Национальный исследовательский университет

«Высшая школа экономики»

Факультет мировой экономики и мировой политики

Образовательная программа «Мировая экономика»

Отчет о самостоятельной работе

по дисциплине «Эконометрика» на тему

«A study on the dependence of GDP on various variables»

Выполнен:

Нгуен Тхюи Зыонг, БМЭ204

Фам Тху Чанг, БМЭ204

Москва 2023

Table of Contents

Introduction

Our study concentrates on how GDP is determined by various economic factors in different countries through time.

Every year, international organizations compile data on each country's GDP and various components that contribute to GDP growth in order to learn the how the variables interact with each other and to compare how economic growth between countries across different continents over time.

It is common knowledge that each country wants to become richer and stronger: developing countries try to become developed as soon as possible while developed countries do not want to be surpassed or by the others. Hence it is also important to acknowledge which economic factors are contributing to GDP and if they hold the same effectiveness across different countries and over time.

In this research, we aim to study how specific variables can affect GDP of 83 countries around the world based on recorded annually recorded data from 2000 to 2019. In a bigger picture, we aim to derive from our study useful conclusions for every country to prepare appropriate policies to reach their GDP target.

In order to achieve the goal, in our study, we set to accomplish the following objectives:

- analyze the chosen variables of the panel data;
- formulate hypotheses and testify them based on different types of regression model;
- choose an optimal model that best estimates the real values of GDP

The main body of our study includes 4 parts which corresponding majorly to the objectives above: part 1 describes the data source and main hypotheses; part 2 describes the data in general; part 3 assesses whether the main hypotheses are confirmed based on the preliminary analysis; part 4 focuses on building models (pooled, fixed effects and random effects); and part 5 testifies hypotheses based on the optimal model.

Part 1: Data description

Source of dataset

"Penn World Table" from University of Groningen, Groningen growth and development centre. URL: https://www.rug.nl/ggdc/productivity/pwt/?lang=en

The data is taken yearly in the span of 20 years from 2000 to 2019 in 83 countries. In particular, for each year from 2000 to 2019, there are 83 observations representing 83 countries. Hence, in total, there are 20 * 83 = 1660 observations in the panel data.

No.	Object's characteristic	Name of variable	Role	Type of data
1	GDP *(million US dollars)*	GDP	dependent	numerical
2	Domestic consumption *(million US dollars)*	consumption	independent	numerical
3	Domestic consumption plus investment *(million US dollars)*	dom_abs	independent	numerical
4	Welfare-relevant TFP 1 – low TFP_w (< 0.75) 2 – medium TFP_w (≥ 0.75, < 1.25) 3 – high TFP_w (≥ 1.25)	TFP_w	independent	categorical

- GDP is the market value of all final goods and services produced in a country in a year and is one of the main indicators reflecting the economic development of a country.
- Domestic consumption is typically defined as the sum of household consumption, gross investment, and government consumption.
- Domestic absorption is the sum of domestic consumption and gross investment (by households and government) within a country.
- TFP (total factor productivity) measures the productive efficiency by determining how much output (or GDP) a country can produce from a certain amount of input. The total factor productivity is used to identify the performance level of a business and its efficiency. It tries to reach a definite value of how well the inputs have been transferred to the output. The welfare factor is included because TFP contributes to the welfare of a nation, but not all nations have the same level of welfare.

Main hypotheses

In the framework of our study, we will take into consideration these following hypotheses:

1. Welfare-relevant TFP has a positive impact on GDP.
2. The dependence of GDP on household consumption is stronger for countries Asia-Pacific region than those in other continents.

   Asia-Pacific is the most dynamic and fast-paced economic environment in the world. Therefore, under the effect of high economic growth, people in this region are spending more, which means household consumption has more impact in this region than elsewhere (people's incentive to spend)
3. Consumption and investment have a positive impact on GDP, and since consumption has more effects on GDP than investment does, negative deference of GDP over consumption (over-consumption) creates more damage to GDP than negative investment does.

