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

- 1. Project Introduction
- 2. Coding the Project in R & Results
- 3. Takeaways and Future Work

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 % of the growth

Table 3: Growth Accounting for the United States

ι	3			 Contributions f 	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

(Schwab, 2018)

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

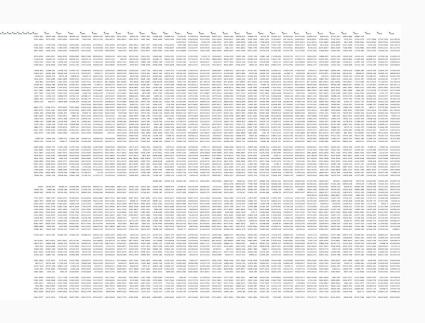
```
'Market Size' = '10th pillar: Market size
                                                                                                                                                                                                                                                                                         . 'Innovation Capability' = '12th pillar: Innovation capability')
 3 # Importing Global Competitiveness Index Report 2018 Data (produced by the World
                                                                                                                                                                                  88 GDPData «
                                                                                                                                                                                                                                                                           119 # Dropping all observations with missing values as we cannot use these observation
34 rm(list=ls(all=T))
                                                                                             as_tibble(GDPData <- read_excel("API_NY.GDP.PCAP.PP.KD_DS2_en_excel_v2_3732022.xls
                                                                                                                                                                                        mutate_at("Country Name", str_replace, "Venezuela, RB", "Venezuela")
                                                                                                                                                                                                                                                                           120 GDPTFP <- drop_na(GDPTFP)
 6 setwd("~/Documents/Sorbonne/Ca'Foscari/R Coding/R Project")
38 library(tidyverse)
                                                                                                                                                                                                                                                                           122 # Changing 'Income Group' variable to a factor variable
                                                                                                                                                                                        mutate_at("Country Name", str_replace, "Vietnam", "Viet Nam")
39 library(readxl)
                                                                                                                                                                                                                                                                           123 GDPTFP <- GDPTFP %>% mutate_at(vars(2), factor)
40 GCIndexFullData <- read excel("GCI 4.0 2018 Dataset.xlsx".</p>
                                                                                                                                                                                  96 GDPData «
                                                                                                                                                                                                                                                                           125 # Removing original datasets from the environment
                                                                                                                                                                                                                                                                           126 rm(GCIndexFullData)
                                                                                                                                                                                                                                                                           127 rm(TFPScore)
                                                                                                                                                                                                                                                                           128 rm(GDPData)
                                                                                               mutate_at("Country Name", str_replace, "Congo, Dem, Rep.", "Congo, Democratic
                                                                                                                                                                                      GDPTFP <- left_join(GDPData, TFPScore, by = "Country Name")
46 # Removing unwanted rows (observations are in columns. I will address this next)
                                                                                                                                                                                                                                                                           130 # Assigning summary to a new dataframe to print it properly, else 2 summaries are
47 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), ]</p>
                                                                                               mutate_at("Country Name", str_replace, "Cabo Verde", "Cape Verde")
49 # Transposing the dataframe to have the observations show up correctly
                                                                                                                                                                                                                                                                                datasummary <- sum_up(GDPTFP)
50 library(sjmisc)
 1 TEPScore <- TEPScore %% rotate df(rn = "Country Name", cn = TRUE)</p>
                                                                                               mutate_at("Country Name", str_replace, "Cote d'Ivoire", "Côte d'Ivoire")
                                                                                                                                                                                                                                                                                Regressors <- data_frame(GDPTFP[ , 4:15])
3 # Converting character variables into numeric
                                                                                                                                                                                                                                                                           136 result <- sum(cor(Regressors) > 0.75)
54 cols.num <= c(2:13)
55 TFPScore[cols.num] <- sapply(TFPScore[cols.num],as.numeric)</p>
                                                                                              GDPData %>%
56 sapply(TFPScore, class)
                                                                                                mutate_at("Country Name", str_replace, "Egypt, Arab Rep.", "Egypt"
```

