Constructing the Empirical Model to Understand the Variance in Gross Domestic Product Per Capita Explained by Competitiveness Factors

An Econometrics Project for ‘R Coding and Finance Analytics’ Course

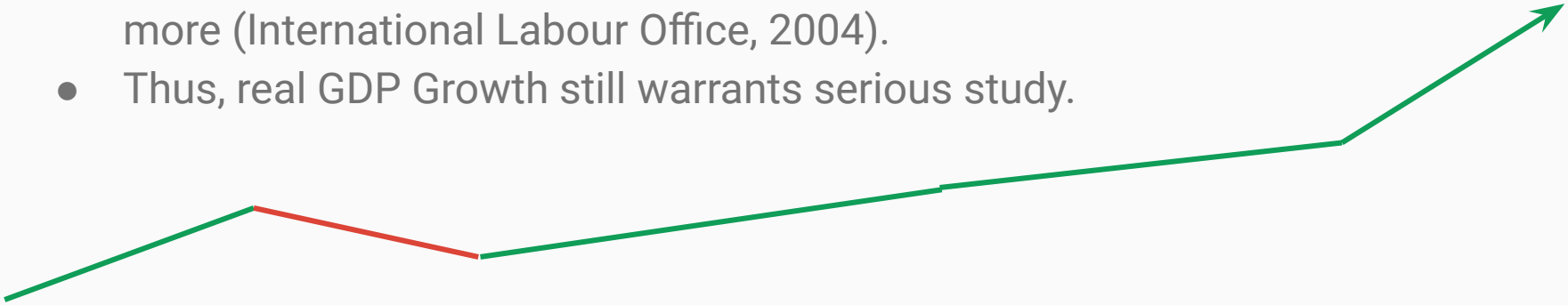
Sarbjit Gill
12 April 2022

Table of Contents

Project Introduction

Why Does the GDP Growth Matter?

- Gross Domestic Product is “the monetary value of final goods and services produced in a country in a given period of time” (Callen, 2020).
- An economic expansion may be unequal, but an economic contraction has big implications, such as widespread job loss, lower quality of life, and more (International Labour Office, 2004).
- Thus, real GDP Growth still warrants serious study.



GDP Growth Components

1. Growth in labour force
2. Growth in physical and natural capital inputs
3. Total factor productivity growth (TFP)
    - The least understood
    - The most influential
        i. represents more than ⅗ of the growth

(Schwab, 2018)

Table 3: Growth Accounting for the United States

		Contributions from		
Period	Output per hour	K/Y	Labor Composition	Labor-Aug. TFP
1948–2013	**2.5**	**0.1**	**0.3**	**2.0**
1948–1973	3.3	-0.2	0.3	3.2
1973–1990	1.6	0.5	0.3	0.8
1990–1995	1.6	0.2	0.7	0.7
1995–2000	3.0	0.3	0.3	2.3
2000–2007	2.7	0.2	0.3	2.2
2007–2013	1.7	0.1	0.5	1.1

Note: Average annual growth rates (in percent) for output per hour and its components for the private business sector, following equation (3). Source: Authors calculations using Bureau of Labor Statistics, *Multifactor Productivity Trends*, August 21, 2014.

Total Factor Productivity (TFP) Growth

- TFP Growth is the **“unexplained part”** of GDP growth.
- The World Economic Forum’s Global Competitiveness Index (GCI), published annually, aims to understand the TFP growth factors of nations to uncover underlying differences propelling unexplained growth.
- The GCI computes an overall competitiveness score for observed economics and a score for 12 main drivers of productivity: **institutions; infrastructure; information and communications technology (ICT) adoption; macroeconomic stability; health; skills; product market; labor market; financial system; market size; business dynamism; and innovation capability.**

Project Purpose

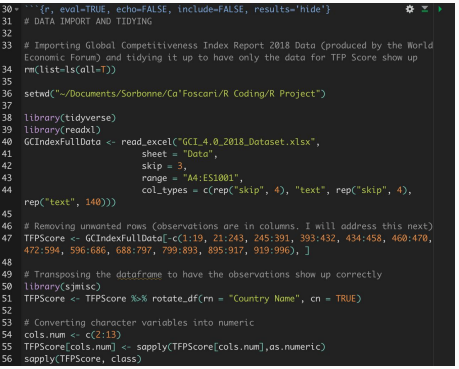
- **THE INTENT:** Conduct a thorough multiple linear regression analysis to understand which total factor productivity (TFP) factors best explain the variance in GDP per capita.
- **THE REASON:** Policy implications: Understanding which competitiveness factors are the most valuable in driving GDP growth is crucial for economies seeking to expand. The information is particularly valuable for government's long-term investment decision-making.