   Overconsumption occurs in two situations. One, the government might have spent too much on social spending such as for pensions, welfares, infrastructures, … and this creates budget deficit. Two, households might have over-spent on products in an ineffective way, which leads to the decrease in the marginal utility of consumption. Negative investment means that invested projects are not producing profits and results in loss, indicating that an economy is doing badly.

   We want to test if the damage of overconsumption is greater than that of investment, to see if GDP is more easily hurt in the short run (with overconsumption) or in the long run (with negative investment).

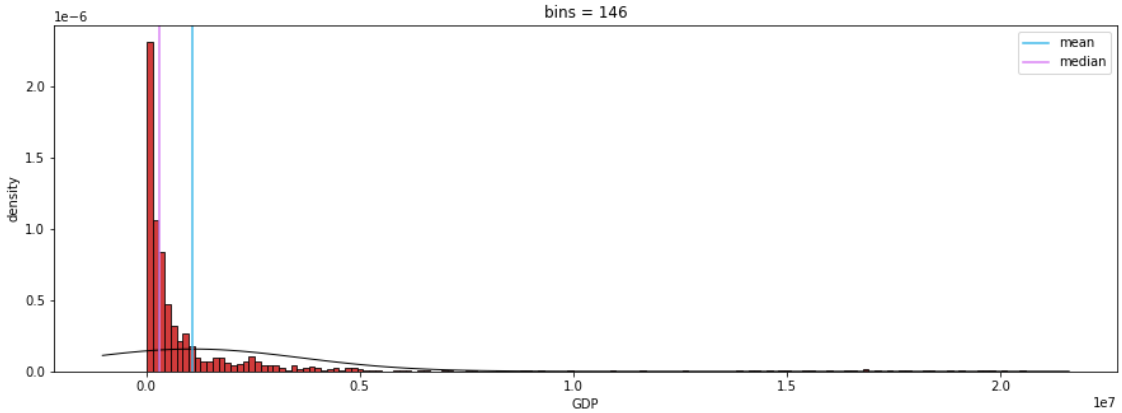
Part 2: Preliminary analysis

Descriptive statistics

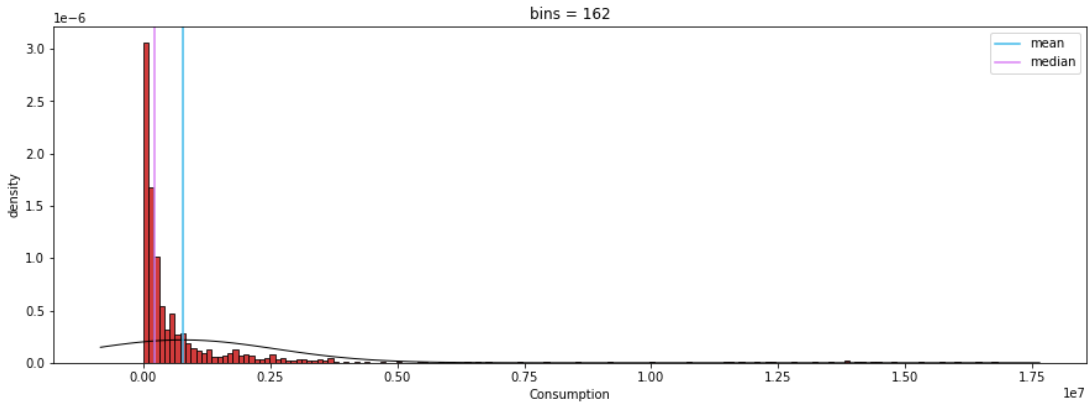
	GDP	consumption	dom_abs	TFP_w
Min	6 670.96	2 655.92	3 416.25	0.09
Max	20 572 606	16 803 152	21 638 008	2.35
Median	308 181.03	219 920.29	303 310.22	0.98
Mean	1 079 407.29	776 253.09	1 054 762.72	0.98
Standard deviation	2 547 695.58	1 827 208.19	2 512 937.29	0.17

