

# Estimating Internet Users: An evidence-based alternative in the absence of survey data

POLICY BRIEF

With the growth of information and communication technologies (ICTs), and their enormous contribution to economic and social progress, there is an increasing demand for accurate measurement of ICT access and use. One of the core indicators proposed by the ITU to measure ICT use is ‘proportion of individuals using the Internet’. It is a base indicator that is used in many composite indices including NRI (Networked Readiness Index), Digital Economy Index (previously e-readiness Index), KEI (Knowledge Economy Index) and IDI (ICT Development Index). Errors in such base indicators ripple through the system and therefore need to be minimized. The recommended method of measuring the ‘proportion of individuals using the Internet’ is through demand-side surveys, however according to the ITU only 37.5% of countries have actually conducted demand-side surveys on ICT use.

There are significant shortcomings in the current method of estimating the proportion of Internet users in countries where demand-side data is unavailable. In the absence of demand-side surveys, governments calculate the proportion of Internet users based on the number of subscriptions, using a multiplier to account for people who use public Internet access points. ITU allows national administrations to use different multipliers at their discretion which are often not evidence based. The reported number of subscriptions is also often inaccurate due to mobile data not being reported in some countries and other countries reporting on potential subscriptions. Due to the importance of this indicator, and these problems, it is necessary to improve the ‘proportion of Internet user’ measurement method.

## RECOMMENDATIONS TO ITU

- 1. Use survey data whenever possible** to estimate the proportion of Internet users in a country, both from surveys conducted by the country and representative surveys conducted by regional organizations such as Research ICT Africa (RIA). If a country has conducted an ICT survey in a previous year, the Internet penetration for the current year should be imputed based on previous survey data.
- 2. Use income and education data to impute missing data:** Derive a model using income and education components of the Human Development Index (HDI) vs. Internet penetration rate for countries which have conducted a survey. Use this model to impute the Internet penetration rate of countries which have never conducted a survey based on their income and education level.
- 3. Use imputed data to override** .... If the Internet penetration rate provided by the country administrator is within +/- 10 percentage point band around the calculated estimate, the figure sent by the country should be used, else the imputed figure based on evidence should be used. This would preserve the current practice of using country data, except in cases where the figure seems unrealistic, in which case the country is recommended to conduct a demand-side survey.

## JUSTIFICATION

Hilbert & Peres conducted a multivariate discriminative analysis of ten attributes (including Education, Income, Household Size, Age, Gender and Ethnicity) in South American countries, testing for household Internet access penetration. They came to the conclusion that out of the variables tested the main two factors driving Internet penetration are income and education (Hilbert & Peres, 2010).

In addition, according to the ITU survey data, countries with high income are the ones with the highest Internet and computer penetration, they also

report that there is a correlation between Internet penetration and education. Therefore the possibility of imputing the Internet penetration rate for countries which have not conducted demand-side surveys, based on their income and education level was explored.

### I METHOD

A regression analysis was carried out on the actual Internet penetration rates (for countries where demand side surveys on Internet use has been conducted) and the Index created only using the Education and Income components of the Human Development

Index (HDI). The HDI was developed by United Nations Development Programme (UNDP) by combining indicators of life expectancy, educational attainment and income into a composite human development index.

A new index (HDI\_EdGNI) was created using only the education and income components of HDI since there is no evidence that Internet penetration is correlated with life expectancy. In HDI\_EdGNI, income and education have been given equal weightage. The education component of the HDI is measured by mean of years of schooling for adults and expected years of schooling for children of school going age. The income component is measured by GNI per capita (PPP\$). The HDI uses the logarithm of income, to reflect the diminishing importance of income with increasing GNI.

## II RESULTS AND DISCUSSION

Figure 1 shows the correlation between HDI\_EdGNI Index and the Internet penetration. A regression analysis was conducted which gives the best fit for the correlation, with adjusted R squared value of 0.85, as in equation 1. where  $y = 0.4 \times e^{6x}$  (1)

x = new index HDI\_EdGNI, y = % of Internet users

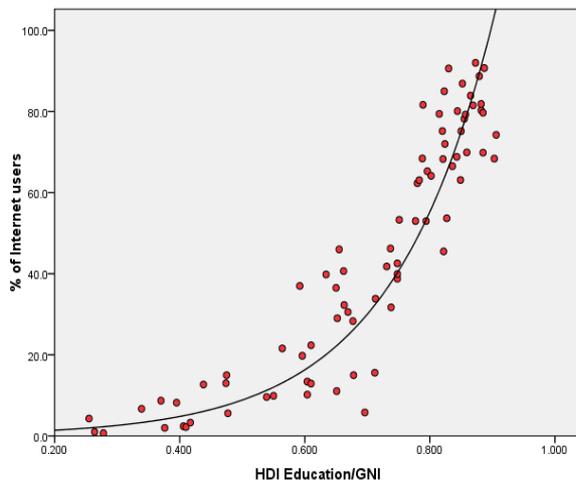


Figure 1: Internet penetration rate correlated with education and income components of HDI Index of countries for which demand-side survey results are available *Source: Authors*

A total of 78 countries have conducted ICT surveys out of 193 UN recognized countries. For countries which conducted a survey, the average variance between the model and survey was found to be around seven percentage points. Therefore it seems reasonable that the Internet penetration rate should fall within a band of +/- 10 percentage points from the figures calculated by the new model. Therefore it is recommended that for countries which have not conducted demand side surveys, the Internet penetration rate estimated by country administrations,

should be restricted to be within a +/- 10 percentage points band from the model.

The proposed method does not drastically change Internet penetration rates of most countries, but it is an intellectually defensible method and will be consistent across all countries without demand-side data. This method implements the often-stated principle that demand-side data is first best and therefore suggests including representative survey results from regional research organizations such as RIA, even if they are not identical to the results provided by the country administrations. It also removes the most egregious use of high multipliers and the problems of accurately estimating the total subscriptions in a country. It is based on income and education levels which have been considered by many researchers as the main drivers of Internet penetration rate.

It creates strong incentives for countries to conduct demand-side surveys because that is the only way they can escape the constraint of the mathematically derived estimate. If a national administration believes that it actually has implemented policies that have resulted in a higher proportion of Internet users than the model predicts, all it has to do is to conduct a demand-side survey to demonstrate its success.

## SOURCES

1. Gunaratne R. L. & Samarajiva R. (2012), Working paper titled "Estimation of Internet Users: An evidence based alternative in the absence of survey data.
2. ITU (2011a) Handbook for the collection of administrative data on telecommunications/ICT, Geneva: ITU
3. ITU (2011b) Measuring the Information Society, Geneva: ITU
4. Hilbert, M., & Peres, W. (Eds.). (2010). Information societies in Latin America and the Caribbean: Development of technologies and technologies for development. ECLAC Books. Santiago: United Nations Economic Commission for Latin America and the Caribbean.
5. Gillwald, A & Stork, C. (2008), Towards Evidence based ICT Policy and Regulation: ICT access and usage in Africa – Volume One 2008 paper two.
6. United Nations Development Programme (2011) Human Development Report 2011, Sustainability and Equity: A Better Future for All, New York: UNDP

## AUTHOR/S

Roshanthi Lucas Gunaratne, Senior Researcher, LIRNEasia, Sri Lanka, [roshanthi@lirneasia.net](mailto:roshanthi@lirneasia.net)

Rohan Samarajiva, Chair and CEO, LIRNEasia, Sri Lanka, [rohan@lirneasia.net](mailto:rohan@lirneasia.net)