Coding the Project in R & Results

Data Used

Regression Model Component	Source
• Regressors	• World Economic Forum: 2018 GCI Index ○ 12 Regressors (TFP Productivity Indicators) ○ 140 observations
• Dependent Variable	• World Bank Indicators ○ 2017 GDP per capita, constant, PPP value ○ 140+ observations

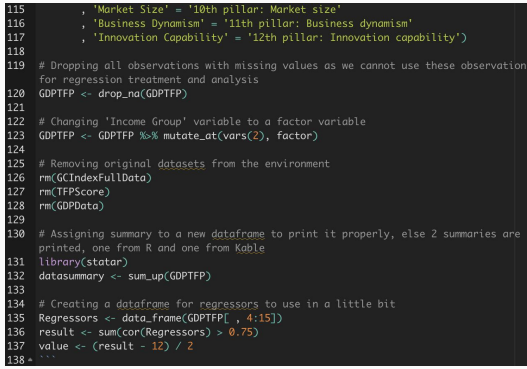
Importing and Tidying Data



```{r, eval=TRUE, echo=FALSE, include=FALSE, results='hide'}
# DATA IMPORT AND TIDYING

# Importing Global Competitiveness Index Report 2018 Data (produced by the World Economic Forum) and tidying it up to have only the data for TFP Score show up
rm(list=ls(all=T))

setwd("~/Documents/Sorbonne/Ca'Foscari/R Coding/R Project")

library(tidyverse)
library(readxl)
GCIndexFullData <- read_excel("GCI_4.0_2018_Dataset.xlsx",
                      sheet = "Data",
                      skip = 3,
                      range = "A4:ES10001",
                      col_types = c(rep("skip", 4), "text", rep("skip", 4), rep("text", 140)))

# Removing unwanted rows (observations are in columns. I will address this next)
TFPScore <- GCIndexFullData[-c(1:19, 21:243, 245:391, 393:432, 434:458, 460:470, 472:594, 596:686, 688:797, 799:893, 895:917, 919:996), ]

# Transposing the dataframe to have the observations show up correctly
library(sjmisc)
TFPScore <- TFPScore %>% rotate_df(rn = "Country Name", cn = TRUE)

# Converting character variables into numeric
cols.num <- c(2:13)
TFPScore[cols.num] <- sapply(TFPScore[cols.num],as.numeric)
sapply(TFPScore, class)
```

```
TFPScore <- as_tibble(TFPScore)
rm(cols.num)

# Importing the second dataset. It comes from the World Bank Website and it contains the 2017 GDP Values (PPP, Constant).
library(readxl)
as_tibble(GDPData <- read_excel("API_NY.GDP.PCAP.PP.KD_DS2_en_excel_v2_3732022.xls",
                                sheet = "Data",
                                skip = 3,
                                col_types = c("text", "skip", "text", rep("skip", 59), "numeric", rep("skip", 3))))

# Renaming country names to match between GDPData and TFPScore dataset. Since TFPScore dataframe is shorter, we'll rename using its values and then drop the observations later with missing values after combibing the datasets.
GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Congo, Dem. Rep.", "Congo, Democratic Rep.")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Cabo Verde", "Cape Verde")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Cote d'Ivoire", "Côte d'Ivoire")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Egypt, Arab Rep.", "Egypt")
```

```
GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "North Macedonia", "Macedonia, FYR")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Venezuela, RB", "Venezuela")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Vietnam", "Viet Nam")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Yemen, Rep.", "Yemen")

# Joining the datasets so as to work within a single dataframe
GDPTFP <- left_join(GDPData, TFPScore, by = "Country Name")

# Renaming Columns
GDPTFP <- rename(GDPTFP, '2017 GDP Per Capita' = '2017'
       , 'Income Group' = 'IncomeGroup'
       , 'Institutions' = '1st pillar: Institutions'
       , 'Infrastructure' = '2nd pillar: Infrastructure'
       , 'ICT Adoption' = '3rd pillar: ICT adoption'
       , 'Macroeconomic Stability' = '4th pillar: Macroeconomic stability'
       , 'Health' = '5th pillar: Health'
       , 'Skills' = '6th pillar: Skills'
       , 'Product Market' = '7th pillar: Product market'
       , 'Labor Market' = '8th pillar: Labor market'
       , 'Financial System' = '9th pillar: Financial system'
```



