Project Report

Pillars of the Global Innovation Index by income level of economies: longitudinal data (2011-2022) for researchers' use

Project Report submitted for the course

Introduction to Statistics and Computation with Data

Nikhil Nagaria (bmat2227)

Rahul Vishwakarma (bmat2230)

Shriyaa Srivastava (bmat2238)

Maulik Pratap Singhal (bmat2123)

Abstract

This is an attempt to replicate and extend the analyses presented in the paper "Pillars of the Global Innovation Index by income level of economies: longitudinal data (2011-2022) for researchers' use" by Gonçalo Rodrigues Brás. This report provides a summary of the methods we used and the sources we used to complete this project. A descriptive analysis is provided and inferences about importance of factors in the GII are made using regression.

Keywords. Innovation, Global Innovation Index, Income Level, Panel Data, Innovation Data, Regression

Contents

1. Introduction

Innovation is important to drive economic progress for both developing and developed countries. Many governments are putting innovation at the centre of their growth strategies. In that light, measuring innovation and providing a rigorous statistical benchmark that attempts to capture the richness of innovation in society is mandate. Innovation is not just about the traditional ideas of research output or R&D laboratories. It includes technological advancements, social models, market and business systems and much more. To capture all these aspects of innovation, Professor Soumitra Dutta launched the GII project in 2007 during his tenure at INSEAD. It has grown considerably since then and gets updated every year to reflect the improved availability of statistics and a better understanding of innovation. Since 2011 though, the changes have been quite minor and the metrics have stabilised. The original paper [3] provides a panel data file [2] with all GII pillars from 2011 to 2022, including scores and ranks, and income data of the economies from 2011 to 2021.

2. About the Data

The GII relies on two sub-indices –
1. *Innovation Input Sub-Index* which includes five pillars capturing elements that enable innovative activities and
2. *Innovation Output Sub-Index* which has two pillars and includes the results of the innovative activities.

The overall GII score is the average of the two sub-indices, which are calculated as the average of the pillars they are comprised of. Each of the seven pillars is divided into three sub-pillars, each of which is composed of individual indicators. The values of these indicators are normalised using several ways for example, by an appropriate scaling $(x \mapsto (x - x_{\min}) / (x_{\max} - x_{\min}))$ or by dividing out by purchasing power parity GDP (in billions), or dividing by population size with some scaling etc. That gives the values for the seven pillars as numbers from 0 to 100, which are then used to compute the GII.

There was a fourth parameter called the *Innovation Efficiency Ratio* [9] which is defined as the ratio of the Output Sub-Index to the Input Sub-Index. It was discontinued in 2019 [10]. Regardless, it is still a very useful measure and can be computed easily (it is not available in the data file).

A brief description of each of the pillars and sub-pillars, along with the corresponding indicators follows.

2.1 The Pillars

Pillars 1-5 comprise the Input Sub-Index, while Pillars 6 and 7 form the Output Sub-Index. Minor changes are made to the indicators each year. The following descriptions are according to the latest GII report of 2022 [12].

Pillar 1: Institutions (I)
An institutional framework that attracts business and fosters growth by good governance and appropriate incentives is essential to innovation. This pillar captures the institutional framework of an economy.

1. *Political Enviroment* – It has two indices: the first is the political, legal, operational or security risk index that measures the likelihood of political, legal, operational or security risks affecting business operations and the second one is the government effectiveness index that reflects the perception of the quality of public and civil services and its independence from political pressures, the quality of policy implementation and the governance around it.
2. *Regulatory Environment* – It has three indices: the first is the regulatory quality index that reflects the perception of the ability of the government to implement sound policies that promote private-sector development, the second is the rule of law index that reflects the amount of confidence in the rules of the society, the police and the courts, and the third is the cost of redundancy dismissal that measures the cost of notice requirements and due severance payments when terminating a redundant worker's employment.
3. *Business Environment* – It has two indices: the first one measures the extent to which the government ensure a stable policy for doing business and the second one measures the perception of experts on entrepreneurial policies and culture.

Pillar 2: Humans, Capital and Research (HC&R)
The standard of education and research activity in an economy are essential to innovation. This pillar tries to gauge the human capital of economies.

1. *Education* – It has five indices measuring the government expenditure on education, government funding per secondary pupil, school life expectancy (that is, the total number of years that a person of school entrance age can expect to spend in the school), PISA scales in reading, maths and science, and the student-to-teacher ratio for secondary school.
2. *Tertiary Education* – It has three indices measuring tertiary school enrollment, proportion of STEM graduates from tertiary graduates, and the tertiary inbound mobility that measures the number of students from abroad.
3. *Research and Development (R&D)* – It has four indices: the proportion of full-time researchers in R&D, the gross expenditure on R&D (GERD), the average expenditure of the top three global companies in the country and the average scores of the top three universities according to the QS world university ranking.

Pillar 3: Infrastructure (Inf)
This pillar measures the infrastructure of an economy and has the followng sub-pillars.

1. *Information and Communication Technologies (ICTs)* – It has four indices: the ICT access index which covers various factors like proportion of the population with access to internet and telephone services, internet bandwidth, number of mobile cellular telephone subscriptions etc., the ICT use index which covers factors like proportion of people using the internet and mobile broadband subscriptions and internet traffic, the government online service index which measures the use of ICTs in delivering government services at the national level, and the e-participation index that measures the use of online services in providing infor-

mation to citizens, interacting with stakeholders and engaging in decision-making processes.

2. *General Infrastructure* – It has three indices: the electricity output, logistics performance index that includes several factors such as ease in international shipments, the timeliness of shipments, the quality of trade and transport infrastructure itself, etc., and the gross capital formation as a percentage of the GDP in local currency.
3. *Ecological Sustainability* – It has three indices: GDP (measured by PPP) per total energy supply measuring the energy productivity of the country, the environmental performance index that measures how close countries are to achieve established environmental policy targets, and the ISO 14001 environment certificates of conformity about enviromental issues per billion PPP$ GDP.

Pillar 4: Market Sophistication (MS)
The availability of credit and an environment that supports investment, access to the international market, competition, and market scale are all critical for businesses to prosper and for innovation to occur. This pillar captures this aspect.

1. *Credit* – It has three indices that measure the perception of experts on finance for starting and growing firms, the domestic credit to the private sector which means the financial support to the private sector by financial corporations, and loans from all microfinance institutions as a percentage of the GDP.
2. *Investment* – It has four indices: market value of domestic companies, number of venture capital deals invested in, number of venture capital deals recieved and the total value of venture capital recieved.
3. *Trade, Diversification and Market Scale* – It has three indices: the weighted average applied tariff which is the average of effectively applied rates weighted by the product import shares corresponding to each partner country, the Herfindahl-Hirschman Index which measures the diversification in the industrial system across different sub-sectors, and the domestic market size of the country.

Pillar 5: Business Sophistication (BS)
The last input pillar tries to capture the level of business sophistication to assess how conducive firms are to innovation activity. Businesses foster the productivity of human capital, their competitiveness, and the innovation potential with the employment of highly qualified professionals and technicians.

1. *Knowledge Workers* – It has five indices that measure the employment in knowledge-intensive services, the number of firms offering formal training, the gross expenditure on R&D (GERD) performed by business enterprises, the GERD financed by business enterprises and the percentage of females employed with advanced degree.
2. *Innovation Linkages* – It has five indices: the extent to which businesses and universities collaborate on R&D, a measure of how much are firms, suppliers, producers and institutions in a particular field spread geographically, the GERD financed by abroad, the number of joint venture/strategic alliance deals and number of patent families (patent for the same invention filed in at least two intellectual property offices).
3. *Knowledge Absorption* – It also has five indices: charges for use of intellectual

property (payments), high-tech imports that contain technical products with a high intensity of R&D defined by the Eurostat classification, ICT service imports as a percentage of the total trade, FDI (foreign direct investment) net inflows which is the three-year average of net investment flows from a different economy to acquire a lasting management interests, and the amount of full-time researchers in the business enterprise.

