UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Understanding the Variance in Gross Domestic Product Per Capita Explained by Competitiveness Factors

Midterm Paper - Applied Econometrics

Sarbjit GILL and Jonathan MORALES

Submitted: March 31, 2021

Table of Contents

Report
1. Introduction & Related Literature - 3 -
2. Data Description - 4 -
3. Econometric Methodology - 6 -
4. Detection of Outliers and Influential Observations - 7 -
5. Comments on Regression - 10 -
6. Heteroskedasticity: Diagnostics & Correction - 11 -
7. Tests and Variable Selection - 11 -
8. Test for Structural Change - 13 -
9. Presentation of the Final Econometric Model - 14 -
10. Conclusions - 18 -

References - 19 -

Appendix I: 2018 Global Competitiveness Index Pillars: Competitiveness Factor Importance and Score Composition - 20 -

Appendix II: Project Stata Files - 23 -

Understanding the Variance in Gross Domestic Product Per Capita Explained by Competitiveness Factors

1. Introduction & Related Literature

Gross Domestic Product (GDP) is defined as "the monetary value of final goods and services produced in a country in a given period of time" (Callen, 2020). In its raw value-form, GDP provides information on the size of an economy and its value for the given period. GDP per capita further breaks down the value of GDP to economic output per person in that economy. Another metric, real GDP, adjusts the raw GDP value for inflation. This metric is used to understand economic growth or decline over time.

The state of an economy has an impact on the lives of the constituents residing in that economy (International Labour Office, 2004), as they form a large part of the labor market that produces goods and services in exchange for wages, which they then use for necessities, leisure and savings. Thus, economic growth is generally a healthy sign—though not without caveats, especially in the presence of substantial inequality, poor laws, and disparate benefits across subpopulations—but an economic decline poses an imminent risk to the livelihoods of people who come to depend on the wages and a certain quality of life. Therefore, real GDP and proponents of growth are still valuable subjects that warrant study.

Economic growth is generally composed of the following: "(1) growth in labour force, (2) growth in physical and natural capital inputs, and (3) total factor productivity growth (TFP) growth, the "unexplained part" of GDP growth, which encompasses all non-physical inputs, such as technological progress, human capital, and institutional and cultural factors" (Schwab, 2018). While (1) and (2) may be easily studied, it is the TFP that is the most convoluted metric and the most influential in driving economic growth. The World Economic Forum's Global Competitiveness Index (GCI), published annually, aims to understand the total factor productivity growth factors of national economies to uncover underlying differences propelling unexplained growth. The 2018 GCI framework is composed of 98 indicators, computing an overall competitiveness score for observed economies 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. Appendix I, on page 20, details each factor’s importance and score composition.

The purpose of our econometric project is to understand which 2018 GCI productivity drivers best explain the variance in 2017 real GDP per capita of national economies (if any). This information is valuable in understanding where the significant competitiveness gaps lie and what investments economies may make to catch up on the TFP growth.

2. Data Description

The data for this econometrics project comes from the World Economic Forum’s 2018 Global Competitiveness Index (Schwab, 2018), where 140 economies were scored across 12 main drivers of productivity via 98 indicators. The score for each productivity driver is numeric, and the allotted score range is 0-100 for each. The real GDP per capita values (constant, 2017 PPP) used in this project, as well as the country income classification, come from the World Bank website (The World Bank Group, 2021). The GCI and World Bank data were combined to create a single dataset, and four observations were discarded. Hong Kong was excluded as it is not a national economy but a special administrative area. Venezuela, Yemen and Taiwan were excluded due to missing GDP values. The final number of observations is 136 national economies, of which 85 are developed and 51 are developing. The significance level of 0.05 is used in our work based on the sample size. Project data and Stata do files can be accessed at the links shared in Appendix II (see page 23).

Upon first checking the data using the summarize command (see figure 1, page 4), we found no missing values. We did, however, uncover variables with a high standard deviation and large gaps between the minimum and maximum values, indicating the potential of outliers and influential variables. These include GDP per capita, health, ICT adoption, market size, and innovation capability.

Variable	Obs	Mean	Std. Dev.	Min	Max
countryname	0				
gdppercapita	136	23049.35	21820.92	773.5718	112666.8
institutions	136	55.3358	11.19791	32.92204	81.55431
infrastruc~e	136	65.25618	15.73323	28.57425	95.70355
ictadoption	136	51.84894	19.44223	12.77239	91.25512
macroecono~y	136	79.86616	15.99119	31.0625	100
health	136	75.0883	19.5816	11.93262	100
skills	136	60.57254	14.79349	28.24768	87.87808
productmar~t	136	56.41957	7.936078	37.51731	81.22964
labormarket	136	59.34682	8.925753	41.99579	81.88546
financials~m	136	61.24389	13.30192	38.71931	92.11703
marketsize	136	53.83298	17.73553	15.99203	100
businessdy~m	136	59.4486	11.13687	14.8974	86.48944
innovation~y	136	42.1531	17.30883	16.78176	87.52204

Figure 1: Data Summary

Next, we checked for correlation among our explanatory variables (see figure 2, page 5), using >0.7 as our reference. We found a number of them to be closely correlated. Thus, near multicollinearity may pose an issue. However, since there are no data entry errors, we'll keep the variables and investigate further.