The most challenging part

Importing and Tidying Data: World Bank Dataset

VARIABLES

OBSERVATIONS



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

OBSERVATIONS -

#PORTANT: The storage on any data re strictly prohibited without prior permise	ion of the data provi																								
ead the Terms of Use in the "About this I	Datasel tab.																								
ndex	▼ Edition	V Series Global D	* Freeze dala	Series name	V Series units	Series order	* Series code (*)	Adribusto	V Angola V	ALB Abenia	ARE United Arab	ARG Argentina	ARM * Armenia	AUS V Australia	AUT Austin	AZE Azerbajan	Burundi 1	BEIL Belgium	BEN Banin	BrA Burking Fas	BGD Bangladeah	BGR Bulgaria	Bahrain 1	Bosnis and	W Bolivia
Hobal Competitiveness Index 4.0		2018 GCM	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)			DATE DESCRIPTION	2018 edition	2018 edition	2018 editor	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition								
llobal Competity eness Index 4.0		2018 GCH	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)		1	SCORE	8 7.056788112	5 8.10159560	073,3719000	84 57,48611008	8859.8614811	14 78.85166974	5 76.33810153	8 50.04414801	3 57.52763496	576.6101553	90 4 39658876	7 43,69803312	34 52,07923884	9 43,5599736	3 53 63291244	8 54.15512388	84 513966
Blobal Competitiveness Index 4.0		2018 GCM	17/10/2018	Global Compelitiveness Index 4.0	0-100 (best)		1	NOTE																	
Nobal Competitiveness Index 4.0		2018 GCH	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)			VALUE																	
Slobal Competitiveness Index 4.0		2018 GCM	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)			SOURCE				om World Econo													
Blobal Competitiveness Index 4.0		2018 GCM	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)			SOURCEDATE	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018				17/10/2018	17/10/2018	17/10/2018	17/10/20
liobal Competitiveness Index 4.0		2018 GCM	17/10/2018	Global Competitiveness Index 4.0	0-100 (best)			RANK	137	76	27	6 1	70	F14	22	69	F 36	21	F23	F24	7 03	51	50	91	105
Slobal Competitiveness Index 4.0		2018 GCH.SUBIDXA	17/10/2018	Enabling environment	0-100 (best)			NOTE																	
Slobal Competitiveness Index 4.0		2018 GCH.SUBIDXA	17/10/2018	Enabling environment	0-100 (best)			SOURCEDATE		17/10/2018	17/10/2018						17/10/2018	17/10/2018		17/10/2018	17/10/2018		17/10/2018	17/10/2018	17/10/20
Global Competitiveness Index 4.0		2018 GCW.SUBIDXA	17/10/2018	Enabling environment	0-100 (best)			DATE DESCRIPTION				2018 edition											2018 edition		
Blobal Competitiveness Index 4.0 Blobal Competitiveness Index 4.0		2018 GCH.SUBDXA 2018 GCH.SUBDXA	17/10/2018	Enabling environment	0-100 (best) 0-100 (best)			SOURCE VALUE	World Econom	World Econo	m World Econ	om World Econo	m World Econo	om World Econor	m World Econor	m World Econo	m World Econor	m World Econo	em World Econo	m World Econo	at World Econo	m World Econo	m World Econor	4 World Econor	em World Ec
Global Competitiveness index 4.0		2018 GCK SUBDXA	17/10/2018	Enabling environment				RANK		w.				w.			Was.		W	West	Was .			W-	
	_	2018 GCM.SUBIDXA 2018 GCM.SUBIDXA	17/10/2018	Enabling environment	0-100 (best) 0-100 (best)			SCORE	na/	44		01 \$5,06536794	98	21	20	78	130	63	725	718	F104	49	90	67	748
Slobal Compositiveness Index 4.0 Slobal Compositiveness Index 4.0		2018 GCM-SUBDXA 2018 GCM-A01	17/10/2018	Enabling environment 1stailar: his litrions	0-100 (best) 0-100 (best)	_		SOURCE DATE	17/10/2018		17/10/2018							17/10/2018		17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/20