Pillar 6: Knowledge and Technology Outputs (K&TO)
This pillar covers all those variables that are traditionally thought to be the fruits of inventions and/or innovations.

1. *Knowledge Creation* – It has five indices: the number of resident (domestic) patent applications filed at a patent office, number of international patent application filed through the WIPO-administered Patent Cooperation Treaty, number of resident utility model applications filed at the national patent office, number of scientific and technical journal articles and the H-index of the economy, measuring the productivity and citation impact of research articles.
2. *Knowledge Impact* – It has five indices: the growth rate of GDP per person employed, the new business density rate which is the number of newly registered firms with limited liability per 1000 working-age people per year, the percentage of GDP spent on computer software, number of ISO 9001 quality certificates, and proportion of high-tech and medium-high-tech manufacturing (on the basis of the OECD classification).
3. *Knowledge Diffusion* – It has four indices: charges for use of intellectual property (receipts), the economic complexity index of the country, high-tech exports as a percentage of total trade and ICT services exports as a percentage of total trade.

Pillar 7: Creative Outputs (CO)
The role of creativity for innovation is still largely underappreciated in innovation measurement and policy debates. The GII has always emphasized measuring creativity as part of its Innovation Output Sub-Index.

1. *Intangible Assets* – It has four indices: the average of the intangible asset value of the top 15 firms, the number of classes in resident trademark applications issued at a given office, the number of designs in resident trademark applications filed at a given office and the global brand value of the top 5000 brands.
2. *Creative Goods and Services* – It has five indices: the cultural and creative services exports, number of national feature films produced per million population (15-69 years old), global entertainment and media market per thousand population (15-69 years old), percentage of printing publications and other media output and creative goods exports as a percentage of total trade.
3. *Online Creativity* – It has four indices: the number of generic top-level domains (TLDs), the number of country-code TLDs, the number of GitHub commit pushes recieved and the number of global downloads of mobile apps.

2.2 Income Level Classification and the Atlas Method

The World Bank's official estimates of the size of economies are based on GNI converted to current US dollars using the Atlas method instead of simply using exchange rates.

The Atlas method smoothes out the exchange rate fluctuations by using a three year moving average and a price-adjusted conversion factor. It is calculated as the average of a country's exchange rate for that year and its exchange rate for the previous two years, adjusted for the difference between the inflation rate in the country and outside. The method is explained in brief.

The *GDP deflator* is a measure of inflation and is denoted by p_t. The inflation rate in a country is measured by the change in its GDP deflator. Thus, the inflation rate between year $t-n$ and t is given by $r_{t-n} = p_t/p_{t-n}$. On the other hand, the international inflation rate is measured by the change in the *SDR deflator* denoted by $p_t^{\text{SDR\$}}$. It is a weighted average of the GDP deflators (in SDR terms) of China, Japan, the United Kingdom, the United States and the Euro Area, weighted by the equivalent of a SDR unit in that currency, and finally converted to US dollars. Thus, international inflation rate between year $t-n$ and year t is given as $r_{t-n}^{\text{SDR\$}} = p_t^{\text{SDR\$}}/p_{t-n}^{\text{SDR\$}}$. Thus, denoting the exchange rate for year t by e_t, we get that the Atlas conversion factor is

$$e_t^{\text{atlas}} = \frac{1}{3}\left(e_t + e_{t-1}\frac{r_{t-1}}{r_{t-1}^{\text{SDR\$}}} + e_{t-2}\frac{r_{t-2}}{r_{t-2}^{\text{SDR\$}}}\right)$$

Finally, we can find the GNI in US dollars using Atlas method as the GNI in local currency divided by e_t^{atlas} and then divide it by country's midyear population to derive GNI per capita.

Based on the Atlas GNI per capita of an economy, the World Bank divides each economy into one of the four categories: low income, lower-middle income, upper-middle income and high income. Note that these thresholds are not invariant in time.

2.3 Purchasing Power Parity

Purchasing power is the amount of goods that can be bought with one unit of currency. *Purchasing power parity* is a measurement of the price of specific goods in different countries, and is used to compare the purchasing powers of different countries. It is based on the *law of one price*, which states that in absence of trade frictions (such as transport costs and tariffs) and under conditions of free competition and price flexibility, identical goods must sell for the same price in different locations.

To account for errors which may arise due to choosing specific goods, PPP uses a basket of goods which is essentially many goods with different quantities. PPP computes the PPP exchange rate as the ratio of the price of the basket of goods in two different locations. For example, if a basket consisting of a computer, a kilogram of rice, and half a ton of steel was 1000 US dollars in New York and the same goods cost 6000 HK dollars in Hong Kong, the PPP exchange rate would be 6 HK dollars for every 1 US dollar.

The PPP exchange rate may not match the market exchange rate, as the market rate is more volatile since it reacts to changes in demand at each location. Also, tariffs and differences in the price of labour can contribute to longer-term differences between the two rates. Due to PPP exchange rates being more stable over time and less affected by tariffs, they are used for many international comparisons especially when time is an important attribute, for example comparing GDP values.

2.4 Dataset

The dataset is a panel data file consisting of the values for each of the seven pillars and the GII as well as the rankings for 149 economies over the years 2011-2022. It also includes the GNI per capita in US dollars computed using the Atlas method and the GDP per capita based on purchasing power parity (constant 2017 international $) and on purchasing power parity (current international $). The file also includes the thresholds for classification of economies by income level for each year in a separate worksheet. Another worksheet has some incomplete data records which were removed from the main database while compiling.
The reason the data was only (and could be) compiled for the years 2011 to 2022 is that since 2011, the metrics and methods have quite stabilised despite minor changes each year. The name of the sixth pillar, for instance, was changed from "Scientific Outputs" in 2011 to "Knowledge & Technology Outputs" in 2012 onwards. The subpillars, however, were the same so the pillars are still comparable. Another major change is the introduction of the "Online Creativity" subpillar in the CO pillar in 2012 [8]. There were other changes in names of subpillars from 2011 which have then stayed the same till date. Those changes are as follows.

Pillar	**Old Name**	**Year of Change**	**Current Name**
Inf	Energy	2012	Ecological Sustainability
MS	Trade & Competition	2016	Trade, Diversification and Market Scale
CO	Creative Intangibles	2013	Intangible Assets

2.5 Data Collection

- The data for the years 2013 to 2022 were collected from the GII excel files using the World Intellectual Property Organization website [13].
- The data from 2011 and 2012 were based on the GII reports ([7] and [8]) and introduced manually in the database.
- Excel commands were used to merge all these data into one single file shaped into a panel data file in long format.
- There were some issues that had to be manually fixed, like different designations given to the same economy (for instance, "Cabo Verde" and "Cape Verde", and "Hong Kong, China" and "Hong Kong (China)") or the name of the economy changing entirely ("Swaziland" to "Eswatini" in 2016 and "The Former Yugoslav Republic of Macedonia" to "North Macedonia" in 2018).
- Some economies were included in excel data files but not in the reports, particularly for the years 2015 and 2016. Those were removed from the main table and put into the "incomplete data - removed" worksheet of the data file.
- After these procedures, only economies presented annually in GII reports were in the analysis.