	instit~s	infras~e	ictado~n	macroe~y	health	skills	produc~t	laborm~t	financ~m	market~e	busine~m	innova~y
institutions	1.0000											
infrastruc~e	0.8159	1.0000										
ictadoption	0.8022	0.8736	1.0000									
macroecono~y	0.7356	0.7181	0.7023	1.0000								
health	0.6598	0.8362	0.7902	0.6143	1.0000							
skills	0.8425	0.8960	0.8962	0.6991	0.8113	1.0000						
productmar~t	0.8626	0.7815	0.7538	0.7014	0.6454	0.7903	1.0000					
labormarket	0.8388	0.6686	0.7051	0.6420	0.5420	0.7835	0.8215	1.0000				
financials~m	0.8285	0.7940	0.7149	0.7483	0.6665	0.7534	0.7624	0.7180	1.0000			
marketsize	0.4023	0.6175	0.4428	0.4521	0.4590	0.4467	0.3574	0.2522	0.5575	1.0000		
businessdy~m	0.8506	0.7860	0.7369	0.6923	0.6274	0.7955	0.8164	0.7876	0.7767	0.5000	1.0000	
innovation~y	0.8583	0.8345	0.7633	0.7250	0.6753	0.8073	0.7624	0.7439	0.8490	0.6227	0.8263	1.0000

Figure 2: Correlation between explanatory variables

Next, we checked kernel densities of all our variables, overlaying a normal curve for comparison to see if our data is close to following a normal distribution. We also generated logs for all variables and checked their kernel densities in the same manner. Then we compared the line of fit (linearity) of all level and log variables against the dependent variable using scatterplots (see figure 3, page 5). We found that the log of our dependent variable was a better fit with our explanatory variables. After doing the graphical and regression tests, logs of the following variables were found to be a better fit with the log of our dependent variable: institutions, financial system, market size, and innovation capability. All of our explanatory variables are significant per the t-test (p-value<0.05) and positively correlated.

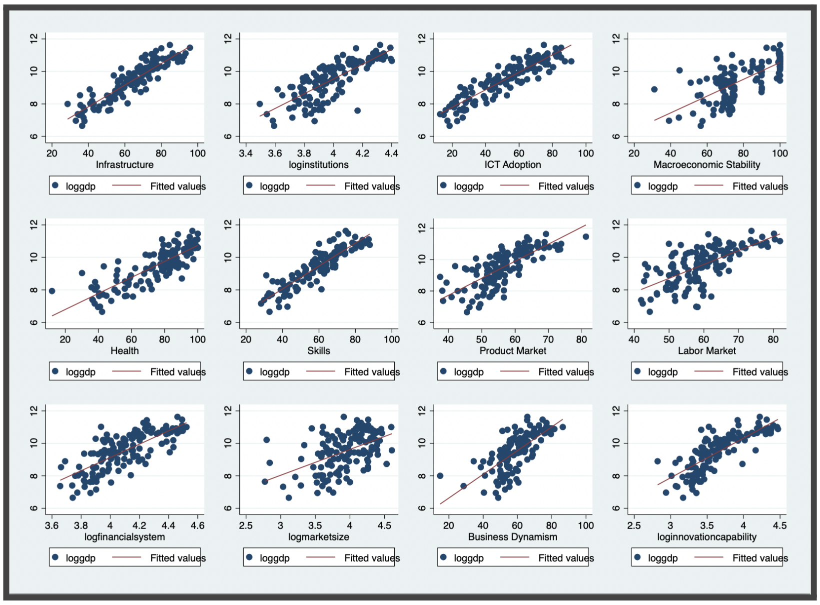

Figure 3: Linearity between loggdp and explanatory variables (best-fitting levels/logs only)

3. Econometric Methodology

Based on our research motive and data transformation of select variables (level to logs), we'll use the following empirical econometric model to understand which 2018 GCI productivity drivers best explain the variance in 2017 real GDP per capita (PPP) of national economies, if any:

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

where, 2017 real GDP per capita (PPP) is our dependent Y variable in the log form, β_0 is the constant, β_1-β_{12} are coefficients of the explanatory variables, u is the error term, and the 12 explanatory variables are: infrastructure, institutions *(log form)*, ICT adoption, macroeconomic stability, health, skills, product market, labor market, financial system *(log form)*, market size, business dynamism, and innovation capability *(log form)*. Our model above is in terms of observation i. Our goal is to predict the best regression line, using our sample and explanatory variables, to explain the variance in *loggdp*:

$$\widehat{loggdp} = \hat{\beta}_0 + \hat{\beta}_1 * infrastructure + \hat{\beta}_2 * loginstitutions + \hat{\beta}_3 * ictadoption + \hat{\beta}_4 * macroeconomicstability + \hat{\beta}_5 * health + \hat{\beta}_6 * skills + \hat{\beta}_7 * productmarket +$$

$\widehat{\beta}_8 * labormarket + \widehat{\beta}_9 * logfinancialsystem + \widehat{\beta}_{10} * logmarketsize +$
$\widehat{\beta}_{11} * businessdynamism + \widehat{\beta}_{12} * loginnovationcapability$

3.1 First Estimation of the Model

The first estimation of our full model (see figure 4, page 7) reveals that our global model is significant (fisher test p-value < 0.05). It has an adjusted R-square of 89.17%, explaining 89% of the variance in our dependent variable. However, 8 of our variables are not significant post-regression (t-test p-value > 0.05): institutions *(log form)*, macroeconomic stability, health, product market, labor market, financial system *(log form)*, market size *(log form)*, and innovation capability *(log form)*. On the other hand, the variable "business dynamism" was positively correlated before regression, but it is now negatively correlated. The VIF analysis shows (see figure 5, page 7) that 8 of our 12 variables have a variance inflation factor of more than five and a mean VIF of 6.16. Near multicollinearity may likely be the issue affecting our model. Finally, our residuals appear mostly well-behaved, but there may be a heteroskedasticity issue (see figure 6, page 7). We will investigate our first regression estimate further to refine it. However, we will not remove any variables at this stage.