Slobal Competitiveness Index 4.0 Slobal Competitiveness Index 4.0	_	2018 GCMA01 2018 GCMA01	17/10/2018	1stpilar: hsitutons	0-100 (best) 0-100 (best)	_		NOTE	17/10/2018	17/14/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/19/2018	17/10/2018	17/10/2018	17710/2018	17/10/2018	1//10/20
Slobal Competitiveness Index 4.0 Slobal Competitiveness Index 4.0		2018 GCHA01 2018 GCHA01	17/10/2018	1stpiler: heliulons 1steller: heliulons	0-100 (best) 0-100 (best)			NOTE SOURCE	Weeld France	Washi Erasa	m World From	om World Econo	or World Conn.	on World Conne	w World France	world Fores	w World Econo	World France	Weeld France	world Engra	World Front	m World From	w World Cooper	Washi Faran	non Milandel E.
Global Competitiveness Index 4.0		2018 GCHA01	17/10/2018	1stellar hallufors	0-100 (best)			VALUE	World Econom	n world Econo	m world Econ	om world Econo	im world Econ	om waria Econor	m world Econor	m world Econo	m World Econor	m World Econo	en wene scono	m world Econo	A World Econo	m world Econo	m world Econor	I Waria Econor	JM WORLD EX
Global Competitiveness Index 4.0		2018 GCH A01	17/10/2018	Istoliar hylkdons	0-100 (best)			DATE DESCRIPTION	2018 edition	2010 -45	2010 15-1	2018 edition	2010 of fee	2016 edition	2010	2018 edition	2018 edition	2018	2018 edition	2018 edition	2018 edition	2010	2018 edition	2015 45	2010 ad
Global Competitiveness index 4.0		2018 GCMA01	17/10/2018	1stellar hsilutors	0-100 (best)			SCORE				11 51 543 14539													
Slobal Competitiveness Index 4.0		2018 GCHA01	17/10/2018	1stpilar: hsBulons	0-100 (best)			RANK		Ea.	F a	H's	E.	Ks	Ke	Ke	F35	E1	K14	Ka .	FOR	Mo.	E2	Ktt	F/29
Slobal Competitiveness Index 4.0		2018 GCHA01.01	17/10/2018	Security	0-100 (best)			VALUE	104	-	-10		41		-10	-	7100	-			100	10			12.0
Slobal Competitiveness Index 4.0		2018 GCHA01.01	17/10/2018	Security	0-100 (best)			RANK	Mos	K1	K	Fox	Eq.	Mr.	F3	Ec.	W13	Es.	Wy.	Wa.	Mon	No.	Fo.	K4	Fig.
Blobal Competitiveness Index 4.0		2018 GCHA01.01	17/10/2018	Security	0-100 (best)		4	SOURCEDATE	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/20
Global Competitiveness Index 4.0		2018 GCHA01.01	17/10/2018	Security	0-100 (bost)		4	SCORE			4 51,0029261	28 48.36877207	0 402155979												33 58 6930
Slobal Competitiveness Index 4.0		2018 GCH A01.01	17/10/2018	Security	0-100 (best)		4	DATE DESCRIPTION	2018 edition	2018 edition	2018 editor	2018 edition	2018 edition	2016 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2018 edition	2016 edition	2018 ed
Giobal Competity eness Index 4.0		2018 GCM A01.01	17/10/2018	Security	0-100 (best)		4	NOTE			20.00	-	-		-	-		-	-		-			-	-
Slobal Competitiveness Index 4.0		2018 GCHA01.01	17/10/2018	Security	0-100 (best)		4	SOURCE	World Econom	World Econo	m World Econ	om World Econo	m World Econe	om World Econor	m World Econor	m World Econo		m World Econo	en World Econo		m World Econo		m World Econor	n World Econor	om World Er
Slobal Competitiveness Index 4.0		2018 EDBQ035	17/10/2018	Organized crime (1-7)	1-7 (best)		6 1.01	RANK	154	F108	F13	Flos	E 7	27	23	59	F19	Mg .	# 3	F123	19 7	F16	F16	F03	6 2
Slobal Competitiveness Index 4.0		2018 EOSQ035	17/10/2018	Organized crime (1-7)	1-7 (best)		5 1.01	SCORE	56.994048745	51,72652403	3.51.7608833	33 52.03025341	659.6105718	66 75.21024545	#7.87103513	3 71.72685465	0 46.27664883	3 78.3794323	6 57.00147151	6 41.30434593	3 55.50560951	6 47.1117893	3 40.25722503	3 53.11087766	66 56.3468
Slobal Competitiveness Index 4.0		2018 E090035	17/10/2018	Organized crime (1-7)	1-7 (best)		5 1.01	DATE DESCRIPTION	2018			rer Weighted av											er Weighted ave		
Global Competitiveness Index 4.0		2018 EOSQ035	17/10/2018	Organized crime (1-7)	1-7 (best)		5 1.01	VALUE	X419642925	M.103591442	5.905653	W.121815205	5.17643431	2 5.512614727	E.660262108	5.30361127	5,77659893	5.70276594	Z ■420088291	5.478260756	₹.33033657°	5.82670736	₹815433502	M.18665266	3.980808