- The income data (GNI per capita and PPP-based GDP) and GNI thresholds for income level classifications were directly imported from the World Bank database, and the classification was assigned to the economies (1 for low income, 2 for lower-middle income, 3 for upper-middle income and 4 for high income).
- However, the income parameters of the economies published by the WB were only available till 2021. This gave rise to some discrepancies like the case of Iraq and Mauritania that were introduced in the 2022 GII edition, so there are no GII or pillar values for these countries despite the availability of the income parameters.
- Evaluation and assessment procedures give consistency through comparisons between the data downloaded and GII reports. The gaps found (for example, even though Morocco did not show any rank in the file, it was in the 2014 GII report) were fixed manually.
- Several data controls were performed through random sampling and amendments were made using some Excel commands to reach a final panel database.

3. Descriptive Statistics

3.1 Summary of the Innovation Variables

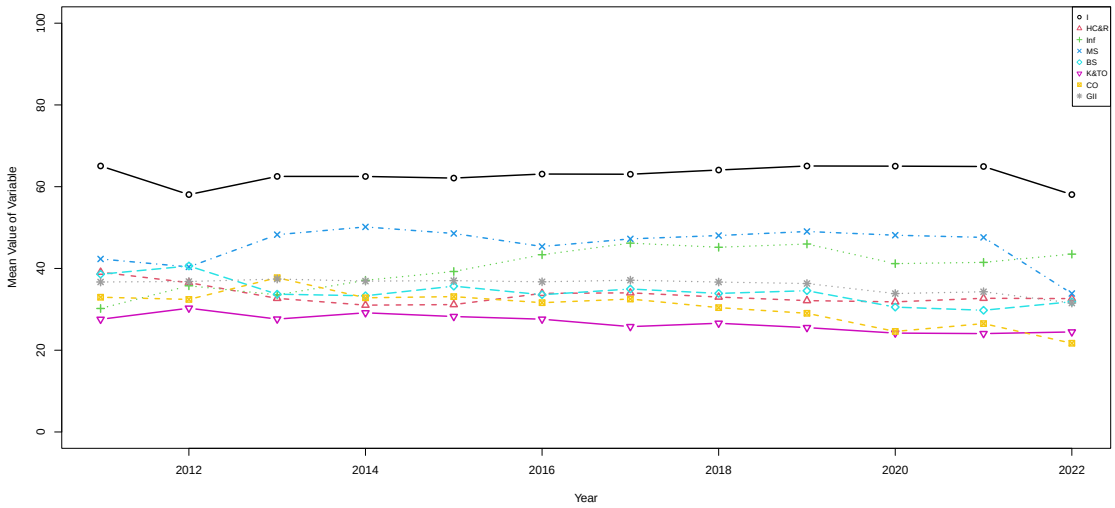
Variable	Mean	Median	Minimum	Maximum	SD	CV
I	62.74458	61.1	15.4	95.9	16.04216	0.255674
HC&R	33.35435	31.4	0.7	74.7	15.10492	0.452862
Inf	40.12361	39.5	6.2	69.9	13.61709	0.339378
MS	45.77514	44.9	4.4	88.6	12.82714	0.280221
BS	34.27232	31.8	8.6	79.1	12.56392	0.366591
K&TO	26.80582	23.9	1.6	74.9	13.64900	0.509180
CO	30.52912	29.4	0.3	73.7	13.77410	0.451179
GII	35.96286	33.6	11.6	68.4	12.06778	0.335562

Variable	Skewness	Ex. Kurtosis	5th %ile	95th %ile	Q1	Q3	IQR
I	0.132470	−0.623498	39.08	90.40	50.8	74.6	23.8
HC&R	0.369223	−0.717770	11.78	60.72	21.1	43.8	22.7
Inf	0.117774	−0.972272	19.58	62.60	28.7	51.1	22.4
MS	0.332069	+0.586109	26.10	68.40	37.7	53.0	15.3
BS	0.680202	−0.141202	17.68	58.10	24.9	42.0	17.1
K&TO	0.807230	+0.156570	8.90	54.80	16.9	34.5	17.6
CO	0.225779	−0.358962	8.68	53.92	20.4	40.1	19.7
GII	0.578352	−0.520494	19.80	58.20	27.0	43.4	16.4

Briefly summarizing the measures of central tendency and variability presented in the above table, the mean and median measures of the variable “Institutions” (I) presents

far higher values in relation to other variables; it also presents greater variability in comparison to other variables with a standard deviation of 16.042. But there is no significant difference in interquartile range.
Regarding univariate normality, based on kurtosis and skewness criteria defined by Huck [4] in which normality may be indicated up to absolute value of 1, this table suggests the absence of severe deviations from normality in data generating process.

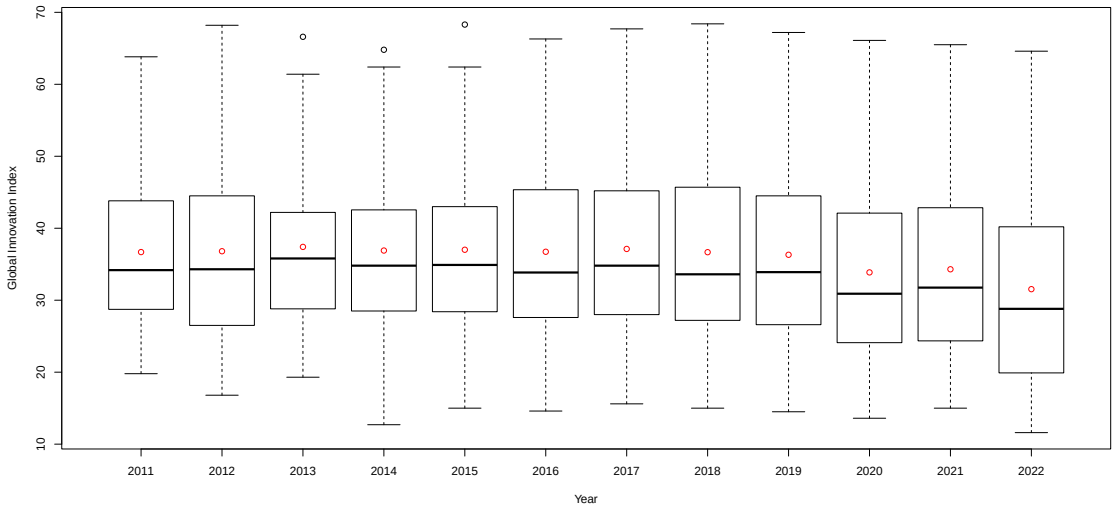
3.2 Mean Values of GII and GII Pillars over the Years



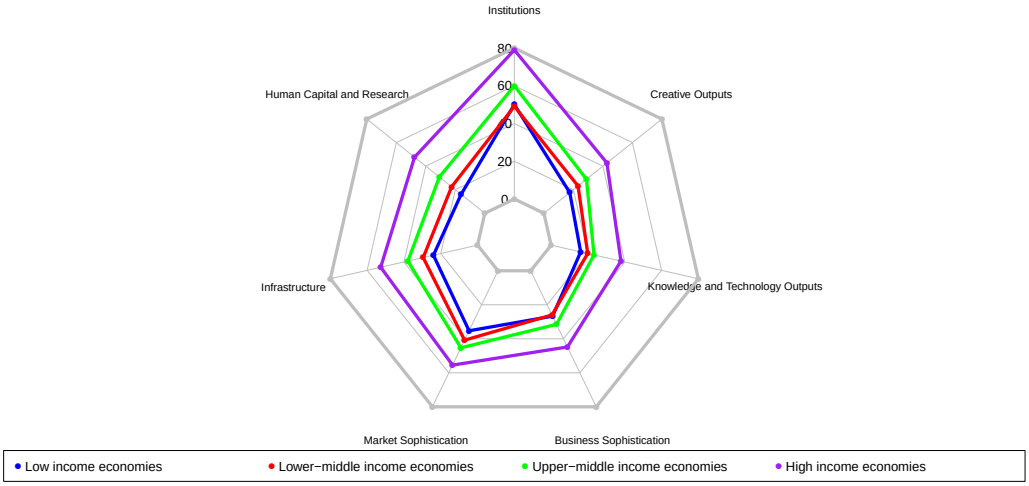
The mean values of all the variables except "Market Sophistication" and "Infrastructure" remain very stable between the years 2011-2021. The above figure also shows a steady decline in some variables in same period, namely "Knowledge and Technology Outputs", "Business Sophistications", "Creative Outputs", "Human Capital and Research" and "Global Innovation Index".
Moreover, there is significant decrease in the "Institutions" and "Market Sophistication" pillars between 2021 and 2022. This significant decrease can be explained by methodological differences in these two pillars in 2021 and 2022 GII editions.