Source	SS	df	MS			
				Number of obs	=	136
				F(12, 123)	=	93.62
Model	162.744579	12	13.5620483	Prob > F	=	0.0000
Residual	17.8177114	123	.144859442	R-squared	=	0.9013
				Adj R-squared	=	0.8917
Total	180.562291	135	1.33749845	Root MSE	=	.3806

loggdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
infrastructure	.0232087	.0070822	3.28	0.001	.0091899	.0372274
loginstitutions	-.0849506	.4712078	-0.18	0.857	-1.017678	.8477764
ictadoption	.0189854	.0042192	4.50	0.000	.0106338	.027337
macroeconomicstability	.0065867	.0034471	1.91	0.058	-.0002365	.01341
health	.0061384	.003353	1.83	0.070	-.0004987	.0127755
skills	.0251572	.0070631	3.56	0.001	.0111762	.0391383
productmarket	.0102906	.0094513	1.09	0.278	-.0084177	.0289988
labormarket	-.0043035	.0082906	-0.52	0.605	-.0207142	.0121072
logfinancialsystem	.0205116	.3376623	0.06	0.952	-.6478702	.6888935
logmarketsize	.1135743	.1375846	0.83	0.411	-.1587659	.3859146
businessdynamism	-.0157529	.0068695	-2.29	0.024	-.0293505	-.0021552
loginnovationcapability	-.1877204	.243946	-0.77	0.443	-.6705966	.2951558
_cons	5.610141	1.597228	3.51	0.001	2.448526	8.771755

Figure 4: First regression estimation of the empirical model

Variable	VIF	1/VIF
infrastruc~e	11.57	0.086426
skills	10.17	0.098282
loginstitu~s	8.25	0.121254
loginnovat~y	7.96	0.125640
ictadoption	6.27	0.159465
businessdy~m	5.45	0.183333
productmar~t	5.24	0.190729
labormarket	5.10	0.195954
logfinanci~m	4.76	0.210122
health	4.02	0.248912
macroecono~y	2.83	0.353142
logmarkets~e	2.31	0.433356
Mean VIF	6.16	

Figure 5: VIF analysis

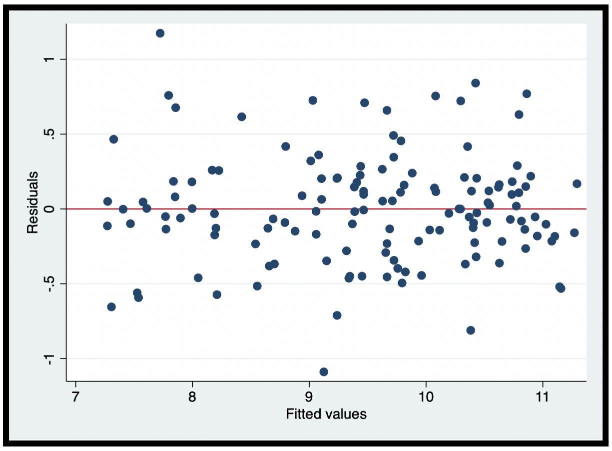

Figure 6: Residual-versus-fitted plot

4. Detection of Outliers and Influential Observations

Using the lvr2plot command post-regression, we carried out the leverage-versus-squared-residual plot test (see figure 7, page 8). There are several influential observations and outliers in the dataset. Haiti has the highest leverage, and Angola has the largest residual squared. We shall investigate the outliers and influential observations further. Chow test, to be conducted later, may also explain influential observations and outliers.