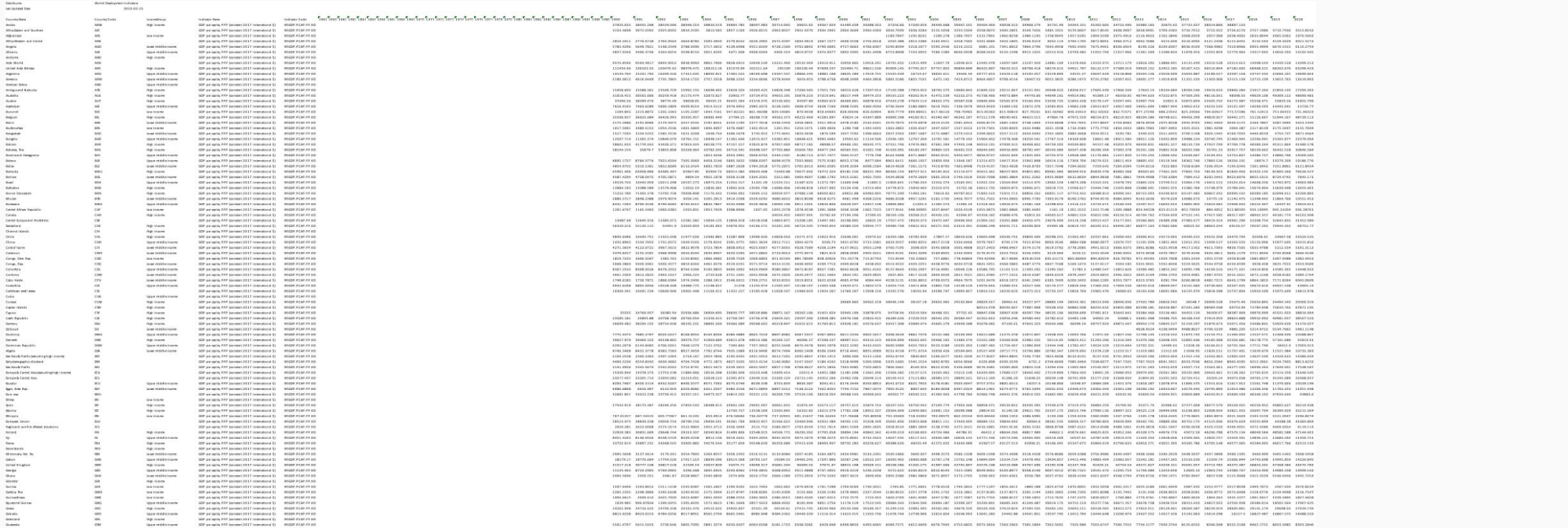
```
       , 'Market Size' = '10th pillar: Market size'
       , 'Business Dynamism' = '11th pillar: Business dynamism'
       , 'Innovation Capability' = '12th pillar: Innovation capability')

# Dropping all observations with missing values as we cannot use these observations for regression treatment and analysis
GDPTFP <- drop_na(GDPTFP)

# Changing 'Income Group' variable to a factor variable
GDPTFP <- GDPTFP %>% mutate_at(vars(2), factor)

# Removing original datasets from the environment
rm(GCIndexFullData)
rm(TFPScore)
rm(GDPData)

# Assigning summary to a new dataframe to print it properly, else 2 summaries are printed, one from R and one from Kable
library(statar)
datasummary <- sum_up(GDPTFP)

# Creating a dataframe for regressors to use in a little bit
Regressors <- data_frame(GDPTFP[ , 4:15])
result <- sum(cor(Regressors) > 0.75)
value <- (result - 12) / 2
```

- The most challenging part

Importing and Tidying Data: World Bank Dataset

VARIABLES

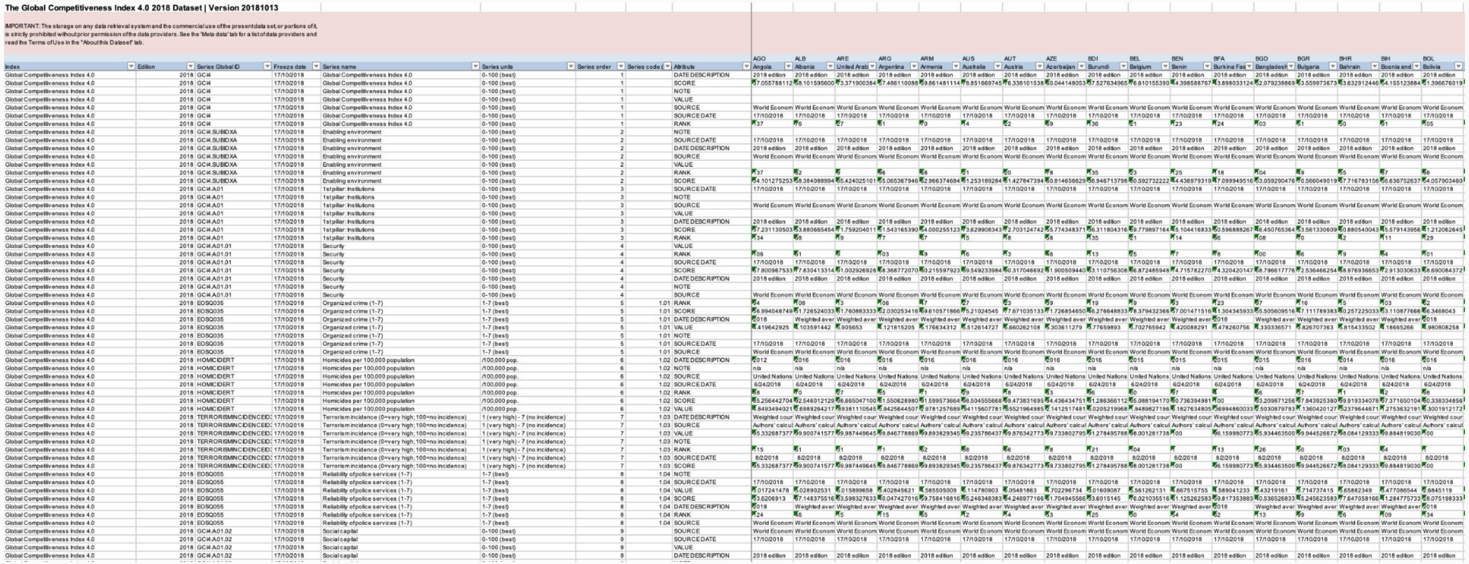
OBSERVATIONS



Importing and Tidying Data: WEF’s Global Competitiveness Index Data

OBSERVATIONS →

VARIABLES ↓



Process: Importing and Tidying Data

```r
```{r, eval=TRUE, echo=FALSE, include=FALSE, results='hide'}
# DATA IMPORT AND TIDYING

