ECON 302 PROJECT

CHINA’S GROWTH BY INVESTMENT AND INNOVATION

A comparative analysis of China with U.S. to explain its growth by innovation level measured by “Patent Fillings” and general investment as “Gross Capital Formation”.

BUĞRA OĞUZ

UYGAR ÖZDEMİR

ABSTRACT

This paper examines the relationship between investment, innovation, and GDP growth of China by comparing the economies of China and the United States. Using patent filings as a proxy for innovation and gross capital formation as a measure of investment, the study investigates how these factors contribute to economic growth in China’s rapid growing economy. Using World Bank data from 1985 to 2021, the research provides a comparative analysis of growth trends and a time-series examination of China’s rapid development. The findings focus on econometric results and show that a nuanced examination has to be done before examining China’s economic growth by these variables. This study offers insights for policymakers for innovation-driven growth with sustainable economic practices.

OUTLINE

1. INTRODUCTION

Economic growth has been a fundamental objective for nations seeking to improve the quality of life for their citizens. Among the many factors driving economic growth, investment and innovation hold an important role. Investment, represented by gross capital formation, provides the necessary physical infrastructure and resources for production, while innovation stimulates technological advancement, improves productivity, and paves the way for new industries and markets. These factors has been extensively studied to understand their impact on GDP growth across different economic contexts.

This paper aims to explore the relationship between investment, innovation, and GDP growth by focusing on the economies of China and the United States, but especially on China. These two nations, while both major global economic powers, represent distinct economic growth models. China's growth has predominantly been driven by state-led strategies, massive infrastructure investments, and a gradual shift toward technological innovation. Conversely, the U.S. economy's growth is rooted in a free-market framework, characterized by private-sector-led innovation, strong institutional frameworks, and a focus on knowledge-based industries.

The research leverages patent fillings as a proxy for innovation. Patents serve as a critical metric to evaluate the innovative capacity of an economy. By comparing patent trends and gross capital formation in China and the U.S., this paper seeks to provide insights into how these factors influence economic growth in China.

The analysis draws on World Bank data from 1985 to 2021 to conduct a comparative study of GDP growth in China and the U.S. Additionally, a detailed time-series analysis of China's economic growth will be undertaken to understand the specific dynamics driving its rapid development. Ultimately, the findings aim to shed light on the strategies that can foster sustainable and inclusive economic growth in both developed and emerging economies.

2. LITERATURE REVIEW

The main purpose of our research is to measure the impact of innovation and gross capital formation on GDP growth. What we will specifically examine here is the number of patents for innovation. We will use the number of patents as a proxy variable to measure innovation. The two countries we use and compare are the US and China. But the specific econometic analysis will be done on only China.

GDP is the total economic value of final goods and services produced within a country's borders in a year. The main reason for GDP growth is the increase in the amount of goods and services produced in that country. Countries also try to increase this amount. Two possible ways to increase this amount is to increase productivity and technology.

We would like to point out the following statement by Robert Solow, whom we benefited from a lot in Growth Economics course in METU on this subject (1956): “A large part of economic growth in developed countries can be attributed to technological progress rather than the accumulation of capital or labor.” With this statement, Solow indicates that economic growth will be possible not with capital increase or population increase but with technological progress. Real growth is possible by increasing production per capital or per worker.

Another Economist we see on Growth Economics, Paul Romer (1990), says: “The most important ideas are not only free; they are inexhaustible. They lead to long-term economic growth.” Here, Romer emphasizes that technology is inexhaustible. Capital has a depreciation rate and economic life. Labor also has an end. Human life is not unlimited. However, ideas, technologies, inventions are immortal. They can live on for generations with different machines. This reveals how important technological progress is for economic growth.

Another topic we will focus on in this project is Innovation. Innovation is the adaptation and application of new creative ideas or inventions to economic areas. Technological development and increased productivity, which are the most important sources of economic growth, are also possible with innovation. Innovation's most important feature is that it increases productivity. With increased productivity, the economy can produce more even if the amount of capital or labor remains the same. Innovation causes new goods and services to emerge. This means employment in that new sector. The economy grows by producing goods and services that it has not produced before. Knowledge accumulation is another important factor. The knowledge accumulation that is formed creates an important infrastructure for innovations that will occur in the future.

The owner of the theory of Creative Destruction, Austrian Economist and Political Scientist Joseph Schumpeter said the following about innovation (1939): "Innovation is the outstanding fact in the economic history of capitalist society or in what is purely economic in that history." As can be understood from this saying, innovation is economic history and even all of history itself.