3.3 Boxplots of Distribution of GII

The following figure shows the distribution of GII by years. Between 2011 to 2019, the central tendency indicators (mean and median values) remain stable followed by a slight downward trend. Regarding the dispersion domain, the distributions in the years 2013, 2014 and 2015 show more homogeneous values than the remaining years, where dispersion is more visible (particularly a right skewed distribution corresponding to a longer upper tail). This right skewness is mainly due to a small number of developed countries outperforming the developing and underdeveloped ones.



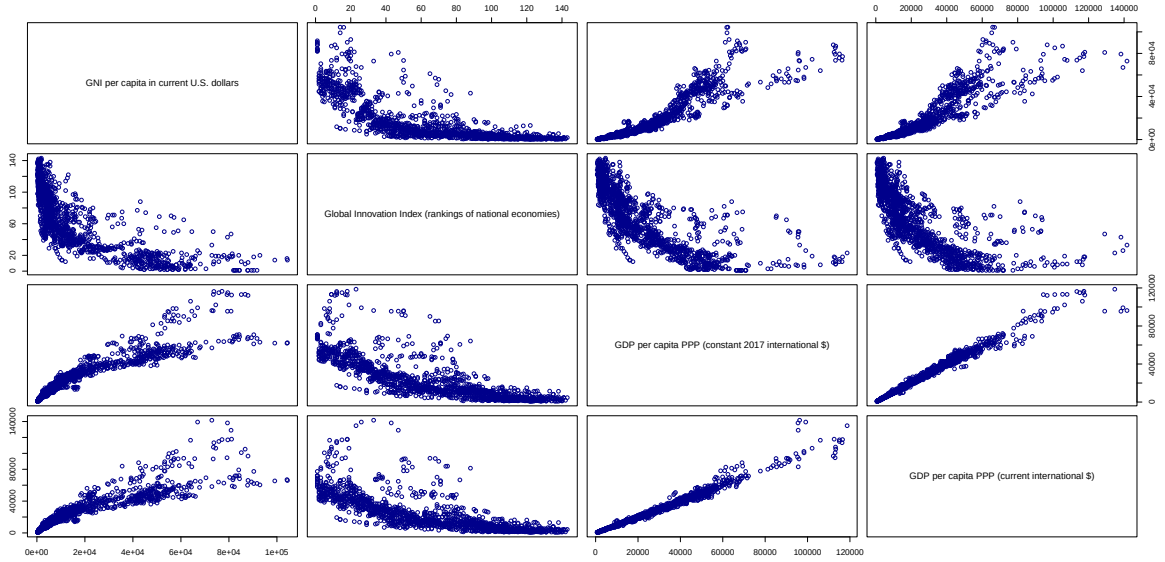
3.4 Mean Values based on Income Level



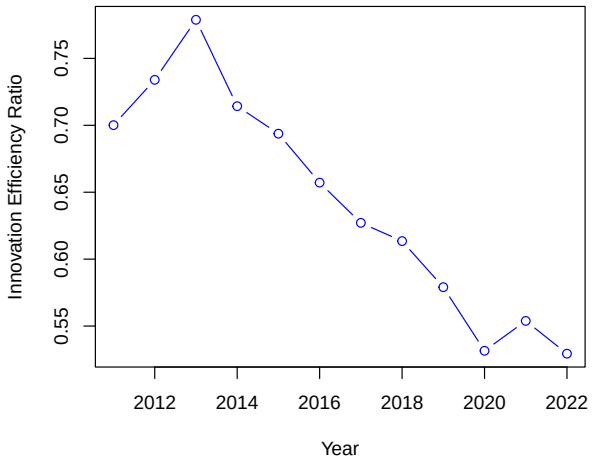
The above radar chart shows that the high income economies clearly outperform the other groups of economies in each GII pillar. It also seems clear that the low-income economies and the lower-middle income economies present similar mean values for all the GII pillars, while there is a significant difference in the performance of upper-middle income economies and high income economies.

3.5 GII Ranks v/s Income Parameters

The following figure describes the distribution of the ranked national economies in GII by income (per capita). The scatterplot clearly shows that the higher income are more likely to achieve top positions of GII than lower-income economies for the years 2011 to 2022.



3.6 Mean of Innovation Efficiency Ratio



The mean value of the IER seems to increase from year 2011 to 2013, then steadily decline till year 2020 and then a increase in 2021 followed by a decrease in 2022. There is an unusual peak in the year 2013. An overall decrease in the mean value of IER can be seen which suggests that the importance of output pillars is decreasing with time in the computation of actual GII.

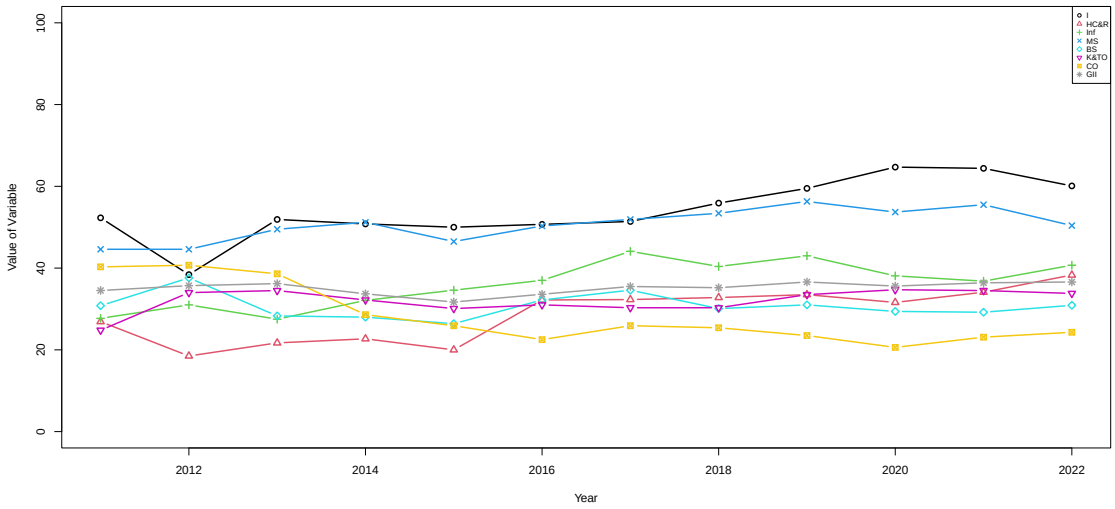
3.7 Top Ten Ranks over the Years

Rank	2011	2012	2013	2014	2015	2016
1	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland
2	Sweden	Sweden	Sweden	United Kingdom	United Kingdom	Sweden
3	Singapore	Singapore	United Kingdom	Sweden	Sweden	United Kingdom
4	Hong Kong (China)	Finland	Netherlands	Finland	Netherlands	USA
5	Finland	United Kingdom	USA	Netherlands	USA	Finland
6	Denmark	Netherlands	Finland	USA	Finland	Singapore
7	USA	Denmark	Hong Kong (China)	Singapore	Singapore	Ireland
8	Canada	Hong Kong (China)	Singapore	Denmark	Ireland	Denmark
9	Netherlands	Ireland	Denmark	Luxembourg	Luxembourg	Netherlands
10	United Kingdom	USA	Ireland	Hong Kong (China)	Denmark	Germany