Slobal Competitiveness Index 4.0		2018 EOSQ035	17/10/2018	Organized crime (1-7)	1-7 (best)		5 1.01		100000000000000000000000000000000000000																
Global Competitiveness Index 4.0		2018 E09Q035	17/10/2018	Organized crime (1-7)	1-7 (best)			90URCEDATE	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/20
Slobal Competitiveness Index 4.0		2018 EOSQ035	17/10/2018	Organized crime (1-7)	1-7 (best)			SOURCE				om World Econo			m World Econor			m World Econo						n World Econor	
Slobal Competitiveness Index 4.0		2018 HOMICIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pop.			DATE DESCRIPTION	2012	2016	2016	2016	2016	2016	2016	2016	2016	2015	2 015	2015	2015	₹016	型014	2016	2016
Global Competitiveness Index 4.0		2018 HOMCIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pep.		6 1.02		nia	nis	rula	mis	nis	n/a	n/a	nis	nia	n/a	rula	nis	mia	mis	nis	n/a	n/a
Blobal Competitiveness Index 4.0		2018 HOMCIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pop.			SOURCE				ns United Nation													
Global Competitiveness Index 4.0		2018 HOMCIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pop.			SOURCEDATE	624,2018	6/24/2018	6/24/2018	6/24/2018	6/24/2018	6/24/2018	6242018	6242018	6/24/2018	6242018	624/2018	6/24/2018	6/24/2018	6/24/2018	6/24/2018	6/24/2018	624201
Slobal Competitiveness Index 4.0		2018 HOMICIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pop.		6 1.02		68	70	27	95	771	29	F18	63	96	60	F67	R.	66	57	A1	42	98
Slobal Compositiveness Index 4.0		2018 HOMICIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pop.			BCORE				00 1.55062896											099.91933407		
Slobal Competitiveness Index 4.0		2018 HOMCIDERT	17/10/2018	Homicides per 100,000 population	/100,000 pep.			VALUE				45 5.942564450													
Blobal Competitiveness Index 4.0		2018 TERRORISMINCIDENCE		Terrorism incidence (0=very high; 100=no incidence)	1 (very high) - 7 (no incidence			DATE DESCRIPTION				our Weighted cor													
Blobal Compositiveness Index 4.0		2018 TERRORISMNCIDENCE		Terrorism incidence (0=very high; 100=no incidence)	1 (very high) - 7 (no incidence			SOURCE				cul Authors' calc													
Blobal Competitiveness Index 4.0		2018 TERRORISMINCIDENCE		Terrorism incidence (0=very high; 100+no incidence)	1 (very high) - 7 (no incidence			VALUE	95.332667377	99.90074157	7 59,9874496	45 59 84677886	19799.8938293	45.99.23578643	7.99.87634277	3 99.7338027	5 91,27849578	8/98.0012617	38700	96.15996077	73 4 5,93446350	0799,9445266	298.08412933	J 99.88481903	30 M CO
liobal Competitiveness Index 4.0		2018 TERRORISMINCIDENCE		Terrorism incidence (0=very high; 100=no incidence)	1 (very high) - 7 (no incidence		7 1.03	NOTE RANK					-								No.				-
liobal Compositiveness Index 4.0		2018 TERRORISMINCIDENCE		Terrorism incidence (0=very high; 100=no incidence)	1 (very high) - 7 (no incidence				800018	151	81	71	62	488	-06	778	nz1	704	7	M13		50	P103	64	-
lobal Competitiveness Index 4.0 lobal Competitiveness Index 4.0	_	2018 TERRORISMINCIDENCE 2018 TERRORISMINCIDENCE		Terrorism incidence (0=very high; 100=no incidence) Terrorism incidence (0=very high; 100=no incidence)	1 (very high) - 7 (no incidence			SOURCE DATE SCORE		8/2/2018	8/2/2018	8/2/2018 45/89 84677886	8/2/2018	82,2318	8/2/2018	8/2/2018	8/2/2018	8/2/2018	8/2/2018	8/2/2018	82/2018 23 85 93446350	8/2/2018	8/2/2018	8/2/2018	8/2/2011