Table 1: Descriptive statistics of all variables

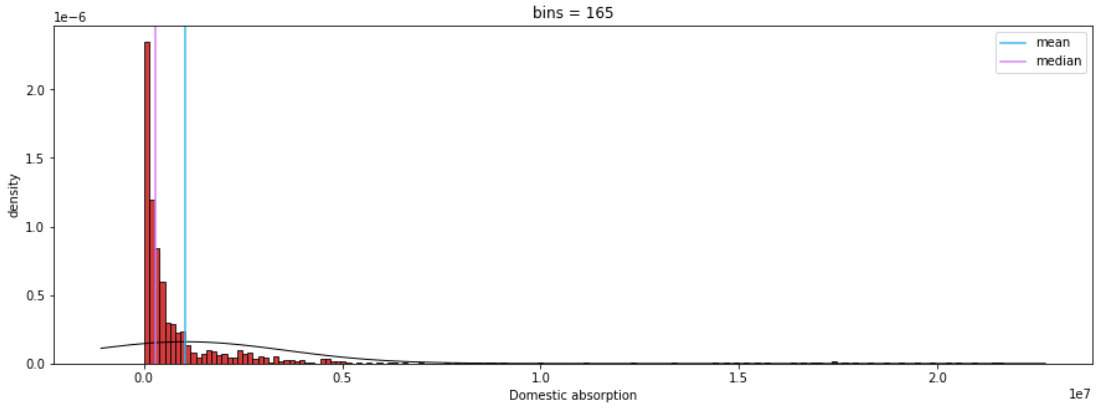
From the graphs, we can see that the distribution of GDP, consumption and domestic absorption is skewed to the right (average > median) and all graphs are quite heavily skewed. We will not go into more details and just simply want to explain that because of this, we will normalized these numeric variables using logarithm.



Graph 1: Histogram of GDP (Freedman – Diaconis's rule)