Rank	2017	2018	2019	2020	2021	2022
1	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland
2	Sweden	Netherlands	Sweden	Sweden	Sweden	USA
3	Netherlands	Sweden	USA	USA	USA	Sweden
4	USA	United Kingdom	Netherlands	United Kingdom	United Kingdom	United Kingdom
5	United Kingdom	Singapore	United Kingdom	Netherlands	Republic of Korea (the)	Netherlands
6	Denmark	USA	Finland	Denmark	Netherlands	Republic of Korea (the)
7	Singapore	Finland	Denmark	Finland	Finland	Singapore
8	Finland	Denmark	Singapore	Singapore	Singapore	Germany
9	Germany	Germany	Germany	Germany	Denmark	Finland
10	Ireland	Ireland	Israel	Republic of Korea (the)	Germany	Denmark

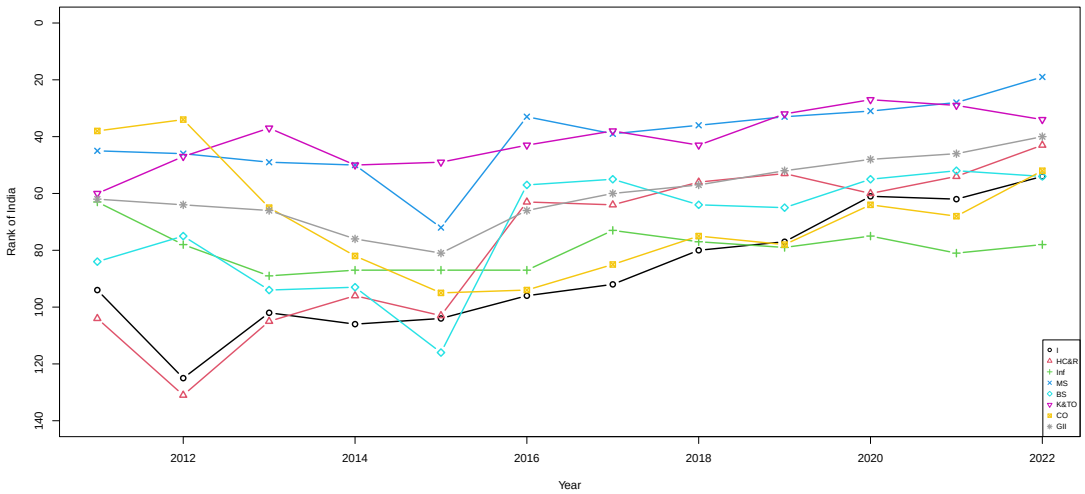
It is visible that Switzerland has been consistently ranked first for all years from 2011 to 2022. Sweden has also maintained its second rank for most of the years. Hong Kong (China) completely disappeared from the top ten ranks after year 2014. The Republic of Korea appeared in 2020 and rose up in following years.

3.8 Innovation Variables for India



It is visible from the above figure that all the variables show an overall increase from 2011 to 2022 except "Creative Outputs" which shows a steady decline. The value of GII does not show any significant change over the years. An unusual fall in the "Institutions" pillar can be seen in the year 2012 and there is a sudden increase in the "Human Capital and Research" pillar from 2015 to 2016.

3.9 Rankings of India in Innovation Variables



It can be seen that the rank of India has improved in every variable except "Creative Outputs" and "Infrastructure". Rank of India in GII has slightly improved over the years with a decrease in first few years followed by a steady increase. The rank of India has significantly improved for "Institutions", "Human Capital and Research" and "Business Sophistication".
There are some unusual falls in "Institutions" and "Human Capital and Research" and "Business Sophistication" for the year 2015 followed by a big leap.

4. Summary of Multiple Linear Regression

In multiple linear regression, we want to model an independent/response variable y as a linear function of some dependent/predictor variables x_1, x_2, ..., x_k plus some random error ε.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

We assume that ε follows $\mathcal{N}(0, \sigma^2)$, and that the values of ε for two different values of y are independent of each other.

4.1 Ordinary Least Squares

We want to fit a model $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \cdots + \widehat{\beta}_k x_k$ to the given sample data $\{(x_{1,i}, x_{2,i}, \ldots, x_{k,i}, y_i)\}_{i=1}^{n}$, where $\widehat{y}$ is an estimator for $\mathrm{E}[y]$ and $\widehat{\beta}_i$ estimates β_i for each i from 1 to n. Suppose $\widehat{y}_i$ is the fitted value for the values $(x_{1,i}, x_{2,i}, \ldots, x_{k,i})$. We use the method of ordinary least squares (OLS) to find $\widehat{\beta}_i$s such that $\mathrm{E}[\widehat{y}] = \mathrm{E}[y]$ and

$$\mathsf{SSE} = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_{1,i} x_{1,i} + \cdots + \widehat{\beta}_{k,i} x_{k,i} \right) \right)^2$$

is minimum. A computational aid is used to find the point estimators $\widehat{\beta}_i$ and the corresponding standard errors $s_{\hat{\beta}_i}$.

4.2 Confidence Intervals

A $100(1-\alpha)\%$ confidence interval for β_i is given by $\widehat{\beta}_i \pm t_{\alpha/2} s_{\hat{\beta}_i}$, where the $t_{\alpha/2}$ value is based on $n - (k+1)$ degrees of freedom.

4.3 Estimating Variance

We can estimate σ^2 using s^2 where

$$s^2 = \frac{\mathsf{SSE}}{n - \text{number of } \beta_i\text{s}} = \frac{\mathsf{SSE}}{n - (k+1)}$$

4.4 Individual t-tests

We can test whether or not an independent variable x_i is significant for predicting y using the following t-test with the test statistic $t_{\mathrm{c}} = \widehat{\beta}_i \big/ s_{\hat{\beta}_i}$. The same assumptions are made about ε.

	Lower-Tailed	**Upper-Tailed**	**Two-Tailed**
Null Hypothesis	$H_0: \ \beta_i = 0$	$H_0: \ \beta_i = 0$	$H_0: \ \beta_i = 0$
Alternate Hypothesis	$H_\text{a}: \ \beta_i < 0$	$H_\text{a}: \ \beta_i > 0$	$H_\text{a}: \ \beta_i \neq 0$
Rejection Region	$t_\text{c} < -\lvert t_\alpha \rvert$	$t_\text{c} > \lvert t_\alpha \rvert$	$\lvert t_\text{c} \rvert > \lvert t_{\alpha/2} \rvert$
p-Value	$\Pr(t < t_\text{c})$	$\Pr(t > t_\text{c})$	$2\Pr(t > \lvert t_\text{c} \rvert)$

If the null is not rejected, it means one of the following three things.

- There is no relationship between y and x_i.
- There is a linear relationship between y and x_i but a Type-II error occured.
- There is a relationship between y and x_i but it is not linear.

Therefore, the most we can say about a β_i parameter is that there is sufficient/insufficient evidence to reject the possibility of a linear relationship between y and x_i.