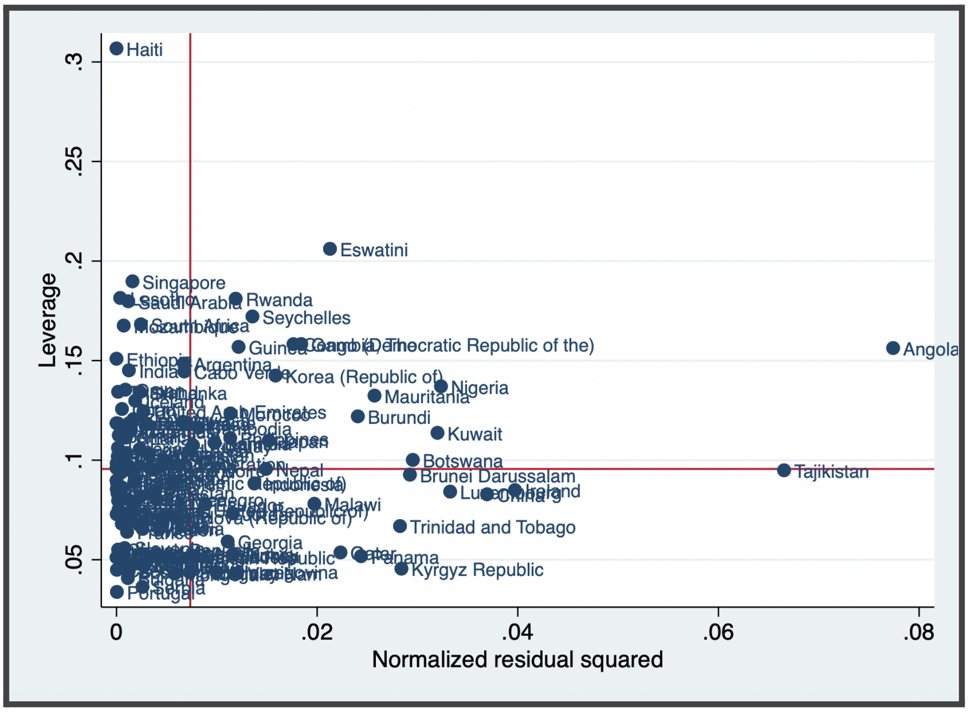

Figure 7: Leverage-versus-squared-residual plot test

4.1 Outlier Diagnostics

4.1.1. Observations with high leverage

After running the “predict lev, leverage” command and displaying the observations by leverage on the stem plot using “stem lev,” we ran the “list countryname lev if lev > .19117647” command using (2k+2)/n as the cutoff, where k is the number of variables in our model and n is the sample size (see figure 8, page 8). Haiti and Eswatini have high leverage, and we’ll examine them further in a bit.

```
. list countryname lev if lev > .19117647

      +----------------------+
      | countr~e         lev |
      |----------------------|
 17.  |    Haiti    .3067962 |
126.  | Eswatini    .2061377 |
      +----------------------+
```

Figure 8: Observations with high leverage

4.1.2. Observations with high studentized residuals

Next, we ran the “predict r, rstudent” command and chose to display the observations that had studentized residuals of more than +2 or less than -2. Studentized residuals provide an estimate of error varying between points. We used the “list countryname r if r > 2” and “list countryname r if r < -2” commands to display observations exceeding our criteria and 9 appeared (see figure 9, page 9). We’ll examine these observations as well.

```
. list countryname r if r > 2

     +--------------------------------+
     |       countryname            r |
     |--------------------------------|
121. |        Luxembourg     2.144136 |
129. |           Ireland     2.352905 |
131. |          Botswana     2.034766 |
132. |           Nigeria      2.17953 |
133. | Brunei Darussalam     2.015929 |
     |--------------------------------|
135. |            Kuwait      2.13754 |
136. |            Angola     3.510591 |
     +--------------------------------+

. list countryname r if r < -2

     +-------------------------+
     | countryn~e            r |
     |-------------------------|
  1. | Tajikistan    -3.111456 |
  2. |      China    -2.262674 |
     +-------------------------+
```

Figure 9: Observations with high studentized residuals

4.1.3. Comparison of the econometric model with and without outliers

Next, we created dummies for all our outliers to test the impact of removing all outliers on our regression model coefficients. After creating the dummies, we ran two regression models, one with the outliers and one without outliers. The adjusted R-square improved by +4.02%. Then we used the "suest withoutliers withoutoutliers" simultaneous testing command to compare our stored estimates of regressions. Then we tested each variable using commands similar to "test [withoutliers_mean]infrastructure=[withoutoutliers_mean]infrastructure." For each explanatory variable in our model, we could not reject the hypothesis: "The coefficient value for our variable is the same between two models" (see table 1, page 10). Thus, we conclude that removing outliers will not significantly impact our regression coefficients, even if it improves the line slightly, from 89% to 93%. Therefore, we move forward with all observations in our model to avoid potential bias by eliminating outliers.

Table 1: Coefficient comparison results between models with and without outliers

Explanatory variable	**Coefficient value (all observations, n=136)**	**Coefficient value (without outliers, n=125)**	**Coefficient comparison chi-statistic p-value**
infrastructure	.0232087	.0229222	0.9515
loginstitutions	-.0849506	-.5434772	0.0926
ictadoption	.0189854	.0162235	0.4788

macroeconomicstability	.0065867	.0074444	0.7851
health	.0061384	.0064587	0.8906
skills	.0251572	.0296172	0.3813
productmarket	.0102906	.0144264	0.3820
labormarket	-.0043035	-.0127962	0.1300
logfinancialsystem	.0205116	.0134866	0.9754
logmarketsize	.1135743	-.0038316	0.2742
businessdynamism	-.0157529	-.0104729	0.1457
loginnovationcapability	-.1877204	.0540405	0.2010
Constant	**Coefficient value (all observations)**	**Coefficient value (without outliers)**	**Coefficient comparison chi-statistic p-value**
_cons	5.610141	6.759915	0.1837

5. Comments on Regression

We only have one new comment to add to our first regression estimate comments in section 3.1 (see page 6). Our model does contain outliers. However, they don't have a significant impact on the coefficients of the explanatory variables when removed. Thus, to avoid bias, we will not exclude these observations.

To continue to improve our model and to make it more precise, we'll move forward with heteroskedasticity diagnostics (and adjustment if relevant), examine subsamples through Chow test (and adjust, if needed), and at the end, run a backward regression to clean our model and reduce VIF.

6. Heteroskedasticity: Diagnostics & Correction

Next, we check if our model is heteroskedastic as the linear regression assumption requires residuals to be mostly well-behaved (homoskedastic). After running the regression, commands “rvfplot, yline(0)” and “estat hettest” were run. The first gives a visual idea of the distribution of the residuals around the regression line (see figure 6, page 7). The other, the

Breusch-Pagan test, gives us a chi-statistic and p-value under the hypothesis: “constant variance.” Since our p-value for the chi-statistic is 0.0785 (see figure 10, page 11), we fail to reject the null hypothesis and conclude that our model does not meet the criterion for heteroskedasticity. We do not need to make any adjustments to make it homoskedastic.

Neither we have to adjust for autocorrelation as we are not dealing with time-series data.