This year's Nobel Prize in Economics winner Turkish-American Economist Daron Acemoğlu also says the following about Innovation (2012): "Technological innovation is critical for sustained economic growth, but it requires strong institutions that incentivize and protect new ideas." Here Acemoğlu emphasizes the importance of innovation, but he also talks about the things that are necessary for innovation to occur. Encouraging innovation, and more importantly, protecting it, is at least as important as the innovation itself. Here, the task falls on the state and the consciousness of the people. Societies that manage to establish the necessary infrastructure for these innovations manage to increase their production and economic welfare. History has shown us this.

Using patent filings is adequate to mesure innovation as Edith Penrose (1951) said, "The patent system serves as a reward mechanism for innovation and plays a central role in facilitating the commercialization of new technologies." Patents have been a source of motivation for inventors.

Another element of economic growth is Gross Capital Formation. Gross Capital Formation refers to the net increase in the total physical capital stock produced in an economy in a given period. This includes investments made to create machinery, buildings, infrastructure and other means of production. If a country makes large investments in infrastructure projects such as roads, power plants, factory buildings, this increases Gross Capital Formation. While Gross Capital Formation contributes greatly to economic growth in countries such as China, Gross Capital Formation rates are relatively lower in some developed countries because these countries focus more on the consumption and service sectors.

In this article, we will use the US and Chinese economies for comparison. The Chinese economy has become one of the fastest growing economies in the world in the past few decades. The reform and opening-up policies initiated under the leadership of Deng Xiaoping in 1978 have laid the foundation for this growth. The transition from central planning to a market-oriented economic structure has increased efficiency and opened up a large space for the private sector. With its export-oriented industrialization strategy, China has assumed a critical role in the global supply chain and positioned itself as a low-cost manufacturing center. High levels of infrastructure investment contributed to its growth as much as a large population advantage, and rapidly increasing urbanization. Furthermore, with the emphasis on R&D investment and technological innovation, China has evolved from an economy based solely on cheap labor to one that exports high-tech products.

The role of innovation in the recent growth of the Chinese economy is undoubtedly huge even if it was not as much in the beginning of China’s economic turnover. Justin Yifu Lin (2012) “China’s future growth depends not on cheap labor or exports, but on its ability to innovate and transition to a high-income economy driven by technological advances.” As stated here, the growth of the economy depends on technological development.

3. ANALYSIS AND THE RESULTS

The following econometric analysis will be made by using World Bank (WB) World Developments Indicator (WDI) data for China and United States for the years 1985-2021 (37 years). There will be a time series analysis for China in this paper. We will try to understand China's growth. The data used in the analysis could be found online in my GitHub repository here in csv format.

The main aim of the paper is to understand and examine the role of investment and innovation in GDP Growth of China. To assess innovation we will use patent fillings. We will start with comparing China's growth to U.S. and then will move on understanding China's economic growth individiually. There might be control variables used as well if necessary. Those variables and their descriptive statistics could be seen below.

Descriptive Stats and Graphs:

For China:

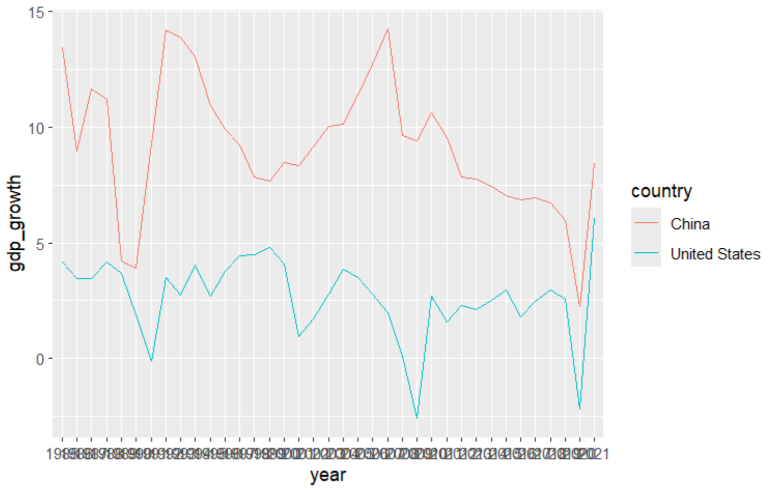
Variable	Mean	Min	Max
gdp_{growth}	9.2	2.2	14.2
$patent_{total}$	408597	8009	1585663
$patent_{total\ per\ population}$	2.969e-04	7.434e-06	1.123e-03
$patent_{resident}$	343773	3494	1426644
$patent_{resident\ per\ population}$	2.489e-04	3.275e-06	1.010e-03
$patent_{non\text{-}resident}$	64824	4051	159019
$patent_{non\text{-}resident\ per\ population}$	4.806e-05	3.520e-06	1.126e-04
Investment	40.33	33.57	46.66
inflation	5.138	-1.263	20.617
net trade	1.690e+11	-2.257e+10	5.762e+11
unemployment	3.539	1.800	5.610