4.5 Global F-Test

Conducting a series of t-tests on each β_i is not the best way to test the overall utility of the model because as the number of variables increase, the possibility of including an insignifcant variable as significant increases.
Suppose we do a t-test on each of the β_is at a significance level of α, and that all β_is are actually zero. Then, the probability that we accept the null all the times is $(1-\alpha)^k$. Thus, the probability that we reject the null at least once is $1-(1-\alpha)^k \sim k\alpha$. Therefore, a global F-test is performed using the following test statistic.

$$F_\text{c} = \frac{\mathsf{SS}_{yy} - \mathsf{SSE}}{k} \Bigg/ \frac{\mathsf{SSE}}{n-(k+1)}$$

where $\mathsf{SS}_{yy} = \sum (y_i - \overline{y})^2$. The test is as follows.

$$\begin{aligned} \text{Null Hypothesis, } H_0 &: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ \text{Alternate Hypothesis, } H_\text{a} &: \text{at least one } \beta_i = 0 \\ \text{Rejection Region} &: F_\text{c} > F_\alpha \\ p\text{-Value} &: \Pr(F > F_\text{c}) \end{aligned}$$

Rejecting the null in this case means that the model is actually adequate and useful.

4.6 Coefficients of Determination

The **multiple coefficient of determination** is given by

$$R^2 = 1 - \frac{\mathsf{SSE}}{\mathsf{SS}_{yy}} = \frac{\text{Explained Variability}}{\text{Total Variability}}$$

The problem with R^2 is that it increases with the number of β_i parameters, and thus is a good measure if $n \gg k$. To deal with this problem, we scale down SSE and SS_{yy} so that increasing k does not increase the coefficient, and hence define the **adjusted coefficient of determination** as

$$R^2_\text{a} = 1 - \frac{\mathsf{SSE}}{n-(k+1)} \Bigg/ \frac{\mathsf{SS}_{yy}}{n-1} = 1 - \left(\frac{n-1}{n-(k+1)} \right) \left(1 - R^2 \right)$$

4.7 Final Check

To check the utility of a MLR model,

- Use F-test to check the adequacy of the model. If the model is deemed adequate,
- Conduct t-tests on the β_i parameters of particular interest, and limit the number of such t-tests.
- Use R^2 and R^2_{a} to decide how good the fit is.

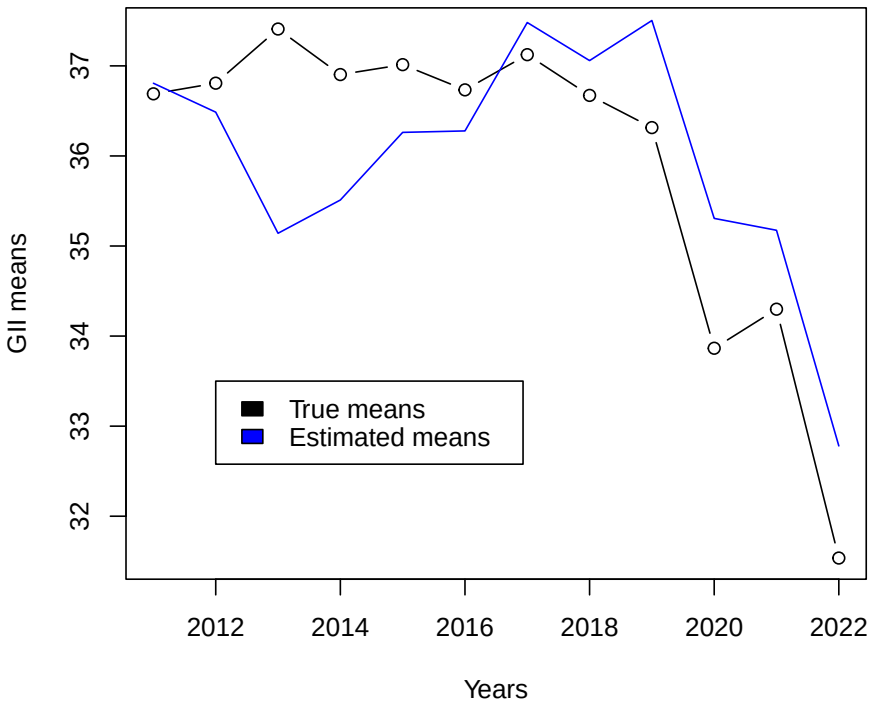
5. Inferential Statistics

The Global Innovation Index is calculated as the average of the input and output subindices. But the question is, are both of these equally significant? To check this, we try to estimate the GII using only the input pillars. Since the GII has an exact linear dependence on the seven pillars, we shall try to fit a linear model by considering only the five input pillars, *i.e.*, we shall conduct an MLR with the GII as the dependent variable and the five input pillars as the independent variables.

$$\mathsf{GII} = \beta_0 + \beta_1 \cdot \mathsf{I} + \beta_2 \cdot \mathsf{HC\&R} + \beta_3 \cdot \mathsf{Inf} + \beta_4 \cdot \mathsf{MS} + \beta_5 \cdot \mathsf{BS}$$

5.1 Overall Model

We find the coefficients β_0 to β_5 using all our rows of data (for all years) as the sample points. We use this model to find the estimated GII values and hence the resultant estimates for the GII means for the years. Plotting these estimates against the years along with the true means, we get the following plot.

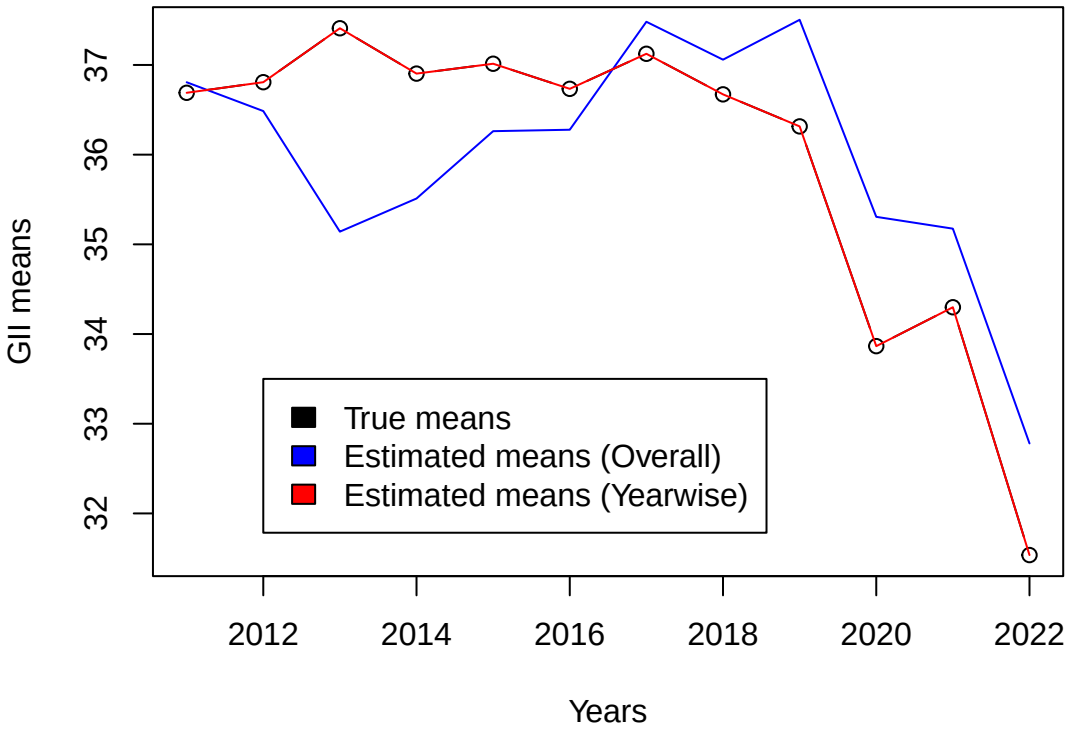


While we see some similarity, this is clearly not a good fit. The following reasons may have caused the discrepancies observed.