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of loggdp

         chi2(1)      =     3.10
         Prob > chi2  =   0.0785
```

Figure 10: Breusch-Pagan test for heteroskedasticity

7. Tests & Variable Selection

7.1. Test for normality of regression

Since our tests usually require normal error terms, the rudimentary tests we ran tested this hypothesis. After re-running the regression, we first used the “predict e, resid” command to create residual values and then looked at the following graphs: “kdensity e, normal,” “qnorm e,” and “pnorm e.” Our data looks close to normal on all graphs, except on the qnorm graph, where our residuals on the far right do not look well behaved. After running the Shapiro-Wilk W test for normal data, we get a p-value output of 0.07981 (see figure 11, page 12). Thus we fail to reject the null hypothesis: "the distribution of the residuals is normal." We conclude that residuals are normally distributed.

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	136	0.98257	1.866	1.406	0.07981

Figure 11: Shapiro-Wilk W test for normal data

7.2. Global F-test

Our global Fisher test is still the same as shared in section 3.1 analysis (see figure 4, page 7). We fail to reject the null hypothesis. At least one of our variables is significant in explaining the variance in our dependent variable.

7.3. Coefficient test

We will save the coefficient testing and variable selection until after we run the Chow test for structural change—as else our constrained model may not provide us with a good understanding of structural differences between developed and developing countries.

7.4. Omitted variable bias test

Using the command ovtest, we checked for the omitted variable bias in our model. The Ramsey RESET test has a p-value of 0.2879 (see figure 12, page 12). Thus, we cannot reject our hypothesis: "model has no omitted variables." So we conclude that our model does not need more variables.

```
Ramsey RESET test using powers of the fitted values of loggdp
       Ho:  model has no omitted variables
                  F(3, 120) =      1.27
                   Prob > F =      0.2879
```

Figure 12: Ramsey RESET test for omitted variable bias

7.5. Model specification test

Using the “linktest” command, we tested for model specificity (see figure 13, page 13). The p-value of _hatsq (0.860) is not significant, and _hat has more explanatory power. Therefore, we cannot reject our hypothesis: "there is no specification error." So we conclude that our model is correctly specified, and we do not need to include or omit variables.

```
      Source |       SS           df       MS      Number of obs   =       136
-------------+----------------------------------   F(2, 133)       =    607.56
       Model |  162.74878          2   81.37439    Prob > F        =    0.0000
    Residual | 17.8135106        133  .13393617    R-squared       =    0.9013
-------------+----------------------------------   Adj R-squared   =    0.8999
       Total | 180.562291        135 1.33749845    Root MSE        =    .36597

------------------------------------------------------------------------------
      loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |    1.08803   .4977826     2.19   0.031      .1034353    2.072625
      _hatsq |  -.0047414   .0267667    -0.18   0.860     -.0576849     .048202
       _cons |  -.4026761   2.289754    -0.18   0.861     -4.931721    4.126369
```

Figure 13: Model specification test

7.6. Test for multicollinearity

The VIF analysis shows (see figure 5, page 7) that 8 of our 12 variables have a VIF score of more than 5. Near multicollinearity is likely an issue affecting our model. We will revisit multicollinearity after the Chow test and variable selection for our final model.

7.7. Select variables

We'll first run the Chow test with all variables to test for structural change and later select final variables. We'll use backward stepwise regression to clean our model.

8. Test for Structural Change

To run the Chow test on our two subsamples: developed and developing economies, we opted for the augmented regression Chow test by estimating one single model using interacted variables. We first used the following taxonomy to classify countries as developed and developing economies using the World Bank's income classification categories for our observations: "Low- and middle-income economies are usually referred to as developing economies, and the Upper Middle Income and the High Income are referred to as Developed Countries" (A4ID, 2021). After creating the developed and developing economy dummy variables, we next created developed economy dummy variables for each explanatory variable using commands such as "generate ddinfrastructure = developed_economy*infrastructure." We then inputted these dummy variables into our regression model and later tested whether the dummies were significantly structurally different (see figure 14, page 14). Our test indicated (fisher test, p-value = 0.0005), that globally, developed and developing economies differ in explaining the variance in GDP per capita through our explanatory variables. Thus, due to this structural change, our final model will be composed of 2 regression models, one for the developed economies and the other for the developing.

```
. test ddinfrastructure ddloginstitutions ddictadoption ddmacroeconomicstab
> ility ddhealth ddskills ddproductmarket ddlabormarket ddlogfinancialsyste
> m ddlogmarketsize ddbusinessdynamism ddloginnovationcapability developed_
> economy

 ( 1)  ddinfrastructure = 0
 ( 2)  ddloginstitutions = 0
 ( 3)  ddictadoption = 0
 ( 4)  ddmacroeconomicstability = 0
 ( 5)  ddhealth = 0
 ( 6)  ddskills = 0
 ( 7)  ddproductmarket = 0
 ( 8)  ddlabormarket = 0
 ( 9)  ddlogfinancialsystem = 0
 (10)  ddlogmarketsize = 0
 (11)  ddbusinessdynamism = 0
 (12)  ddloginnovationcapability = 0
 (13)  developed_economy = 0