Slobal Competitiveness Index 4.0 Slobal Competitiveness Index 4.0	_	2018 TERRORISMINCEDENCES 2018 EDSQ055	17/10/2018		1 (very high) - 7 (no incidence 1-7 (best))	7 1.03 8 1.04		P05.332687377	199.90074157	7709.9874496	45/99/54677886	59799.8938293	45 99.23578643	1799.87634211	3 99 7338027	0 9127849578	8 58.0012817	38700	96.15998077	3 65.93446360	0 99.9445266	298.08412933	P99.88481903	J0 P100
lobal Competitiveness Index 4.0 lobal Competitiveness Index 4.0	_	2018 EOSQ005 2018 EOSQ005	17/10/2018	Reliability of police services (1-7) Reliability of police services (1-7)	1-7 (best) 1-7 (best)	_		NOTE SOURCEDATE	17/10/2018	17/10/2018	17/10/2014	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	12/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	17/10/2018	478000
Robal Competitiveness Index 4.0		2018 EDSQUSS 2018 FDSQUSS	17/10/2018	Reliability of police services (1-7)	1-7 (best)			VALUE				8 5.402845621													
Nobal Competitiveness Index 4.0	_	2018 EUSQUSS 2018 FORQUSS	17/10/2018	Reliability of police services (1-7) Reliability of police services (1-7)	1-7 (best)	_		WALUE SCORE				33 40 04742701													
Hobal Competity eness Index 4.0 Hobal Competity eness Index 4.0	_	2018 ED9Q055 2018 ED9Q055	17/10/2018	Reliability of police services (1-7) Reliability of police services (1-7)	1-7 (best) 1-7 (best)			DATE DESCRIPTION	03.6206913			rer Weighted av											or Weighted ave		
lobal Competitiveness Index 4.0 lobal Competitiveness Index 4.0		2018 EDSQ055 2018 EDSQ055	17/10/2018	Reliability of police services (1-7) Reliability of police services (1-7)	1-7 (best) 1-7 (best)			RANK	E24	No.	Ke Teensell a	rer Weighted av	or Weighted av	er Weighted are	K4	er Weighted av	F25	Ko Ko	K.	E2	Moighted avo	Mo	Ke Negranday	Kon Son Sare	E24
lobal Compatity eness index 4.0		2018 ED9Q055	17/10/2018	Reliability of police services (1-7)	1-7 (best)			SOURCE	World France	Waste France	- World From	om World Econo						World France	or World France	Washi France			w World France	World France	World E
labal Competitiveness Index 4.0		2018 GCH A01.02	17/10/2018	Relability of police services (1-7)	0-100 (best)			SOURCE				om World Econo													
lobal Competitiveness Index 4.0	_	2018 GCWA01.02	17/10/2018	Social capital	0-100 (best) 0-100 (best)			SOURCEDATE				17/10/2018		17/10/2018				17/10/2018		17/10/2018		17/10/2018	17/10/2018		
lobal Competitiveness Index 4.0 lobal Competitiveness Index 4.0		2018 GCWA01.02	17/10/2018	Social capital	0-100 (best) 0-100 (best)			WALUE	11/10/2018	11/14/2018	17710/2018	17/13/2018	177702018	177102018	177702018	177742018	1777442018	17/10/2018	1771-02018	1771422018	177192018	177102018	17716/2018	17710/2018	17710(2)
lobal Competitiveness index 4.0	_	2018 GCMA01.02	17/10/2018	Social capital	0-100 (best) 0-100 (best)			DATE DESCRIPTION	2010 - 47	2010 - 48	2040 - 48-	2018 edition	2010 - 45	2010 - 45	2010 - 450-	2242 - 45	2242 - 45	2011	2010 - 48	2010 - 17	2040 - 47	2010 491-	2010 - 45	2010 - 45	2010

```
`{r, eval=TRUE, echo=FALSE, include=FALSE, results='hide'}
31 # DATA IMPORT AND TIDYING
33 # 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
34 rm(list=ls(all=T))
36 setwd("~/Documents/Sorbonne/Ca'Foscari/R Coding/R Project")
38 library(tidyverse)
39 library(readxl)
40 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)))
46 # Removing unwanted rows (observations are in columns. I will address this next)
47 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),
49 # Transposing the dataframe to have the observations show up correctly
51 TFPScore <- TFPScore %>% rotate_df(rn = "Country Name", cn = TRUE)
53 # Converting character variables into numeric
54 cols.num <- c(2:13)
55 TFPScore[cols.num] <- sapply(TFPScore[cols.num],as.numeric)
56 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"