- Ignoring the output pillars leads to loss of information.
- The pillars, sub-pillars and the indicators change over the years *i.e.*, there is a great variation between the data for different years. This model, however, does not consider yearwise changes.

5.2 Yearwise Model

In an attempt to account for the variation between the years, we shall now try to fit a separate linear model for each of the years 2011 to 2022. For this model, we find 12 different sets of coefficients β_0 to β_5 for each of the years from 2011 to 2022. We plot these estimates against the years and compare with the Overall Model.



Unexpectedly, the yearwise model estimates agree very closely to the true values.

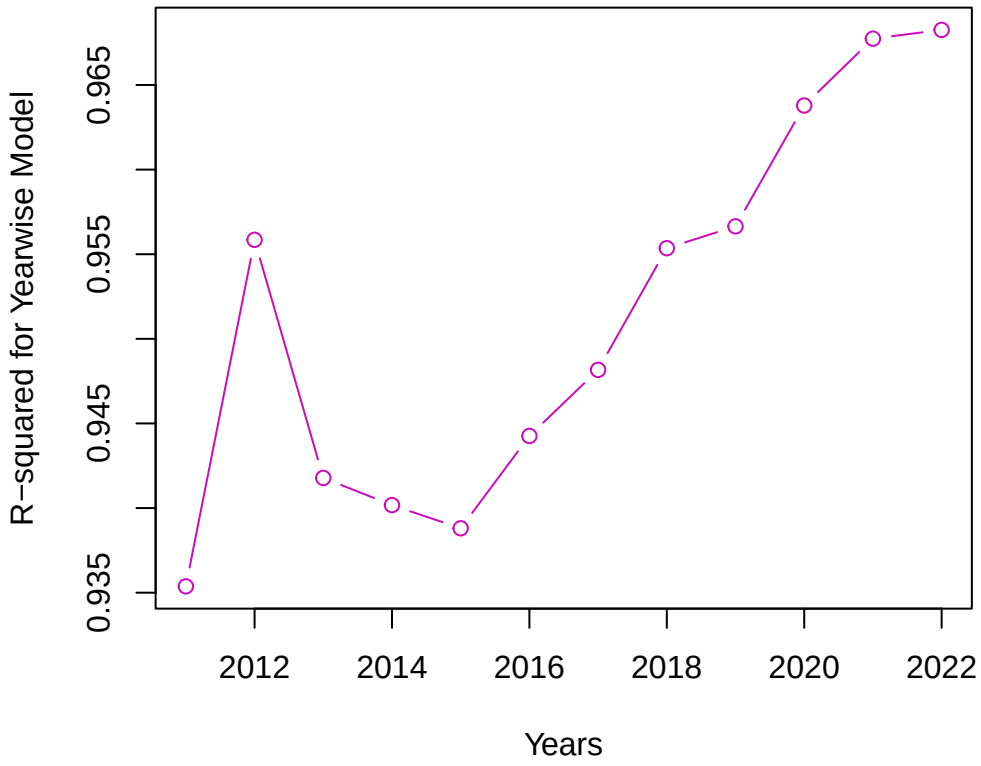
5.3 Coefficients of Determination for Yearwise Model

For a more concrete measure of the quality of this fit, we analyse the coefficients of determination. We tabulate the R^2 and the adjusted R^2 values for each of the years.

Year	R^2	R^2_a
2011	0.9353734	0.9326580
2012	0.9558503	0.9542152
2013	0.9417804	0.9396400
2014	0.9401746	0.9379912
2015	0.9388046	0.9365381
2016	0.9442601	0.9419757

Year	R^2	R^2_a
2017	0.9481662	0.9460243
2018	0.9553562	0.9534960
2019	0.9566461	0.9548837
2020	0.9637872	0.9623387
2021	0.9677380	0.9664577
2022	0.9682551	0.9669953

All of these values are well above 0.9. Thus, for any given year, the input pillars alone (without the output pillars) provide a good estimate for the GII. Also note that the R^2 values are seen to increase steadily from about 0.94 in 2016 to about 0.97 in 2022. This is also evident from the plot of the R^2 values.



This increasing trend of R^2 values suggest that as the years progress, the input pillars get increasingly better at estimating the GII independently of any contribution from the output pillars. This aligns well with the fact that the Innovation Efficiency Ratio, which is the ratio of the output subindex to the input subindex, has also been decreasing with the years.

5.4 Utility of Yearwise Model – *F*-Test

While we have seen how good our model is, for the sake of completion, we shall conduct an F-test to check that not all of the 5 pillars are redundant *i.e.*, to check whether our model is adequate or not. Below are the tabulated values for the F-statistic and the

F_α values for each of the years (at the 1% significance level).

Year	F_c	F_α
2011	344.4692	3.174896
2012	584.5563	3.155743
2013	439.9966	3.154699
2014	430.5995	3.153671
2015	414.2093	3.155743
2016	413.3478	3.170912

Year	F_c	F_α
2017	442.6771	3.172217
2018	513.5887	3.173545
2019	542.8226	3.169628
2020	665.3646	3.167124
2021	755.9037	3.165902
2022	768.6273	3.165902

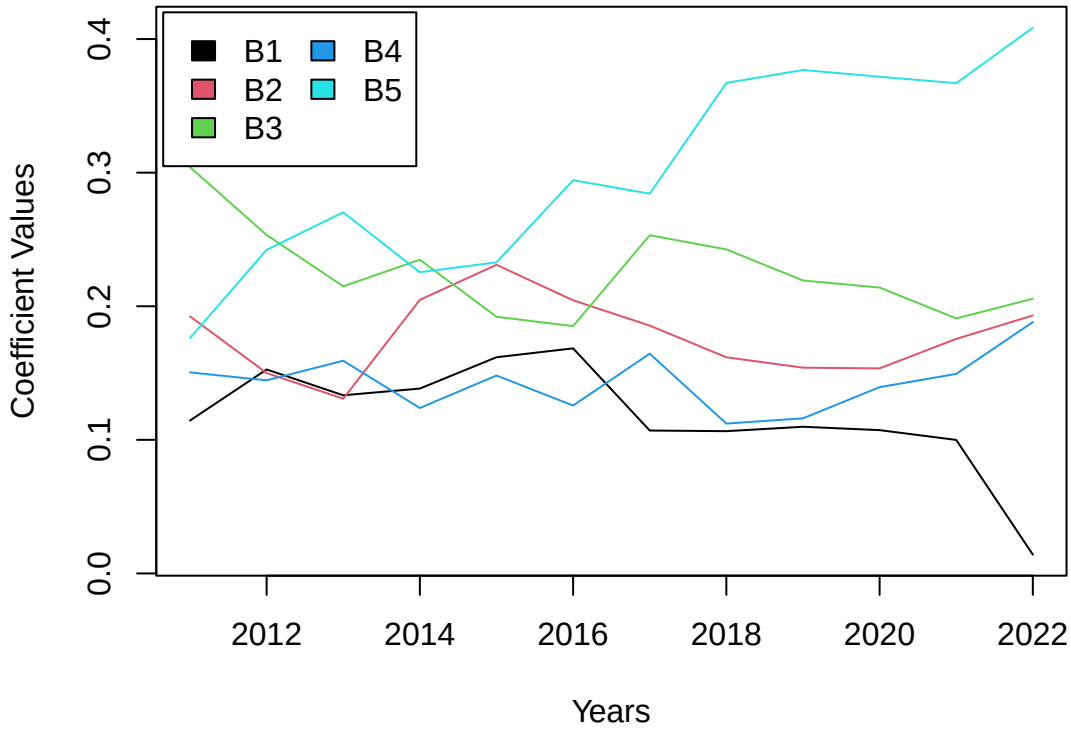
Note that F_α must be calculated separately for each year. This is because, due to the way the data is collected, not all years have the same number of economies and so the degrees of freedom differ for each year.
From the above table, we can observe that $F_c \gg F_\alpha$ so we can safely reject the null hypothesis. In fact, the test works even when we take $\alpha = 10^{-8}$. Thus, we can conclude that our model is adequate and useful, and as seen from the R^2 values, highly accurate.