       F( 13,   110) =    3.17
            Prob > F =    0.0005
```

Figure 14: Chow test for structural change

9. Presentation of the Final Econometric Model

Please note that our models were not revised after the Chow test. To stick to the requirements of this paper—the instructions and the paper length—we are presenting our final models after merely running the backward stepwise method on our full model for our subsamples. Ideally, we would investigate the new models and subsamples after the Chow test, however, this is out of the scope for this assignment.

9.1. Econometric Model for the Developed Economies

After running the backward stepwise regression on our full model for developed economies (see figure 15, page 15)—to eliminate insignificant variables one by one until those left in our model are significant—our final model for developed economies best-explaining the variance in our dependent variable is as follows:

$$\widehat{loggdp}_{\text{developed}} = 3.085099 + 0.0188293 * infrastructure + 1.195333 * loginstitutions + 0.0139077 * ictadoption$$

```
. stepwise, pr(0.05): reg loggdp infrastructure loginstitutions ictadoption macroeconomicstabi
> lity health skills productmarket labormarket logfinancialsystem logmarketsize businessdynami
> sm loginnovationcapability if developed_economy
                      begin with full model
p = 0.9461 >= 0.0500  removing loginnovationcapability
p = 0.7933 >= 0.0500  removing labormarket
p = 0.6607 >= 0.0500  removing logfinancialsystem
p = 0.5377 >= 0.0500  removing logmarketsize
p = 0.2846 >= 0.0500  removing productmarket
p = 0.1971 >= 0.0500  removing skills
p = 0.0787 >= 0.0500  removing health
p = 0.0748 >= 0.0500  removing macroeconomicstability
p = 0.0885 >= 0.0500  removing businessdynamism

      Source |       SS           df       MS      Number of obs   =        85
-------------+----------------------------------   F(3, 81)        =     72.22
       Model |  23.0291117         3  7.67637057   Prob > F        =    0.0000
    Residual |  8.60910459        81  .106285242   R-squared       =    0.7279
-------------+----------------------------------   Adj R-squared   =    0.7178
       Total |  31.6382163        84  .376645432   Root MSE        =    .32601

--------------------------------------------------------------------------------
         loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
 infrastructure |   .0188293   .0059338     3.17   0.002     .0070228    .0306357
loginstitutions |   1.195333   .3841557     3.11   0.003     .4309832    1.959682
    ictadoption |   .0139077   .0044901     3.10   0.003     .0049737    .0228417
          _cons |   3.085099   1.248562     2.47   0.016     .6008529    5.569346
--------------------------------------------------------------------------------
```

Figure 15: Backward stepwise regression for the developed economies model

Our final model is globally significant (fisher test p-value < 0.05) and it explains 71.78% (adjusted r-squared) of the variance in our dependent variable: 2017 real GDP per capita (PPP). Each of our final variables is significant (t-statistic p-value < 0.05) and positively correlated with GDP. Our model has a mean VIF of 2.94 and VIF is less than 5 for each of our final variables. Thus near multicollinearity is low in our model. Model coefficient estimates and variables are to be interpreted in the following manner for developed economies:

1. **Infrastructure:** One unit of increase in 2017 GCI infrastructure score (scores range: 0-100, continuous), correlates to an **estimated** 1.88% increase in 2017 GDP per capita (PPP) for developed economies in our sample.
2. **ICT Adoption:** One unit of increase in 2017 GCI ICT adoption score (scores range: 0-100, continuous), correlates to an **estimated** 1.39% increase in 2017 GDP per capita (PPP) for developed economies in our sample.
3. **Institutions:** One percent of the increase in the 2017 GCI institutions score, correlates to an **estimated** 1.20% increase in 2017 GDP per capita (PPP) for developed economies in our sample.

9.2. Econometric Model for the Developing Economies

After running the backward stepwise regression on our full model for developing economies (see figure 16, page 17)—to eliminate insignificant variables one by one until those left in our model are significant—our final model for developing economies best-explaining the variance in our dependent variable is as follows:

$$\widehat{loggdp}_{\text{developing}} = 10.4953 + 0.0395495*\textit{infrastructure} - 1.303035*\textit{loginstitutions} + 0.024033*\textit{ictadoption}$$

Our final model is globally significant (fisher test p-value < 0.05) and it explains 71.30% (adjusted r-squared) of the variance in our dependent variable: 2017 real GDP per capita (PPP). Each of our final variables is significant (t-statistic p-value < 0.05). Two of these variables are positively correlated with GDP: infrastructure and ICT adoption. One is negatively correlated: loginstitutions, which is unexpected as one-on-one it is positively correlated. Ideally, this variable and model would be further investigated and tested. However, due to the above-mentioned constraints, we will not be able to refine the predicted model further but we'll offer this brief explanation: high leverage observations, those exceeding the leverage value of .11764706 (see figures 17 and 18, page 17), are likely influencing the model fit. We'll now continue with our final model description.

Our model has a mean VIF of 2.01 and VIF is less than 5 for each of our final variables. Thus near multicollinearity is low in our model. Model coefficient estimates and variables are to be interpreted in the following manner for developing economies:

1. **Infrastructure:** One unit of increase in 2017 GCI infrastructure score (scores range: 0-100, continuous) correlates to an **estimated** 3.95% increase in 2017 GDP per capita (PPP) for developing economies in our sample.
2. **ICT Adoption:** One unit of increase in 2017 GCI ICT adoption score (scores range: 0-100, continuous) correlates to an **estimated** 2.40% increase in 2017 GDP per capita (PPP) for developing economies sample.
3. **Institutions:** One percent of the increase in the 2017 GCI institutions score correlates to an **estimated** 1.30% decrease in 2017 GDP per capita (PPP) for developing economies in our sample.