# Importing Global Competitiveness Index Report 2018 Data (produced by the World Economic Forum) and tidying it up to have only the data for TFP Score show up
rm(list=ls(all=T))

setwd("~/Documents/Sorbonne/Ca'Foscari/R Coding/R Project")

library(tidyverse)
library(readxl)
GCIndexFullData <- read_excel("GCI_4.0_2018_Dataset.xlsx",
                      sheet = "Data",
                      skip = 3,
                      range = "A4:ES1001",
                      col_types = c(rep("skip", 4), "text", rep("skip", 4), rep("text", 140)))

# Removing unwanted rows (observations are in columns. I will address this next)
TFPScore <- GCIndexFullData[-c(1:19, 21:243, 245:391, 393:432, 434:458, 460:470, 472:594, 596:686, 688:797, 799:893, 895:917, 919:996), ]

# Transposing the dataframe to have the observations show up correctly
library(sjmisc)
TFPScore <- TFPScore %>% rotate_df(rn = "Country Name", cn = TRUE)

# Converting character variables into numeric
cols.num <- c(2:13)
TFPScore[cols.num] <- sapply(TFPScore[cols.num],as.numeric)
sapply(TFPScore, class)
```

← Imported WEF data by select columns

← Deleted unwanted rows, as variables are in rows instead of columns in this dataset

← Transposed dataset to show variables in colum

← Transformed these variables to “numeric”
```

Process: Importing and Tidying Data

```
57  TFPScore <- as_tibble(TFPScore)
58  rm(cols.num)
59
60  # Importing the second dataset. It comes from the World Bank Website and it contains
    the 2017 GDP Values (PPP, Constant).
61  library(readxl)
62  as_tibble(GDPData <- read_excel("API_NY.GDP.PCAP.PP.KD_DS2_en_excel_v2_3732022.xls",
63                                  sheet = "Data",
64                                  skip = 3,
65                                  col_types = c("text", "skip", "text", rep("skip",
    59), "numeric", rep("skip", 3))))
66
67  # Renaming country names to match between GDPData and TFPScore dataset. Since
    TFPScore dataframe is shorter, we'll rename using its values and then drop the
    observations later with missing values after comibing the datasets.
68  GDPData <-
69    GDPData %>%
70    mutate_at("Country Name", str_replace, "Congo, Dem. Rep.", "Congo, Democratic
    Rep.")
71
72  GDPData <-
73    GDPData %>%
74    mutate_at("Country Name", str_replace, "Cabo Verde", "Cape Verde")
75
76  GDPData <-
77    GDPData %>%
78    mutate_at("Country Name", str_replace, "Cote d'Ivoire", "Côte d'Ivoire")
79
80  GDPData <-
81    GDPData %>%
82    mutate_at("Country Name", str_replace, "Egypt, Arab Rep.", "Egypt")
```

← Imported World Bank Dataset

← Changed observation names in this dataset so they match the WEF one

Process: Importing and Tidying Data

```r
GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "North Macedonia", "Macedonia, FYR")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Venezuela, RB", "Venezuela")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Vietnam", "Viet Nam")

GDPData <-
  GDPData %>%
  mutate_at("Country Name", str_replace, "Yemen, Rep.", "Yemen")

# Joining the datasets so as to work within a single dataframe
GDPTFP <- left_join(GDPData, TFPScore, by = "Country Name")

# Renaming Columns
GDPTFP <- rename(GDPTFP, '2017 GDP Per Capita' = '2017'
       , 'Income Group' = 'IncomeGroup'
       , 'Institutions' = '1st pillar: Institutions'
       , 'Infrastructure' = '2nd pillar: Infrastructure'
       , 'ICT Adoption' = '3rd pillar: ICT adoption'
       , 'Macroeconomic Stability' = '4th pillar: Macroeconomic stability'
       , 'Health' = '5th pillar: Health'
       , 'Skills' = '6th pillar: Skills'
       , 'Product Market' = '7th pillar: Product market'
       , 'Labor Market' = '8th pillar: Labor market'
       , 'Financial System' = '9th pillar: Financial system'
```

← Combined the two dataframes into one

← Renamed columns of the new dataset

Process: Importing and Tidying Data

```
115         , 'Market Size' = '10th pillar: Market size'
116         , 'Business Dynamism' = '11th pillar: Business dynamism'
117         , 'Innovation Capability' = '12th pillar: Innovation capability')
118
119 # Dropping all observations with missing values as we cannot use these observations
    for regression treatment and analysis
120 GDPTFP <- drop_na(GDPTFP)
121
122 # Changing 'Income Group' variable to a factor variable
123 GDPTFP <- GDPTFP %>% mutate_at(vars(2), factor)
124
125 # Removing original datasets from the environment
126 rm(GCIndexFullData)
127 rm(TFPScore)
128 rm(GDPData)
129
130 # Assigning summary to a new dataframe to print it properly, else 2 summaries are
    printed, one from R and one from Kable
131 library(statar)
132 datasummary <- sum_up(GDPTFP)
133
134 # Creating a dataframe for regressors to use in a little bit
135 Regressors <- data_frame(GDPTFP[ , 4:15])
136 result <- sum(cor(Regressors) > 0.75)
137 value <- (result - 12) / 2
138 ```
```

← Dropped observations with missing values

← Mutated a character variable into factor

← Removed dataframes not needed

Data Description

- 136 observations
  - 85 developed economies and 51 developing
- 14 variables
  - 2017 GDP per Capita Data (PPP, Constant) [numeric]
  - Income Group Classification [factor]
  - 12 Regressors: TFP Indicator Scores (0-100) [numeric]
- For analysis, significance level of 0.05 is used in this project

Data Summary

Upon running the summary, we uncover variables with a high standard deviation and large gaps between the minimum and maximum values, indicating the potential of outliers and influential variables.