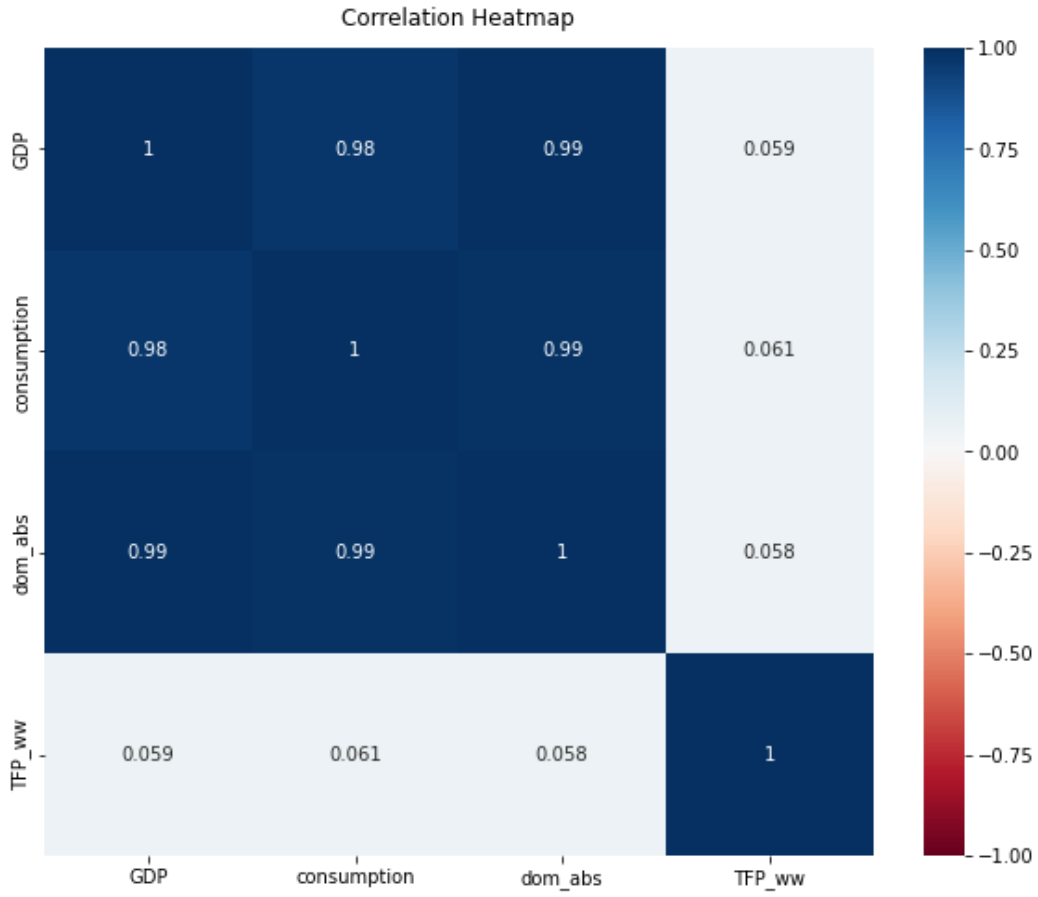


Graph 2: Histogram of *consumption* (Freedman – Diaconis's rule)



Graph 3: Histogram of *domestic absorption* (Freedman – Diaconis's rule)

Pearson's correlation coefficients



Graph 4: Pearson correlation for all variables

Among variables measured on a ratio scale, consumption and domestic absorption show a strong positive correlation with each other and with GDP.

This can be explained by the importance of spending and investment in GDP. Spending increases aggregate demand by generating more profits and incentives for businesses, which pay taxes to the state, which in turn creates more spending in different ways. Investment increases the capital of the business, which has a positive effect not only in the short term by increasing revenue,

but also in the long term through new research, new equipment, ... and hence increasing productivity, which directly affects GDP.

Kruskal-Wallis test

H_0: GDP does not depend on welfare-relevant TFP, i.e. medians are equal

H_1: Price does depend on welfare-relevant TFP

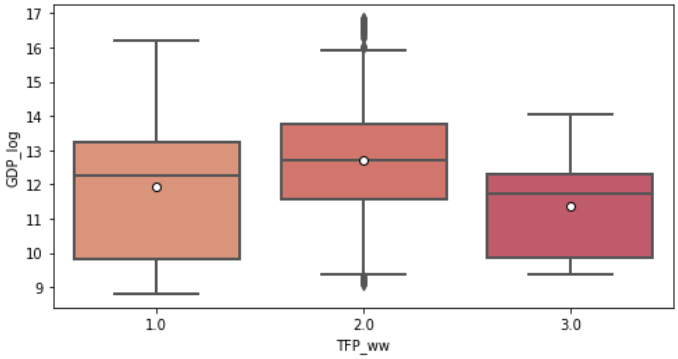
Value of Kruskal-Wallis test	426.389
p-value	0

Based on the received result, we conclude that hypothesis H_0 is rejected. There are significant differences in statistics between GDP depending on the level of welfare-relevant TFP. This means different levels of TFP affect differently on GDP.

Part 3: Preliminary testing of hypotheses

01 Welfare-relevant TFP has a positive impact on GDP. **rejected**

To check this hypothesis, we built a box plot of GDP and welfare-relevant TFP.



Graph 5: Box plot of the GDP – welfare-relevant TFP pair

It can be observed that the average GDP of countries with low and medium TFP is higher than average GDP of those with high TFP. This hypothesis is not approved, as the box plot does not show a positive relationship between them.

02 The dependence of GDP on household consumption is stronger for countries Asia-Pacific region than those in Europe.

rejected

The dependence of GDP on household consumption is approved by our preliminary analysis, as the correlation between them is strong.

In order to check the second condition of the hypothesis, we divided the countries into APAC and non-APAC countries. We then calculated the correlation of GDP and consumption in each group of countries and get the following results.

The correlation coefficient between GDP and consumption for countries in the APAC region is less the same coefficient for countries not in the APAC region (*0.976 < 0.999*)

Therefore, in this part, we reject hypothesis 02.