```
. stepwise, pr(0.05): reg loggdp infrastructure loginstitutions ictadoption macroeconomicstabi
> lity health skills productmarket labormarket logfinancialsystem logmarketsize businessdynami
> sm loginnovationcapability if developing_economy
                      begin with full model
p = 0.9868 >= 0.0500  removing health
p = 0.6362 >= 0.0500  removing productmarket
p = 0.5103 >= 0.0500  removing businessdynamism
p = 0.4598 >= 0.0500  removing labormarket
p = 0.3946 >= 0.0500  removing macroeconomicstability
p = 0.4213 >= 0.0500  removing logfinancialsystem
p = 0.2628 >= 0.0500  removing skills
p = 0.1292 >= 0.0500  removing loginnovationcapability
p = 0.2704 >= 0.0500  removing logmarketsize

      Source |       SS           df       MS      Number of obs   =        51
-------------+----------------------------------   F(3, 47)        =     42.40
       Model |  18.1066048         3  6.03553493   Prob > F        =    0.0000
    Residual |  6.69063994        47  .142354041   R-squared       =    0.7302
-------------+----------------------------------   Adj R-squared   =    0.7130
       Total |  24.7972447        50  .495944894   Root MSE        =     .3773

---------------------------------------------------------------------------------
         loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
 infrastructure |   .0395495   .0074489     5.31   0.000     .0245643    .0545347
loginstitutions |  -1.303035   .4878598    -2.67   0.010    -2.284482   -.3215875
    ictadoption |    .024033    .006076     3.96   0.000     .0118097    .0362563
          _cons |    10.4953   1.708666     6.14   0.000      7.05791     13.9327
---------------------------------------------------------------------------------
```

Figure 16: Backward stepwise regression for the developing economies model

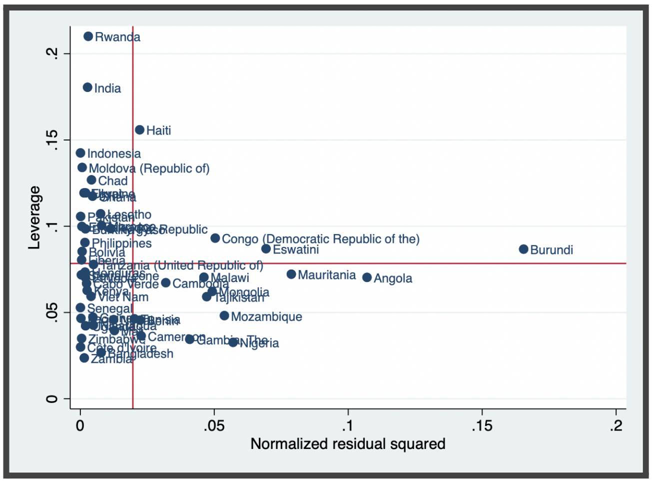

Figure 17: Leverage-versus-squared-residual plot test, model: developing economies

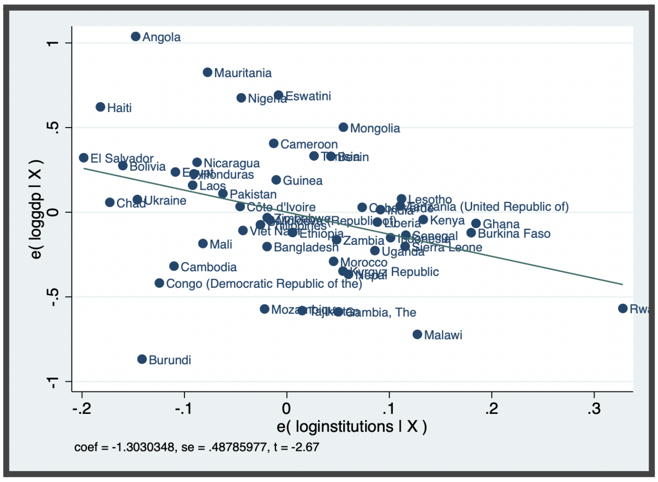

Figure 18: Avplot for the loginstitutions variable, model: developing economies

10. Conclusions

The purpose of our econometrics project was to understand which 2018 GCI productivity drivers best explain the variance in 2017 real GDP per capita (PPP) of national economies (if any). Our findings reveal the following to be the most significant competitiveness variables in explaining differences across both the developed and developing economies: infrastructure, ICT adoption, and institutions.

Infrastructure score is the most important and influential predictor of GDP per capita in both our models, estimated to bring a benefit of 3.95% economic output per person in developed economies and 1.88% in developing economies with 1 unit of increase in score. Therefore, building, improving, and upkeeping transport (road, rail, water, air) and utility (power and water) infrastructure should be a vital consideration for each economy seeking to grow. New transport infrastructure can open new trade routes, improve regional and global connectivity, help facilitate the movement of goods and people, and bring new business to the nation. Reliable utility infrastructure helps economic agents produce economic output efficiently. Upkeep and improvement of the infrastructure lends to the same purpose.

Another vital competitiveness factor is ICT adoption. Our models estimate that 1 unit of increase in ICT adoption score lends to an increase of 2.40% economic output per worker in developing economies and 1.39% in developed economies. ICTs reduce transaction costs, improve efficiency, speed up exchange of information and ideas, and are catalysts for sparking innovation. Therefore, investments in ICT infrastructure and efficient and effective technologies increasing economic output should be a routine consideration and part of any national development plan.

Institutions score is another significant predictor of GDP. For developed economies, 1% of increase in score lends to an estimated 1.20% increase in GDP per capita per our model. Thus, strong public-sector, legal system, social capital, and corporate governance institutions facilitating economic activity are a boon to the output per worker in developed economies. Transparency, ethics, checks and balances, security, and rights reduce uncertainties and risks for participating economic agents. Reforming or continuously improving national institutions to build and upkeep strong foundations is a worthwhile routine endeavor for developed economies to engage in.

Finally, our paper should be viewed as incomplete work. It has limitations, mentioned therein. We invite readers to read current literature on the topic and the latest GCI report.

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).

Appendix I: 2018 Global Competitiveness Index Pillars: Competitiveness Factor Importance and Score Composition

The content in the table below is republished from the 2018 Global Competitiveness Index report (Schwab, 2018) to provide researchers' rationale for the importance of each competitiveness factor assessed and its score composition.

Competitiveness Factor	What does it capture?	Why does it matter?
Institutions	Security, property rights, social capital, checks and balances, transparency and ethics, public-sector performance and corporate governance.	By establishing constraints, both legal (laws and enforcement mechanisms) and informal (norms of behaviors), institutions determine the context in which individuals organize themselves and their economic activity. Institutions impact productivity, mainly through providing incentives and reducing uncertainties.
Infrastructure	The quality and extension of transport infrastructure (road, rail, water and air) and utility infrastructure.	Better-connected geographic areas have generally been more prosperous. Well-developed infrastructure lowers transportation and transaction costs, and facilitates the movement of goods and people and the transfer of information within a country and across borders. It also ensures access to power and water—both necessary conditions for modern economic activity.