Figure 1: Data Summary

Variable	Obs	Missing	Mean	StdDev	Min	Max
2017 GDP Per Capita	136	0	23136.64823	21910.903398	773.57286	114985.84224
Business Dynamism	136	0	59.44860	11.136866	14.89740	86.48944
Financial System	136	0	61.24389	13.301915	38.71931	92.11703
Health	136	0	75.08830	19.581599	11.93262	100.00000
ICT Adoption	136	0	51.84894	19.442234	12.77239	91.25512
Infrastructure	136	0	65.25618	15.733234	28.57425	95.70355
Innovation Capability	136	0	42.15310	17.308829	16.78176	87.52204
Institutions	136	0	55.33580	11.197910	32.92204	81.55430
Labor Market	136	0	59.34682	8.925753	41.99579	81.88546
Macroeconomic Stability	136	0	79.86616	15.991186	31.06250	100.00000
Market Size	136	0	53.83298	17.735528	15.99203	100.00000
Product Market	136	0	56.41957	7.936078	37.51731	81.22964
Skills	136	0	60.57254	14.793489	28.24768	87.87808

```
library(statar)
datasummary <- sum_up(GDPTFP)
```

Correlation Among Regressors

Test reference: >0.7

We uncover that a number of variables are closely correlated based on the test reference (33 unique values). Thus, near multicollinearity may pose an issue in regression analysis. However, since there are no data entry errors, variables are kept as is and they will be investigated further.

Figure 2: Correlation between explanatory variables

	Institutions	Infrastructure	ICT Adoption	Macroeconomic Stability	Health	Skills	Product Market	Labor Market	Financial System	Market Size	Business Dynamism	Innovation Capability
Institutions	1.00	0.82	0.80	0.74	0.66	0.84	0.86	0.84	0.83	0.40	0.85	0.86
Infrastructure	0.82	1.00	0.87	0.72	0.84	0.90	0.78	0.67	0.79	0.62	0.79	0.83
ICT Adoption	0.80	0.87	1.00	0.70	0.79	0.90	0.75	0.71	0.71	0.44	0.74	0.76
Macroeconomic Stability	0.74	0.72	0.70	1.00	0.61	0.70	0.70	0.64	0.75	0.45	0.69	0.73
Health	0.66	0.84	0.79	0.61	1.00	0.81	0.65	0.54	0.67	0.46	0.63	0.68
Skills	0.84	0.90	0.90	0.70	0.81	1.00	0.79	0.78	0.75	0.45	0.80	0.81
Product Market	0.86	0.78	0.75	0.70	0.65	0.79	1.00	0.82	0.76	0.36	0.82	0.76
Labor Market	0.84	0.67	0.71	0.64	0.54	0.78	0.82	1.00	0.72	0.25	0.79	0.74
Financial System	0.83	0.79	0.71	0.75	0.67	0.75	0.76	0.72	1.00	0.56	0.78	0.85
Market Size	0.40	0.62	0.44	0.45	0.46	0.45	0.36	0.25	0.56	1.00	0.50	0.62
Business Dynamism	0.85	0.79	0.74	0.69	0.63	0.80	0.82	0.79	0.78	0.50	1.00	0.83
Innovation Capability	0.86	0.83	0.76	0.73	0.68	0.81	0.76	0.74	0.85	0.62	0.83	1.00

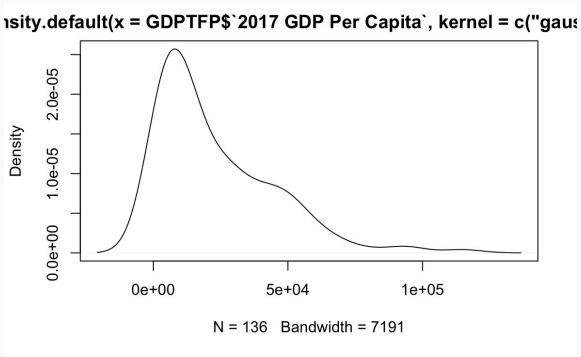
```
Regressors <- data_frame(GDPTFP[ , 4:15])
result <- sum(cor(Regressors) > 0.75)
value <- (result - 12) / 2
```

(`r {value}` unique values).

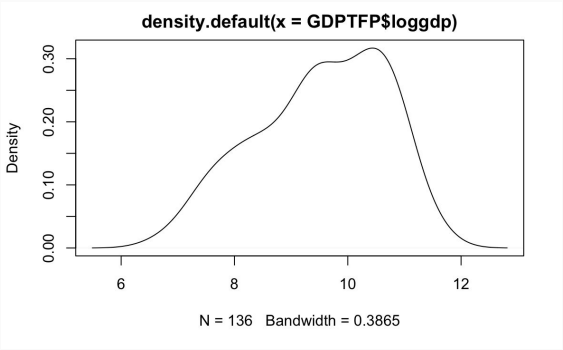
```
library(kableExtra)
knitr::kable(cor(Regressors), digits=2)
```

Test for Normality of Data

- Normality: a key assumption for conducting regression analysis
- Kernel densities were checked to see if variables follow a normal distribution
- The dependent variable and regressors were transformed into log form and added to the dataframe for comparison