03 Consumption and investment have a positive impact on GDP, and since consumption has more effects on GDP than investment does, negative deference of GDP over consumption (over-consumption) creates more damage to GDP than negative investment does.

rejected

First, Pearson coefficients of correlation for consumption and domestic absorption are both positive.

Second, the share of consumption in GDP is generally greater than the share of investment in GDP (with larger range of values, higher median and average, higher minimum and maximum values), indicating higher contribution to GDP.

However, based on the preliminary statistics, there is not enough evidence to proof the last statement, which also carries the main meaning of the hypothesis. Hence, we temporarily reject hypothesis 03.

Variables	Min	Max	Mean	Median
Consumption	15.18	138.76	77.34	77.79
Investment	-31.73	65.98	23.68	24.06

Part 4: Assessment of models

As mentioned above, 3 numerical variables in our dataset, both dependent and independent variables, are right-skewed. Therefore, in this part, we used logarithm in order to normalize the variables.

1. Pooled model

Variable	Coefficients estimators	Standard error	p-value
const	0.124	0.035	0.000
dom_abs_log	1.319	0.025	0.000
consumption_log	-0.309	0.025	0.000
TFP_ww	0.172	0.012	0.000

R-squared = 0.99316 Adj R-squared = 0.99315 AIC = -1936.97

Table 2: Result of pooled-effects model

Formula

As such, the pooled model takes the following form:

GDP_log = 0.124 + 1.319 * dom_abs_log – 0.309 * consumption_log + 0.172 * TFP_ww + ε_{it},
where ε_{it} = errors corresponding to the t (2000 – 2019) observations for country i (1 – 83).

Significance: All coefficients are above significant level because their p-values are lower than 0.05, which means we can include them in the model.

Functions of variables upon their entry into model

- GDP increases by 0.124% when all independent variables equal zero.
- An increase of 1.319% in domestic absorption is associated with an increase of 1% in GDP
- A increase of 0.309% in household consumption is associated with a decrease of 1% in GDP
- An increase of 0.172% in the level of welfare-relevant TFP is associated with an increase of 1% in GDP

Measure of goodness-of-fit

Adjested R-squared = 0.9932

99.32% of the variability observed in the target variable (GDP) is explained by the pooled regression model.

2. Fixed model (within estimators)

Variable	Coefficients estimators	Standard error	p-value
dom_abs_log	0.602	0.031	0.000
consumption_log	0.264	0.033	0.000
TFP_ww	0.05	0.008	0.000

R-squared = 0.92578 Adj R-squared = 0.92177 AIC = -4211.466

Table 3: Result of fixed-effects model

Formula

As such, the fixed model takes the following form:

$$GDP\_log = 0.602 * dom\_abs\_log + 0.264 * consumption\_log + 0.05 * TFP\_ww + \varepsilon_{it},$$

where ε_{it} = errors corresponding to the t (2000 – 2019) observations for country i (1 – 83).

Significance: All coefficients are above significant level because their p-values are lower than 0.05, which means we can include them in the model.

Functions of variables upon their entry into model

- An increase of 0.602% in domestic absorption is associated with an increase of 1% in GDP
- An increase of 0.264% in household consumption is associated with an increase of 1% in GDP
- An increase of 0.05% in the level of welfare-relevant TFP is associated with an increase of 1% in GDP

Note: Since our studies take into consideration 83 countries, we decided not to use dummy variables.

Measure of goodness-of-fit

Adjested R-squared = 0.9217

92.17% of the variability observed in the target variable (GDP) is explained by the fixed regression model.

3. F-test

H_0: ai = 0 (no individual specific effects)

H_1: ai # 0 (individual specific effects are significant)

F-test for individual effects	
F-statistics	64.203
Degree of freedom 1	82
Degree of freedom 2	1574
p-value	0.000

Table 4: Result of F-test to compare between pooled model and fixed model

p-value < 0.05 => H0 is rejected. The temporal structure of the data should be taken into account, therefore, the fixed effect model should be chosen.