ICT adoption	The degree of diffusion of specific information and communication technologies (ICTs).	ICTs reduce transaction costs and speed up information and idea exchange, improving efficiency and sparking innovation. As ICTs are general purpose technologies increasingly embedded in the structure of the economy, they are becoming as necessary as power and transport infrastructure for all economies.
Macroeconomic stability	The level of inflation and the sustainability of fiscal policy.	Moderate and predictable inflation and sustainable public budgets reduce uncertainties, set returns expectations for investments and increase business confidence—all of which boost productivity. Also, in an increasingly interconnected world where capital can move quickly, loss of confidence in macroeconomic stability can trigger capital flight, with destabilizing economic effects.
Health	Health-adjusted life expectancy (HALE)—the average number of years a newborn can expect to live in good health.	Healthier individuals have more physical and mental capabilities, are more productive and creative, and tend to invest more in education as life expectancy increases. Healthier children develop into adults with stronger cognitive abilities.
		(Continued…)

Competitiveness Factor	What does it capture?	Why does it matter?
Skills	The general level of skills of the workforce and the quantity and quality of education. While the concept of educational quality is constantly evolving, important quality factors today include: developing digital literacy, interpersonal skills, and the ability to think critically and creatively.	Education embeds skills and competencies in the labour force. Highlyeducated populations are more productive because they possess greater collective ability to perform tasks and transfer knowledge quickly, and create new knowledge and applications.
Product market	The extent to which a country provides an even playing field for companies to participate in its markets. It is measured in terms of extent of market power, openness to foreign firms and the degree of market distortions.	Competition supports productivity gains by incentivizing companies to innovate; update their products, services and organization; and supply the best possible products at the fairest price.
Labour market	It encompasses "flexibility", namely, the extent to which human resources can be reorganized and "talent management", namely, the extent to which human resources are leveraged.	Well-functioning labour markets foster productivity by matching workers with the most suitable jobs for their skillset and developing talent to reach their full potential. By combining flexibility with protection of workers' basic rights, well-functioning labour markets allow countries to be more resilient to shocks and re-allocate production to emerging segments; incentivize workers to take risks; attract and retain talent; and motivate workers.
Financial system	The depth, namely the availability of credit, equity, debt, insurance and other financial products, and the stability, namely, the mitigation of excessive risk-taking and opportunistic behavior of the financial system.	A developed financial sector fosters productivity in mainly three ways: pooling savings into productive investments; improving the allocation of capital to the most promising investments through monitoring borrowers, reducing information asymmetries; and providing an efficient payment system. At the same time, appropriate regulation of financial institutions is needed to avoid financial crises that may cause long-lasting negative effects on investments and productivity.
Market size	The size of the domestic and foreign markets to which a country's firms have access. It is proxied by the sum of the value of consumption, investment and exports.	Larger markets lift productivity through economies of scale: the unit cost of production tends to decrease with the amount of output produced. Large markets also incentivize innovation. As ideas are non-rival, more potential users means greater potential returns on a new idea. Moreover, large markets create positive externalities as accumulation of human capital and transmission of knowledge increase the returns to scale embedded in the creation of technology or knowledge.
		(Continued…)

Competitiveness Factor	What does it capture?	Why does it matter?
Business dynamism	The private sector's capacity to generate and adopt new technologies and new ways to organize work, through a culture that embraces change, risk, new business models, and administrative rules that allow firms to enter and exit the market easily.	An agile and dynamic private sector increases productivity by taking business risks, testing new ideas and creating innovative products and services. In an environment characterized by frequent disruption and redefinition of businesses and sectors, successful economic systems are resilient to technological shocks and are able to constantly re-invent themselves.
Innovation capability	The quantity and quality of formal research and development; the extent to which a country's environment encourages collaboration, connectivity, creativity, diversity and confrontation across different visions and angles; and the capacity to turn ideas into new goods and services.	Countries that can generate greater knowledge accumulation and that offer better collaborative or interdisciplinary opportunities tend to have more capacity to generate innovative ideas and new business models, which are widely considered the engines of economic growth.

Appendix II: Project Stata Files

Project data files can be accessed the following links:

1. Project dataset (format .dta; Google Drive link)
2. Project do file (format .do; Google Drive link)