**Aim:** The idea is to find the effect of technological advancements on GDP and what impact access to internet has on the whole phenomena.

**Target Variable**: Y - %Increase in GDP per capita.

**Independent Variable**:

X1 -Technology Independent Variables

X2 -High-technology exports (% of manufactured exports)

X3- High-technology exports (current US$)

X4-Technicians in R&D (per million people)

X5-Researchers in R&D (per million people)

X6-Trademark applications, total

X7-Trademark applications, direct resident

X8-Trademark applications, direct nonresident

X9-Patent applications, residents

X10-Patent applications, nonresidents

X11-Scientific and technical journal articles

X12-Research and development expenditure (% of GDP)

X13-Charges for the use of intellectual property, receipts (BoP, current US$)

X14-Charges for the use of intellectual property, payments (BoP, current US$)

X15/after-Time dummy variable (1 for 2007/2008 0 for 2017)

X16 - internet users as a % of total population

**Analogy to Incinerator Problem**

Target variable:

The target variable in the incinerator problem was the price of the house. Similarly, our target variable is GDP Growth.

Price of the house ⬄ GDP Growth

Independent variables:

The incinerator problem shows independent variables of size in sqft, no of bedrooms, etc. Our independent variables are Technology Independent Variables.

Measuring analogy: Building Incinerator ⬄ Internet Boom

In the incinerator problem, building of the incinerator, was similar for every sample. There was a stepwise similar rate of growth throughout all samples. For our GDP problem, the internet usage is not increasing in the same measurement for all countries in the years we are considering. This is because the internet boom happened at different time in each country (or samples) so “internet users as a % of total population” is not fixed over time. Hence, we can either study the effect on similar segments (only developing countries or developed countries), or we need to account for this difference using some method.

One Solution Could be to compare between two time periods where internet population grew drastically for one group and where it remained relatively low.

Distance from incinerator(Near,Far) ⬄ growth in Internet population(High i.e.200-300%, Low i.e.15-20%)

**Approach 1**

We can create Interaction variables for suitable Science and Technology Independent Variables.

For example.

X5(Researchers in R&D) \* X16 (internet users as a % of total population) and so on for all variables.

The co-efficient of this will tell us how the effect of “No of Researchers in R&D” on “GDP Growth” of a country changes for different values of internet users (X16).

**Equation:** GDP = b0 + a1 \* X1 \* X16 + …… + a14 \* X14 \* X16.

We can use all time variables (2000-2018) for this.

**Approach 2**

For Difference in Difference we have to define control and treatment group.

Control group – Developed Countries (Who did not have significant change in internet users from year1 – year2)

**Treatment group – Developing or Underdeveloped countries who had significant change in internet user from Year1-Year2)**

**Based on some observation I think year 1 should be (2007/2008) and Year 2 (2017). During this period there the developing countries saw around 300% increase in internet population, on the other hand developed countries had about 15-20% growth.**

**Switzerland - 2007/2008-2017(77/79-89)**

**Canada– 2007/2008-2017(73/76-91)**

**Germany – 2007/2008-2017(75/78-84)**

**USA - 2007/2008-2017(75/74-87)**

**Poland - 2007/2008-2017(48/53-74)**

**India - 2007/2008-2017(3/4 -34)**

**China - 2007/2008-2017(16-54)**

**Iran – 2007/2008-2017 (9.47/12.02-64)**

**Colombia 2007/2008-2017(21/25-62)**

**Argentina – 2007/2008-2017 (25/28-74)**

**Equation**: GDP = b0 + a1 \* after (2017) + b1 \* internet\_penetration (1 for developing countries, 0 for developed) + d1 after \* internet\_penetration

**Approach 3**

use panel data

Consider only developing countries.

For example.

X5(Researchers in R&D) \* X14 (1 for 2007/2008, 0 for 2017)

Equation

GDP = b0 + a1 \* X1 \* X5 + …… + a14 \* X14 \* X5.

**Questions**

Which Approach is better?

Can there be a better way to choose Year1 and Year2?

Which GDP to choose GDP Nominal or GDP PPP.

**UPDATE**

GDP is too unpredictable as there could be lots of other factors influencing it. Also Instead of Internet Usage affecting GDP, GDP could be a factor in increasing internet usage, and then it becomes very difficult to determine the causal effect.

So we are limiting our approach to how internet usage affected scientific research In different countries.

**Step 1**

**Scientific\_progress**(y)= Trademark\_applications\_total + Patent\_applications \_residents + Scientific\_and\_technical\_journal\_articles

**Internet\_Usage\_Interaction = Internet\_Penetration \* Y2008**

Y = B0 + B1 \* Y2008 + B2 \* Internet \_Penetration + D1 \* Internet\_Usage\_Interaction

Co-eficients: 'Y2008', 'Internet\_Users', 'Internet\_Usage\_Interaction'

30100.44, 4454.73 , -1373.52

**Step 2(Optional)**

Use binary variable Internet\_Penetration = 1 or 0 instead of Internet users

If internet users % is greater than say 80% , then Internet\_Penetration = 1

Though we loose some information, it’ll be intereting to see what the result look like.

**Step 3**

Use Panel data for all the years

Example – 17 dummy time variables for 2000-2017 (Y2000, Y2001…Y2017)

Questions – Do we need any interaction variable?

What should the equation look like?

Do we need to define control and treatment group? Or is there a scope to use difference in difference method.

If we use interaction variable, (**Internet Penetration \* Y2008**) then how do we interpret it since Internet Penetration is not binary?

How internet penetration affected scientific research over the years?

Or should the interaction term be (**Developed \* Y2008**), where Developed is binary?

But we lose information.

Interpretation: measure how the effect of internet penetration on scientific research is different in developed and developing countries.

In cross sectional data we need interaction variable to see how the overall effect of an independent variable has changed over time for example. (x\*y1987)

But for panel data we can have fixed effect for each observation with respect to time, so we don’t need interaction variables.

Instead now that we have different time period for each observation, first we calculate difference of each observation over time and then we regress on those differences. So the co-efficient of x (independent variable) now gives us the change in policy effect, which was given by co-efficient of Interaction variables in cross-sectional data.

We define fixed effect for each observation for each time period.

Are there control and treatment group exist in fixed effect?

Y(total research) = b0 + b1 \* x1 (internet penetration) + a1 (fixed effect for each country) + a2 (fixed effect for each year) + d1 \* Year2008