There are about 62.65% of total individual effects that turned out to be significant with p-values lower than 0.05 *(52 significant out of 83 in total).*

4. Random effect model

Variable	Coefficients estimators	Standard error	p-value
const	1.05	0.066	0.000
dom_abs_log	0.656	0.032	0.000
consumption_log	0.28	0.033	0.000
TFP_ww	0.083	0.008	0.000

R-squared = 0.95165 Adj R-squared = 0.95156 AIC =

Table 5: Result of random-effect variables

Formula

As such, the pooled model takes the following form:

$$GDP\_log = 1.05 + 0.656 * dom\_abs\_log + 0.28 * consumption\_log + 0.083 * TFP\_ww + \varepsilon_{it},$$

where ε_{it} = errors corresponding to the t (2000 – 2019) observations for country i (1 – 83).

p-values

All coefficients are above significant level because their p-values are lower than 0.05, which means we can include them in the model.

Functions of variables upon their entry into model

- GDP increases by 1.05% when all independent variables equal zero.
- An increase of 0.656% in domestic absorption is associated with an increase of 1% in GDP
- An increase of 0.28% in household consumption is associated with an increase of 1% in GDP
- An increase of 0.083% in the level of welfare-relevant TFP is associated with an increase of 1% in GDP

Measure of goodness-of-fit

Adjested R-squared = 0.9516

95.16% of the variability observed in the target variable (GDP) is explained by the random regression model.

5. Hausman test

H_0: No correlation between individual effects and independent variables.

H_1: Correlation between individual effects and independent variables exists.

Hausman test	
χ^2 statistics	149.6
Degree of freedom	3
p-value	0.000

Table 6: Result of Hausman test to compare fixed-effects model and random-effects model

p-value < 0.05 => H0 is rejected. The individual effects are correlated with independent variables.

Since it turns out that the fixed effect estimator is the only consistent one between the two estimators, the fixed model should be chosen over the random model.

Part 5: Model selection and hypothesis testing

Model selection

F-test result: fixed-effects model is preferred

Hausman result: random-effects model is preferred

Model	R-squared	Adjusted R-squared	AIC
Pool model	0.993	0.993	-1936.97
Fixed-effects model	0.926	0.922	-4211.47
Random-effects model	0.951	0.952	-2849.23

We decided to choose the fixed-effects model because except for R-squared parameter, the rest suggest us that this model is the optimal one. As for R-square and adjusted R-squared values, we decided that if our model can predict 92.6% of the variability in GDP, it is also a good enough option in exchange of low AIC.

The meaning of choosing the fixed-effects model over the others is that:

- The structure of panel data (including temporal and cross-dimensional characters), because the pooled model was proven to be less effective.
- Dependencies are observed within individuals, individual effects of unobserved and independent variables are constant (fixed) over time, heterogeneity can exist within the model, because the random model was proven to be less effective.

For our panel dataset, however, we expected the optimal model should be the random-effects model, because in the span of 20 years, with the advancement of technology as well as the increasing level of globalization, dependency of GDP on various variables should be different over time and across different continents, meaning the model should represent both within and between individual estimators.

However, we came the conclusion that it's because our choice of independent variables are very typical drivers of GDP growth (consumption, domestic absorption and total factor productivity), it is understandable that the effects do not differ much between individuals, but rather over time because of the development of globalization, digitalization and technology. Hence, this model is still suitable for this specific studying and can be used to test the formulated hypotheses.

Hypotheses testing

Hypothesis 01: Welfare-relevant TFP has a positive impact on GDP. **confirmed**

The coefficient reflecting GDP's dependence of TFP is positive, so we can confirm hypothesis 01 almost immediately.