For US:

Variable	Mean	Min	Max
$\mathbf{gdp}_{\mathbf{growth}}$	2.655	-2.576	6.055
$\mathbf{patent}_{\mathbf{total}}$	364463	115235	621453
$\mathbf{patent}_{\mathbf{total\ per\ population}}$	0.0012171	0.0004843	0.0018928
$\mathbf{patent}_{\mathbf{resident}}$	184070	63673	295327
$\mathbf{patent}_{\mathbf{resident\ per\ population}}$	0.0006172	0.0002676	0.0009141
$\mathbf{patent}_{\mathbf{non\text{-}resident}}$	180393	51562	336340
$\mathbf{patent}_{\mathbf{non\text{-}resident\ per\ population}}$	0.0005999	0.0002167	0.0010244
Investment	21.64	17.77	24.19
inflation	2.2027	0.6168	4.5660
net trade	-4.963e+11	-1.083e+12	-7.694e+10
unemployment	5.946	3.669	9.633

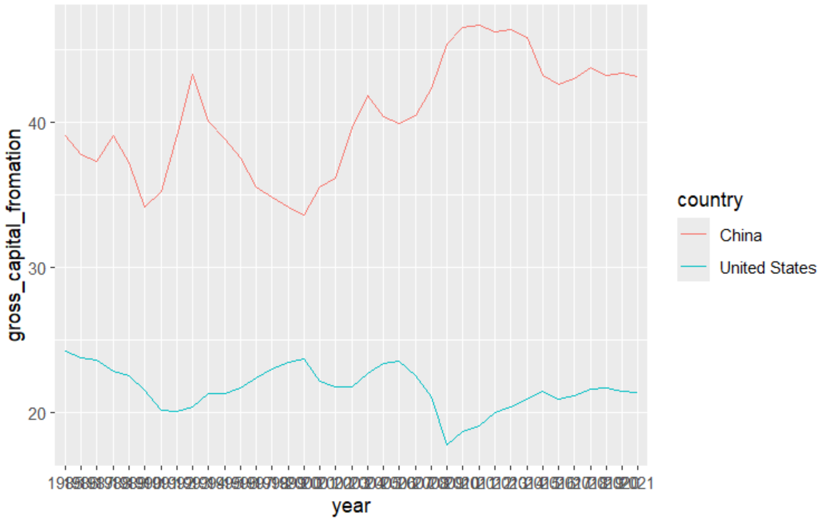
Comparison:

GDP Growth by years:



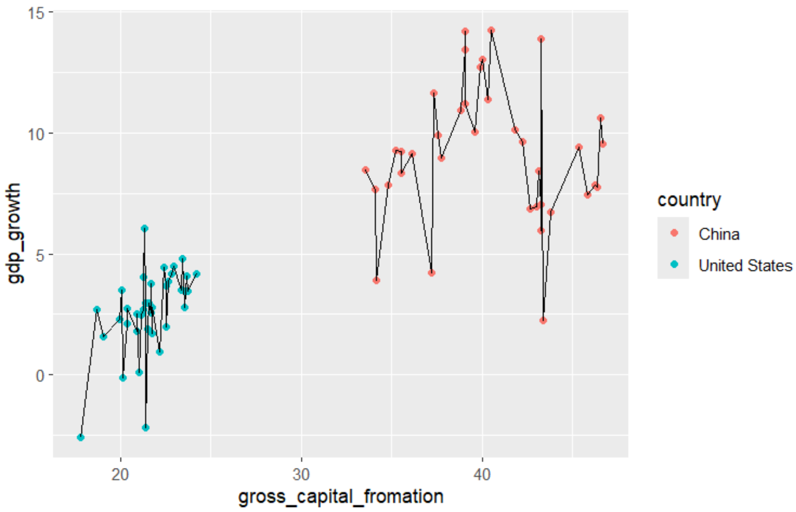
This Image shows how much China exceeds US in economic growth.