```
57 TFPScore <- as_tibble(TFPScore)</p>
58 rm(cols.num)
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"
                                    sheet = "Data".
                                    col_types = c("text", "skip", "text", rep("skip",
    59), "numeric", rep("skip", 3))))
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 <-
     mutate_at("Country Name", str_replace, "Congo, Dem. Rep.", "Congo, Democratic
    Rep.")
   GDPData <-
     mutate_at("Country Name", str_replace, "Cabo Verde", "Cape Verde")
76 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")
```

← Imported World Bank Dataset

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

```
GDPData <-
85
      GDPData %>%
      mutate_at("Country Name", str_replace, "North Macedonia", "Macedonia, FYR")
87
88
    GDPData <-
89
       GDPData %>%
      mutate_at("Country Name", str_replace, "Venezuela, RB", "Venezuela")
92
    GDPData <-
       GDPData %>%
94
      mutate_at("Country Name", str_replace, "Vietnam", "Viet Nam")
    GDPData <-
97
       GDPData %>%
      mutate_at("Country Name", str_replace, "Yemen, Rep.", "Yemen")
99
   # Joining the datasets so as to work within a single dataframe
    GDPTFP <- left_join(GDPData, TFPScore, by = "Country Name")
102
103 # Renamina Columns
    GDPTFP <- rename(GDPTFP, '2017 GDP Per Capita' = '2017'
105
              'Income Group' = 'IncomeGroup'
              'Institutions' = '1st pillar: Institutions'
107
              'Infrastructure' = '2nd pillar: Infrastructure'
108
              'ICT Adoption' = '3rd pillar: ICT adoption'
              'Macroeconomic Stability' = '4th pillar: Macroeconomic stability'
109
110
              '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

```
115
              'Market Size' = '10th pillar: Market size'
             'Business Dynamism' = '11th pillar: Business dynamism'
             'Innovation Capability' = '12th pillar: Innovation capability')
119 # Dropping all observations with missing values as we cannot use these observations
     for regression treatment and analysis
120 GDPTFP <- drop_na(GDPTFP)
122 # Changing 'Income Group' variable to a factor variable
123 GDPTFP <- GDPTFP %>% mutate_at(vars(2), factor)
125 # Removing original datasets from the environment
126 rm(GCIndexFullData)
127 rm(TFPScore)
128 rm(GDPData)
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)
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
```

117

124

129

- ← 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						

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	explanator	y variables
-----------	-------------	---------	------------	-------------

	InstitutionsInfrast	tructure	ICTMa	acroeconomic Stability	lealth	Skills	Product				BusinessInno	
		Au	option	Stability			ividi itotii	narket	System	SizeL	ynamism Cap	ability
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

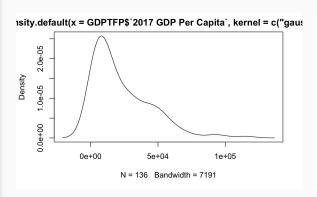
```
Regressors <- data_frame(GDPTFP[ , 4:15])
result <- sum(cor(Regressors) > 0.75)
value <- (result - 12) / 2</pre>
```

```
(`r {value}` unique values).
```