Hypothesis 02: The dependence of GDP on household consumption is stronger for countries Asia-Pacific region than those in Europe. **confirmed**

```
Call:
plm(formula = GDP_log ~ consumption_log, data = pan_hypo1_noapac,
    effect = "individual", model = "within", index = c("country",
        "year"))

Balanced Panel: n = 69, T = 20, N = 1380

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.3073467 -0.0291174  0.0011773  0.0287001  1.3751698

Coefficients:
                Estimate Std. Error t-value  Pr(>|t|)
consumption_log 0.8575076  0.0081548  105.15 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    70.269
Residual Sum of Squares: 7.4431
R-Squared:      0.89408
Adj. R-Squared: 0.8885
F-statistic: 11057.4 on 1 and 1310 DF, p-value: < 2.22e-16
```

Table 7: Result of fixed-effects model for APAC countries

```
Call:
plm(formula = GDP_log ~ consumption_log, data = pan_hypo1_apac,
    effect = "individual", model = "within", index = c("country",
        "year"))

Balanced Panel: n = 14, T = 20, N = 280

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.2152949 -0.0250934  0.0012859  0.0347296  0.1689923

Coefficients:
                Estimate Std. Error t-value  Pr(>|t|)
consumption_log 0.951983   0.012519  76.045 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    23.008
Residual Sum of Squares: 1.0081
R-Squared:      0.95618
Adj. R-Squared: 0.95387
F-statistic: 5782.9 on 1 and 265 DF, p-value: < 2.22e-16
```

Table 8: Result of fixed-effects model for non-APAC countries

We tested this hypothesis by diving our observations into two parts: the first parts includes only 14 countries that are situated in the Asia-Pacific region, the second part includes the remaining 69 countries.

Next, our hypothesis is based on the assumption that people in the Asia-Pacific region are more incentivized to spend because they are growing through a stage of fast development with higher economic growth than other parts of the world. Because we assume that these incentives are correlated with consumption, we consider them individual specific effects that were not observed or measured; and since these individual effects have a correlation with our independent variables, it is more suitable to choose a fixed model rather than a random model.

In order to optimize the model, we removed all irrelevant variables and left only the *consumption* variable. We ran separate models for the two parts of the dataset. From the obtained results, we can see that the coefficient representing consumption's impact on GDP in greater in Asia-Pacific region than elsewhere. *(coefficient APAC: 0.952 > coefficient non-APAC: 0.857)*. Hence, we have grounds to confirm hypothesis 02.

Result: Hypothesis 02 is confirmed.

Hypothesis 03: Consumption and investment have a positive impact on GDP, and since consumption has more effects on GDP than investment does, negative deference of GDP over consumption (over-consumption) creates more damage to GDP than negative investment does. **confirmed**

Again, to test this hypothesis we used the fixed-effects model, which means that we take into account the temporal characteristic of panel data. We double checked by running the three types of model with the same set of regressors and comparing them with each other using the F-test and the Hausman. The result also indicated that it is optimal to use the fixed-effects model.

In order to test this hypothesis, we included two original variables and created three more variables based on the original:

- GDP
- consumption
- investment (the difference between domestic absorption and consumption)

- negative investment index: if investment is negative, the index takes the 1 value; if not, it takes the zero value.
- over-consumption index: if government is smaller than consumption, the index takes the 1 value; if not, it takes the zero value.

We break down the hypothesis into three parts.

Firstly, we need to test if consumption and investment have positive effects on GDP. We can confirm this statement easily as the coefficients for these two variables in all three models are both positive and significant.

Secondly, we need to see if our assumption that consumption (while being positive) has a greater effect on GDP than (positive) investment does. Again, we can confirm this statement easily because the coefficient representing the dependence of GDP on investment is very modest compared to the coefficient of consumption.

Thirdly, based on the result of the adjusted fixed-effects model, we conclude that negative investment in an economy has fewer devastating effects on GDP than negative consumption does. At the same time, the coefficient representing the effect of negative investment on GDP is not statistically significant. Hence, we confirm hypothesis 03 that negative consumption causes larger damage to GDP than negative investment.

Variable	**Coefficients estimators**	**Standard error**	**p-value**
consumption	0.877	0.007	0.000
investment	0	0	0.874
gr_inv_less_0	-0.042	0.027	0.124
cons_over_gov_negative	-0.17	0.013	0.000

R-squared = 0.91717 Adj R-squared = 0.91264 AIC = -1936.97

Table 9: Result of fixed-effects model