Investment:



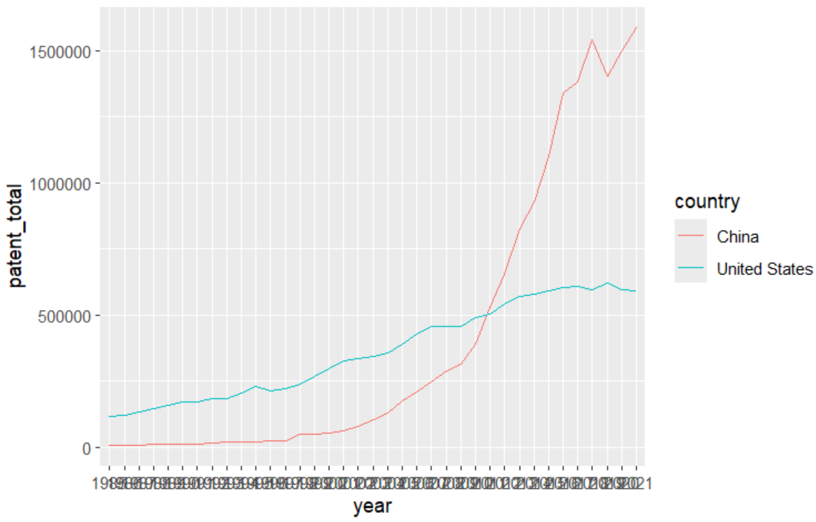
This image shows how China’s gross capital formation is higher than U.S. The graph looks very similar to GDP growth, thus there could be a relation.

Investment Differences:



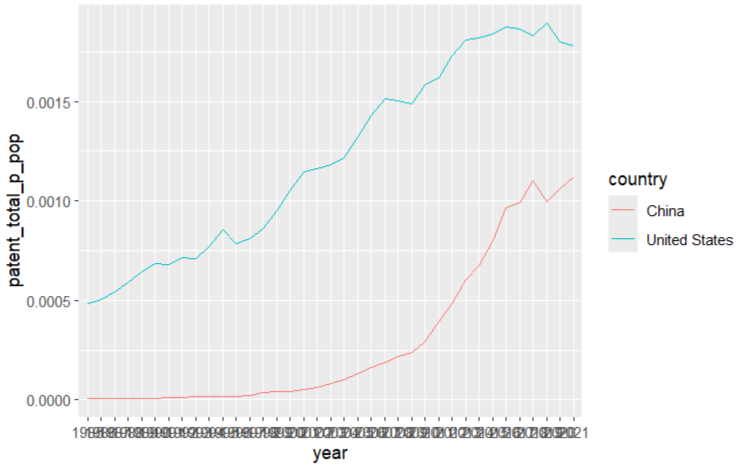
This image shows the different levels of investment in those countries indiviually. Both looks like there is a trend.

Total Patents:



This image shows China’s increasing patent filling in recent years.

Total Patents per population:



This image shows that altough China has increased its patent fillings significantly it is still behind U.S. in per capita comparison. Moreover, U.S. increases its per capita patent fillings significantly as well.

Now we have examined the data and we can move on to the econometric analysis.

Econometric Analysis, China Specific Examination:

$$\boldsymbol{model\ 1}: gdp_{growth} = \alpha + \beta_1 patent_{total} + \beta_2 Investment$$

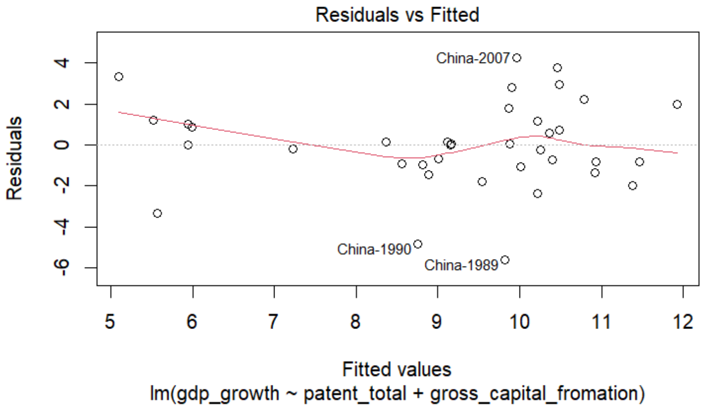
```
Call:
lm(formula = gdp_growth ~ patent_total + gross_capital_fromation,
    data = last_data_for_plm_china)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6182 -0.9719  0.0098  1.1724  4.2712

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -3.239e+00  4.618e+00  -0.701  0.48786
patent_total            -4.322e-06  8.706e-07  -4.965 1.91e-05 ***
gross_capital_fromation  3.522e-01  1.195e-01   2.948  0.00575 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.208 on 34 degrees of freedom
Multiple R-squared:  0.4207,    Adjusted R-squared:  0.3866
F-statistic: 12.35 on 2 and 34 DF,  p-value: 9.321e-05
```

The results:

Median residual value is close to 0, suggesting there is no systemmetical over/underestimation. However, the range of max and min suggests there are significant errors.