GDP per Capita



log(GDP per Capita)

```
##### 1. Dependent Variable: GDP
plot(density(GDPTFP$`2017 GDP Per Capita`, kernel = c("gaussian")))
GDPTFP$loggdp <- log(GDPTFP$`2017 GDP Per Capita`)
plot(density(GDPTFP$loggdp), kernel = c("gaussian"))
```

Test for Normality of Data and Best Regressor

```
##### 2. Regressor Variable: Institutions (Cross-Checked with GDP)
plot(density(GDPTFP$`Institutions`, kernel = c("gaussian")))
GDPTFP$loginstitutions <- log(GDPTFP$`Institutions`)
plot(density(GDPTFP$loginstitutions), kernel = c("gaussian"))
#NOTE: loginstitutions appears more normal. Checking scatterplots ahead.

library(stats)
plot(GDPTFP$`Institutions`, GDPTFP$'2017 GDP Per Capita')
abline(lm(GDPTFP$'2017 GDP Per Capita' ~ GDPTFP$`Institutions`))
#NOTE: Not the best fit

plot(GDPTFP$`loginstitutions`, GDPTFP$'2017 GDP Per Capita')
abline(lm(GDPTFP$'2017 GDP Per Capita' ~ GDPTFP$`loginstitutions`))
#NOTE: Not the best fit. 2017 GDP Per Capita to be discarded

plot(GDPTFP$`Institutions`, GDPTFP$loggdp)
abline(lm(GDPTFP$loggdp ~ GDPTFP$`Institutions`))
#NOTE: loggdp is a better fit with data appearing relatively normal on both sides of
the line

plot(GDPTFP$`loginstitutions`, GDPTFP$loggdp)
abline(lm(GDPTFP$loggdp ~ GDPTFP$`loginstitutions`))
#NOTE: loggdp is a better fit with data appearing relatively normal on both sides of
the line

#NOTE: Now, since both loggdp scatterplots appear to be similar, we'll run a quick
regression to see which variable explains "loggdp" better

summary(lm(GDPTFP$loggdp ~ GDPTFP$`Institutions`))
summary(lm(GDPTFP$loggdp ~ GDPTFP$`loginstitutions`))
#NOTE: "loginstitutions" has a higher R-squared value. Thus, keep "loginstitutions"
and drop "Institutions"
```

← Density plots to compare log and original variables

← Regression Plot (original dependent & regressor variables)

← Regression Plot (original dependent & log regressor)

← Regression Plot (log dependent & original regressor)

← Regression Plot (log dependent & log regressor)

← R-Squared Comparison (log dependent vs log regressor)
(log dependent vs original regressor)

Custom Function to Compare Original and Log-Transformed Regressors

```r
ExplanatoryPower_OrigVsLog <- function(a, b) {
  result <- summary(lm(GDPTFP$loggdp ~ a))$r.squared > summary(lm(GDPTFP$loggdp ~
b))$r.squared
   print(result)
}

ExplanatoryPower_OrigVsLog(GDPTFP$'ICT Adoption', GDPTFP$'logictadoption')
ExplanatoryPower_OrigVsLog(GDPTFP$'Macroeconomic Stability',
GDPTFP$'logmicroeconomicstability')
ExplanatoryPower_OrigVsLog(GDPTFP$'Health', GDPTFP$'loghealth')
ExplanatoryPower_OrigVsLog(GDPTFP$'Skills', GDPTFP$'logskills')
ExplanatoryPower_OrigVsLog(GDPTFP$'Product Market', GDPTFP$'logproductmarket')
ExplanatoryPower_OrigVsLog(GDPTFP$'Labor Market', GDPTFP$'loglabormarket')
ExplanatoryPower_OrigVsLog(GDPTFP$'Financial System', GDPTFP$'logfinancialsystem')
ExplanatoryPower_OrigVsLog(GDPTFP$'Market Size', GDPTFP$'logmarketsize')
ExplanatoryPower_OrigVsLog(GDPTFP$'Business Dynamism', GDPTFP$'logbusinessdynamism')
ExplanatoryPower_OrigVsLog(GDPTFP$'Innovation Capability',
GDPTFP$'loginnovationcapability')

######NOTE: Final variables to keep for the empirical model: loggdp,
loginstitutions, Infrastructure, ICT Adoption, Macroeconomic Stability, Health,
Skills, Product Market, Labor Market, logfinancialsystem, logmarketsize, Business
Dynamism, loginnovationcapability.

FinalRegressors <- GDPTFP[ , c(5:11, 14, 17, 25:26, 28)]
```

← Custom function to compare R-squared results

← Tests (Original, Log Transformation)

← Final Regressors

Custom Function to Compare Significance of Final Regressors in Explaining Y Variable

```r
SignificanceTest <- function(a) {
   result <- summary(lm(GDPTFP$loggdp ~ a))$coefficients[,4] <= 0.05
   print(result)
  }

SignificanceTest(GDPTFP$'loginstitutions')
SignificanceTest(GDPTFP$'Infrastructure')
SignificanceTest(GDPTFP$'ICT Adoption')
SignificanceTest(GDPTFP$'Macroeconomic Stability')
SignificanceTest(GDPTFP$'Health')
SignificanceTest(GDPTFP$'Skills')
SignificanceTest(GDPTFP$'Product Market')
SignificanceTest(GDPTFP$'Labor Market')
SignificanceTest(GDPTFP$'logfinancialsystem')
SignificanceTest(GDPTFP$'logmarketsize')
SignificanceTest(GDPTFP$'Business Dynamism')
SignificanceTest(GDPTFP$'loginnovationcapability')
# All variables are significant in explaining the variation in loggdp, our dependent
variables.
```