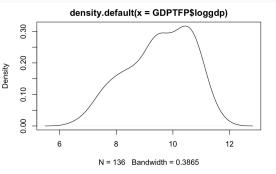
```
library(kableExtra)
knitr::kable(cor(Regressors), digits=2)
```

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"))</pre>
```

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$`Institutions`)</pre>
plot(density(GDPTFP$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 o
plot(GDPTFP$`loginstitutions`, GDPTFP$loggdp)
abline(lm(GDPTFP$loggdp ~ GDPTFP$`loginstitutions`))
#NUIE: loggap is a better fit with data appearing relatively normal on both sides o
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`))
```

#NOTE: "loginstitutions" has a higher R-squared value. Thus, keep "loginstitutions"

summary(lm(GDPTFP\$loggdp ~ GDPTFP\$`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

```
ExplanatoryPower_OrigVsLog <- function(a, b) {</pre>
 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: logadp,
loginstitutions, Infrastructure, ICT Adoption, Macroeconomic Stability, Health,
Skills, Product Market, Labor Market, loafinancialsystem, loamarketsize, Business
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

```
SignificanceTest <- function(a) {</pre>
  result <- summary(lm(GDPTFP$logqdp \sim a))$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 logadp, our dependent
variables.
```

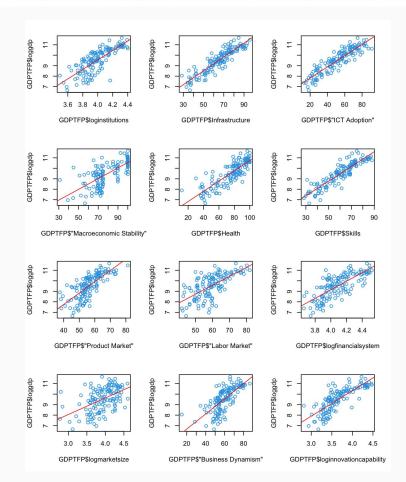
RESULTS

```
(Intercept)
                    TRUE
(Intercept)
                    TRUE
(Intercept)
                    TRUE
(Intercept)
                    TRUE
(Intercept)
       TRUE
                    TRUE
(Intercept)
                    TRUE
(Intercept)
       TRUE
                    TRUE
(Intercept)
                    TRUE
(Intercept)
       TRUE
                    TRUE
(Intercept)
                    TRUE
(Intercept)
                    TRUE
(Intercept)
      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

```
loggdp = \hat{\beta_0} + (\beta_1 * infrastructure) + \beta_2 * loginstitutions) + (\beta_3 * ictadoption) + (\beta_4 * macroeconomicstability)
+ (\beta_5 * \hat{h}ealth) + (\beta_6 * \hat{s}kills) + (\beta_7 * productmarket) + (\beta_8 * labormarket) + (\beta_9 * logfinancial system)
+ (\beta_{10} * logmarket size) + (\beta_{11} * busines s dynamism) + (\beta_{12} * loginnovation capability)
```

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

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$loginnovationcapability)
## Residuals:
                 10 Median
## -1.04269 -0.21592 -0.01117 0.18354 1.13529
## Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                                    5.206548
                                              1.565909
                                                         3.325 0.00117 **
## GDPTFP$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$"Labor Market"
                                   -0.004278
                                               0.008128 -0.526 0.59962
## GDPTFP$logfinancialsvstem
                                    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
```

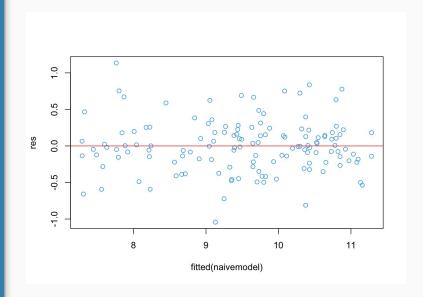
```
naivemodel <- lm(GDPTFP$loggdp ~ GDPTFP$'loginstitutions' + GDPTFP$'Infrastructure'
+ GDPTFP$'ICT Adoption' + GDPTFP$'Macroeconomic Stability' + GDPTFP$'Health' +
GDPTFP$'Skills' + GDPTFP$'Product Market' + GDPTFP$'Labor Market' +
GDPTFP$'logfinancialsystem' + GDPTFP$'logmarketsize' + GDPTFP$'Business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

- 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?