mobile infrastructure	2G network coverage	20%
	3G network coverage	30%
	4G network coverage	25%
	Years since 3G network launch	25%
Network performance	Mobile download speeds	25%
	Mobile upload speeds	25%
	Mobile latencies	50%
Other enabling infrastructure	Access to electricity	35%
	Number of secure servers	25%
	International bandwidth per user	25%
	Number of IXPs per 10 million people	15%
Spectrum	Digital Dividend spectrum	45%
	Other spectrum below 1 GHz	20%
	Spectrum in bands 1–3 GHz	35%
Mobile tariffs	Cost of entry usage basket (100 MB)	40%
	Cost of medium usage basket (500 MB)	40%
	Cost of high usage basket (1 GB)	20%
Handset price	Cost of entry-level handset	100%
Income	GNI per capita	100%
Inequality	Atkinson Measure of Inequality in Income	100%
Taxation	Cost of taxation	50%
	Cost of mobile-specific taxation	50%
Basic skills	Adult literacy rate	25%
	School life expectancy	25%
	Mean years of schooling	25%
	Tertiary enrolment rate	25%
Gender equality	Gender literacy ratio	25%
	Gender mean years of schooling ratio	30%
	Gender account ratio	25%
	Gender labour participation ratio	10%
	Gender GNI per capita ratio	10%
Local relevance	TLDs per capita	20%
	E-government services	20%
	Mobile social media penetration	30%
	Mobile application development	30%
Availability	App accessibility in first language	35%
	Number of apps accessible in first language	35%
	Accessibility of top 400 ranked Google Play apps in any language	15%
	Accessibility of top 400 ranked Apple Store apps in any language	15%

Table 4: Dimension weights for enablers
Source: GSMA Intelligence

Enabler	Dimension	Dimension weight
Infrastructure	Mobile infrastructure	30%
	Network performance	30%
	Other enabling infrastructure	20%
	Spectrum	20%
Affordability	Mobile tariffs	20%
	Handset price	20%
	Income	20%
	Inequality	20%
	Taxation	20%
Consumer	Basic skills	50%
	Gender equality	50%
Content	Local relevance	50%
	Availability	50%

In terms of weighting the enablers for the Index, equal weights are assigned – i.e. each enabler is given a weight of 25%. Table 4 shows the Pearson and Spearman ranking correlation coefficients between the enablers and final index score against mobile internet penetration, demonstrating a high correlation across all enablers.

Table 5 Correlation coefficients with mobile internet penetration
Source: GSMA Intelligence

Enabler/index	Pearson correlation	Spearman correlation
Infrastructure	0.87	0.88
Affordability	0.84	0.85
Consumer	0.75	0.77
Content	0.83	0.84
Final index score	0.89	0.90

An analysis was carried out to assess the impact of adjusting these weights on the correlation between the overall index score and mobile internet penetration rates, including analysis that set weights to optimise both correlation coefficients. Such changes make very small improvements to the final index-penetration correlation (less than 0.01). Equal weights are therefore appropriate.

Aggregation

Two methods of aggregation were considered: arithmetic and geometric. The key consideration when choosing between these is the extent to which indicators, dimensions and enablers are substitutable, with arithmetic aggregation implying perfect substitutability and geometric implying partial substitutability.

At the lower levels of the Mobile Connectivity Index, there is often a greater degree of substitutability than at the higher levels. For example, within the Mobile Infrastructure dimension low 3G network coverage can be compensated by high 4G network coverage. In the Mobile Tariffs dimension, an expensive price for the medium basket could be compensated by a cheap entry basket price. At the index level, such substitutability is unlikely to be perfect – a country with a high infrastructure score is unlikely to achieve high mobile internet penetration if mobile is completely unaffordable or if there is no relevant content. The enabler groups sit somewhere in-between – there is likely to be more substitutability than the index level (e.g. high handset price might be compensated by a low tariff price) but less than at the dimension level (e.g. poor mobile coverage is unlikely to be compensated with high network performance). With this in mind, we have adopted the following aggregation rules:

- dimension aggregation – arithmetic
- enabler aggregation – arithmetic
- index aggregation – geometric.

If geometric aggregation is used at the enabler level, eight countries move more than 10 places in the 2016 rankings. If arithmetic aggregation is used at all levels, no countries move more than 10 places. This shows that the Index rankings are robust to the method of aggregation.

About GSMA Intelligence

GSMA Intelligence is the definitive source of global mobile operator data, analysis and forecasts, and publisher of authoritative industry reports and research.

Our data covers every operator group, network and MVNO in every country worldwide – from Afghanistan to Zimbabwe. It is the most accurate and complete set of industry metrics available, comprising tens of millions of individual data points, updated daily. GSMA Intelligence is relied on by leading operators, vendors, regulators, financial institutions and third-party industry players, to support strategic decision-making and long-term investment planning. The data is used as an industry reference point and is frequently cited by the media and by the industry itself.

Our team of analysts and experts produce regular thought-leading research reports across a range of industry topics.

gsmaintelligence.com • info@gsmaintelligence.com • @GSMAi

Whilst every care is taken to ensure the accuracy of the information contained in this material, the facts, estimates and opinions stated are based on information and sources which, while we believe them to be reliable, are not guaranteed. In particular, it should not be relied upon as the sole source of reference in relation to the subject matter. No liability can be accepted by GSMA Intelligence, its directors or employees for any loss occasioned to any person or entity acting or failing to act as a result of anything contained in or omitted from the content of this material, or our conclusions as stated. The findings are GSMA Intelligence's current opinions; they are subject to change without notice. The views expressed may not be the same as those of the GSM Association. GSMA Intelligence has no obligation to update or amend the research or to let anyone know if our opinions change materially.

© GSMA Intelligence 2017. Unauthorised reproduction prohibited.

Please contact us at info@gsmaintelligence.com or visit gsmaintelligence.com. GSMA Intelligence does not reflect the views of the GSM Association, its subsidiaries or its members. GSMA Intelligence does not endorse companies or their products.

GSMA Intelligence, GSMA, The Walbrook Building, 25 Walbrook, London EC4N 8AF