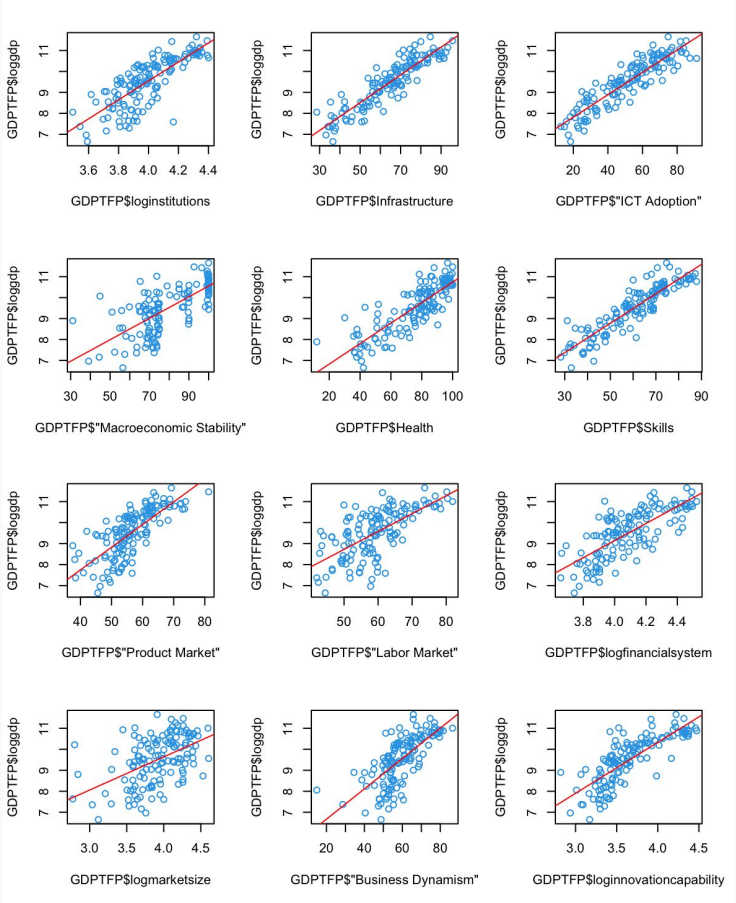
RESULTS

```
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
       TRUE        TRUE
(Intercept)           a
      FALSE        TRUE
```

Scatterplots of Tests

- Our final regressors are all significant in explaining the Y variable and they are all positively correlated

Figure 3: Linearity between loggdp and explanatory variables (best-fitting levels/logs only)



The Empirical Model

$$\hat{loggdp} = \hat{\beta_0} + (\beta_1 * \hat{infrastructure}) + \beta_2 * \hat{loginstitutions}) + (\beta_3 * \hat{ictadoption}) + (\beta_4 * \hat{macroeconomicstability})$$
$$+ (\beta_5 * \hat{health}) + (\beta_6 * \hat{skills}) + (\beta_7 * \hat{productmarket}) + (\beta_8 * \hat{labormarket}) + (\beta_9 * \hat{logfinancialsystem})$$
$$+ (\beta_{10} * \hat{logmarketsize}) + (\beta_{11} * \hat{businessdynamism}) + (\beta_{12} * \hat{loginnovationcapability})$$

```
$$\hat{loggdp} = \hat{\beta_{0}} + \hat{(\beta_{1} * infrastructure)} +
\hat{\beta_{2} * loginstitutions)} + \hat{(\beta_{3} * ictadoption)} +
\hat{(\beta_{4} * macroeconomicstability)} + \hat{(\beta_{5} * health)} +
\hat{(\beta_{6} * skills)} + \hat{(\beta_{7} * productmarket)} + \hat{(\beta_{8} *
labormarket)} + \hat{(\beta_{9} * logfinancialsystem)} + \hat{(\beta_{10} *
logmarketsize)} + \hat{(\beta_{11} * businessdynamism)} + \hat{(\beta_{12} *
loginnovationcapability)}$$
```

Naïve First Estimation of the Model

- The global model is significant
- Model explains 89% of the variance in dependent variable
- 8 regressors are not significant post joint regression
- 3 variables are negatively correlated