5.5 Analysing the Model Parameters

The unexpected good fit of the yearwise model merits a closer look at its coefficients β_1 to β_5. (β_0 is just the intercept and carries no significant meaning in this case.) The parameters are tabulated as below.

	β_1 (I)	β_2 (HC&R)	β_3 (Inf)	β_4 (MS)	β_5 (BS)
2011	0.11444595	0.1924279	0.3041312	0.1504952	0.1761878
2012	0.15267741	0.1500172	0.2533566	0.1445505	0.2422084
2013	0.13334341	0.1307954	0.2149075	0.1592058	0.2702794
2014	0.13838054	0.2047962	0.2348461	0.1238236	0.2254494
2015	0.16183297	0.2310143	0.1920959	0.1481461	0.2328565
2016	0.16848573	0.2044318	0.1851399	0.1256593	0.2943483
2017	0.10696293	0.1855131	0.2531242	0.1644996	0.2842668
2018	0.10649015	0.1617829	0.2425651	0.1121800	0.3671657
2019	0.10978181	0.1540149	0.2192733	0.1161081	0.3768481
2020	0.10728120	0.1534778	0.2140167	0.1394516	0.3717855
2021	0.09993400	0.1755447	0.1908905	0.1493512	0.3669757
2022	0.01409131	0.1931056	0.2056140	0.1881354	0.4084380

Following is a plot of the coefficients β_1 to β_5 over the years.



We can make the following observations.

- β_5, which corresponds to the BS pillar, shows a significant increase of about 131% over the years.
- β_1, which corresponds to the I pillar, show a significant decrease of about 87% over the years.
- The other coefficients show no particular trends.

Hence, we can infer that the relative importance of the Business Sophistication pillar has grown steadily and significantly over the years; on the other hand, the Institutions pillar has shown a relative decrease in its contribution to the GII model.

6. Conclusion

We looked at the Overall Model which was not a very good fit due to the yearwise differences not being taken into consideration. However, accounting for the yearwise variations in the Yearwise Model, we found out that the five input pillars (I, HC&R, Inf, MS and BS) without any contribution from the output pillars (K&TO and CO) give a good estimate of the GII. This linear fit of GII based on the five pillars gets increasingly better with years. In particular, the Business Sophistication pillar is gaining more relative importance as years progress, while the relative contribution of the Institutions pillar to the GII is decreasing.

7. Extensions

Here is the list of extensions we have done which were not there in the original paper.

- Plot of the mean values of the Innovation Efficiency Ratio over the years in subsection 3.6.
- Tabulating the top ten ranks in the GII over the years in subsection 3.7.
- Descriptive statistics for India, namely the innovation variables in subsection 3.8 and the rankings in subsection 3.9.
- Multiple linear regression for the GII as a linear function of the five pillars of the input sub-index in Section 5 to infer the growing importance of the input pillars, and in particular the Business Sophistication pillar.

List of Abbreviations

- **BS**: Business Sophistication, fifth pillar of the GII
- **CO**: Creative Outputs, seventh pillar of the GII
- **FDI**: Foreign Direct Investment, the investment made by a firm or an individual for business interests in a different country than their own.
- **GDP**: Gross Domestic Product, a measure of the market value of final goods produced in a country over a period of time, avoiding double counting the intermediate goods.
- **GERD**: Gross Domestic Expenditure on R&D, expenditure made by a country in the Research and Development sector over a period of time
- **GNI**: Gross National Income, a measure of overall domestic and international income of the residents of a country. It may be computed by adjusting the GDP for foreign residents and nonresidents, or by Atlas method etc.
- **GII**: Global Innovation Index
- **HC&R**: Human Capital and Research, second pillar of the GII
- **I**: Institutions, first pillar of the GII
- **ICT**: Information and Communication Technology, an umbrella term for any communication device like television, cell phones, etc.
- **IER**: Innovation Efficiency Ratio
- **Inf**: Infrastructure, third pillar of the GII
- **INSEAD**: Institut Européen d'Administration des Affaires
- **ISO**: International Organisation for Standardisation
- **K&TO**: Knowledge and Technology Outputs, sixth pillar of the GII
- **MS**: Market Sophistication, fourth pillar of the GII
- **OECD**: Organisation for Economic Cooperation and Development
- **OLS**: Ordinary Least Squares, a method for choosing the unknown parameters in a linear regression model
- **PISA**: Programme for International Student Assessment, a study to evaluate educational systems by measuring 15-year-old school students' performance on mathematics, science and reading
- **PPP**: Purchasing Power Parity
- **QS**: Quacquarelli Symonds, the company which publishes QS world university rankings
- **R&D**: Research and Development, the set of activities that companies undertake to innovate and introduce new products and services, or improve existing ones

- **SDR**: Special Drawing Rights, an supplement reserve asset to the official assets of countries assigned by the International Monetary Fund (IMF)
- **SSE**: Sum of Squares of Errors
- **STEM**: Science, Technology, Engineering and Mathematics
- **TLD**: Top Level Domain, the highest domain name in the hierarchy of the Domain Name System, for example, `.com`, `.edu`, `.in`, `.ru`, etc.
- **WIPO**: World Intellectual Property Organization

References

[1] The World Bank. The world bank atlas method - detailed methodology. https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-the-world-bank-atlas-method-detailed-methodology, 2023.

[2] Gonçalo Rodrigues Brás. Global innovation index: panel data (2011-2022). https://data.mendeley.com/datasets/cvkdzr8tv3/4, 2022.

[3] Gonçalo Rodrigues Brás. Pillars of the global innovation index by income level of economies: longitudinal data (2011-2022) for researchers' use. *Data in Brief*, 46:108818, 2023. ISSN 2352-3409. doi: 10.1016/j.dib.2022.108818. URL https://www.sciencedirect.com/science/article/pii/S2352340922010216.

[4] Schuyler W. Huck. *Reading Statistics and Research*. Pearson Education, Boston, MA, 2012. URL http://refhub.elsevier.com/S2352-3409(22)01021-6/sbref0001.

[5] J.T. McClave and T.T. Sincich. *Statistics*. Pearson Education, 2016. ISBN 9780134080598. URL https://books.google.co.in/books?id=AFvDCwAAQBAJ.

[6] Wikipedia contributors. Purchasing power parity — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Purchasing_power_parity&oldid=1147092000, 2023. [Online; accessed 28-March-2023].

[7] WIPO. Global innovation index 2011. https://www.wipo.int/edocs/pubdocs/en/economics/gii/gii_2011.pdf, 2011.

[8] WIPO. Global innovation index 2012. https://www.wipo.int/edocs/pubdocs/en/economics/gii/gii_2012.pdf, 2012.

[9] WIPO. Global innovation index 2018. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2018.pdf, 2018.

[10] WIPO. Global innovation index 2019. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2019.pdf, 2019.

[11] WIPO. Appendix i: The global innovation index conceptual framework. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020-appendix1.pdf, 2020.

[12] WIPO. Global innovation index 2022. https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2022-en-main-report-global-innovation-index-2022-15th-edition.pdf, 2022.

[13] WIPO. Global innovation index database. https://www.globalinnovationindex.org/analysis-indicator, 2022.

[14] WIPO. About the gii. https://www.globalinnovationindex.org/about-gii, 2023.