Figure 4: First regression estimation of the empirical model

```
##
## Call:
## lm(formula = GDPTFP$loggdp ~ GDPTFP$loginstitutions + GDPTFP$Infrastructure +
##     GDPTFP$"ICT Adoption" + GDPTFP$"Macroeconomic Stability" +
##     GDPTFP$Health + GDPTFP$Skills + GDPTFP$"Product Market" +
##     GDPTFP$"Labor Market" + GDPTFP$logfinancialsystem + GDPTFP$logmarketsize +
##     GDPTFP$"Business Dynamism" + GDPTFP$loginnovationcapability)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04269 -0.21592 -0.01117  0.18354  1.13529
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      5.206548   1.565909   3.325  0.00117 **
## GDPTFP$loginstitutions           0.016531   0.461968   0.036  0.97151
## GDPTFP$Infrastructure            0.022504   0.006943   3.241  0.00153 **
## GDPTFP$"ICT Adoption"            0.018918   0.004136   4.574 1.15e-05 ***
## GDPTFP$"Macroeconomic Stability" 0.006046   0.003379   1.789  0.07605 .
## GDPTFP$Health                    0.005782   0.003287   1.759  0.08109 .
## GDPTFP$Skills                    0.025706   0.006925   3.712  0.00031 ***
## GDPTFP$"Product Market"          0.009005   0.009266   0.972  0.33305
## GDPTFP$"Labor Market"           -0.004278   0.008128  -0.526  0.59962
## GDPTFP$logfinancialsystem        0.071034   0.331041   0.215  0.83045
## GDPTFP$logmarketsize             0.117360   0.134887   0.870  0.38596
## GDPTFP$"Business Dynamism"      -0.016970   0.006735  -2.520  0.01302 *
## GDPTFP$loginnovationcapability  -0.183609   0.239163  -0.768  0.44413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3731 on 123 degrees of freedom
## Multiple R-squared:  0.9036, Adjusted R-squared:  0.8942
## F-statistic:  96.1 on 12 and 123 DF,  p-value: < 2.2e-16
```

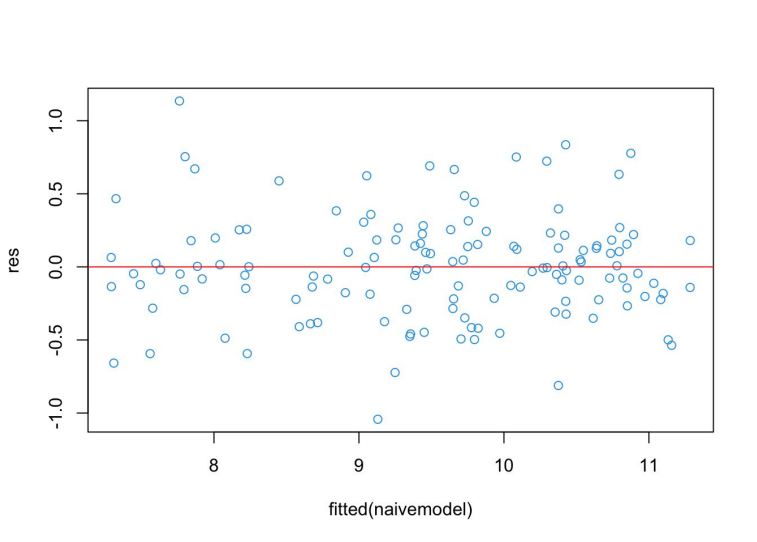
```
naivemodel <- lm(GDPTFP$loggdp ~ GDPTFP$'loginstitutions' + GDPTFP$'Infrastructure'
+ GDPTFP$'ICT Adoption' + GDPTFP$'Macroeconomic Stability' + GDPTFP$'Health' +
GDPTFP$'Skills' + GDPTFP$'Product Market' + GDPTFP$'Labor Market' +
GDPTFP$'logfinancialsystem' + GDPTFP$'logmarketsize' + GDPTFP$'Business Dynamism' +
GDPTFP$'loginnovationcapability')
print(summary(naivemodel))
```

VIF & Fitted Residuals Plot

- 8 regressors have a VIF of more than 5; near multicollinearity likely an issue
- Residuals, however, appear mostly well-behaved

Figure 5: VIF analysis

```
##           GDPTFP$loginstitutions            GDPTFP$Infrastructure
##                         8.247161                        11.570637
##          GDPTFP$"ICT Adoption" GDPTFP$"Macroeconomic Stability"
##                         6.270975                         2.831720
##                  GDPTFP$Health                    GDPTFP$Skills
##                         4.017492                        10.174774
##         GDPTFP$"Product Market"           GDPTFP$"Labor Market"
##                         5.243028                         5.103245
##       GDPTFP$logfinancialsystem              GDPTFP$logmarketsize
##                         4.759135                         2.307573
##      GDPTFP$"Business Dynamism"   GDPTFP$loginnovationcapability
##                         5.454547                         7.959280
```



Further Analysis to be Conducted at a Later Date

- Detection of outliers and influential observations
- Heteroskedasticity: Diagnostics & Correction
- Tests and Variable Selection for the Final Model
- Test for Structural Change
- Presentation of the Final Econometric Model

Takeaways and Future Work

Takeaways

- A rich curiosity
- Desire to learn more advanced coding techniques
- Problem-solving skills

Thank You

References

1. Callen, T. (2020, February 24). “Gross Domestic Product: An Economy’s All. https://www.imf.org/external/pubs/ft/fandd/basics/gdp.html. (2021, 26 March).
2. International Labour Office. (2004). “Economic Security for a Better World.” ILO Socio-Economic Security Programme
3. Schwab, K. (2018). “The Global Competitiveness Report.” World Economic Forum. pp. 43.
4. A4ID. (2016, April 25). “Understanding the Developed/Developing Country Taxonomy.”A4ID. www.a4id.org/policy/understanding-the-developeddeveloping-country-taxonomy/. (2021, 26 March).
5. The World Bank Group. (2017) “GDP per capita, PPP (constant 2017 international $)” The World Bank, data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD?end=2017. (2021, 26 March).

Questions?