Coefficients are highly significantfor independent variables. Patent parameter is negative. As shown in descriptive statistics part, especially considering China's recent patent surge, it is not surprising it's parameter is negative as China's gdp didn't grow similarly in the same period. Also, As we have seen in comparison with US, gross capital formation having positive and significant effect is not surprising as well.

Despite significant coefficients there is low R^2. F-stat is high.



The results seem okay, however, there might be problems like non-stationarity, autocorrelation, multicollinearity, heteroskadasticity etc. For that, we need to test the model before coming to conclusions. It might be good to test for stationarity first with unit-roots via Augemented Dickey-Fuller Test (ADF).

STATIONARITY

ADF for $model\ 1$:

Hypothesis test for ADF:

H0: Series is non-stationary (There are Unit Root)

H1: Series is stationary (There is no Unit Root)

Thus, if a p-value is lower than significance levels, than the null hypothesis is rejected and the series is proved to be stationary under respective significance level.

Test results could be seen in the next page.

Test for gdp_growth variable:

Lag order 0:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gdp_growth
Dickey-Fuller = -3.2822, Lag order = 0, p-value = 0.08999
alternative hypothesis: stationary
```

Lag order 1:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gdp_growth
Dickey-Fuller = -3.67, Lag order = 1, p-value = 0.04131
alternative hypothesis: stationary
```

Lag order 2:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gdp_growth
Dickey-Fuller = -2.789, Lag order = 2, p-value = 0.2656
alternative hypothesis: stationary
```

Results show that GDP growth has a p-value lower than 0.1 when lag order is set to 0. Meaning that gdp_growth is stationary under 0.1 significance level. At First lag difference p-value is 0.04 < 0.05. At first lag it is even stationary under 0.05 level. At lag order = 2, it loses it's stationarity as p-value rises above 0.1 level. We can also check for differentiated versions like:

Differenced 1 lag order 0:

```
        Augmented Dickey-Fuller Test

data:  diff_growth_china
Dickey-Fuller = -5.3413, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Differenced 1 lag order 1:

```
        Augmented Dickey-Fuller Test

data:  diff_growth_china
Dickey-Fuller = -5.3999, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary
```

Differenced 1 lag order 2:

```
        Augmented Dickey-Fuller Test

data:  diff_growth_china
Dickey-Fuller = -3.5117, Lag order = 2, p-value = 0.05702
alternative hypothesis: stationary
```

Conclusion: gdp_growth variable is stationary under lag orders 0 and 1but not 2. One time differenced gdp_growth (I(1)), is stationary under lag orders 0, 1 and 2 but it loses its significancy under lag order 2. Since it significant even without differencing it might be adequate to use I(0) however, in some cases I(1) could be helpful as well.

Test for patent_total variable:

Lag order 0:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$patent_total
Dickey-Fuller = -0.58252, Lag order = 0, p-value = 0.971
alternative hypothesis: stationary
```

Lag order 1:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$patent_total
Dickey-Fuller = -0.72559, Lag order = 1, p-value = 0.9589
alternative hypothesis: stationary
```

Lag order 2:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$patent_total
Dickey-Fuller = -1.2314, Lag order = 2, p-value = 0.8729
alternative hypothesis: stationary
```

For patent_total vairable all lag orders are highly insignificant we need to check for differences.

Difference 1 lag order 0:

```
        Augmented Dickey-Fuller Test

data:  diff_patent_china
Dickey-Fuller = -5.3523, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Difference 1 lag order 1:

```
        Augmented Dickey-Fuller Test

data:  diff_patent_china
Dickey-Fuller = -2.599, Lag order = 1, p-value = 0.3401
alternative hypothesis: stationary
```

Difference 1 lag order 2:

```
        Augmented Dickey-Fuller Test

data:  diff_patent_china
Dickey-Fuller = -2.7301, Lag order = 2, p-value = 0.289
alternative hypothesis: stationary
```

Difference 1 lag order 3:

```
        Augmented Dickey-Fuller Test

data:  diff_patent_china
Dickey-Fuller = -3.684, Lag order = 3, p-value = 0.04061
alternative hypothesis: stationary
```

We see that first difference with lag order 0 is highly significant at p-value 0.01. In higher lag orders it is not always significant like order 1 and 2. At order 3 it is significant again. Thus, this test shows us we can make patent_total stationary by differencing it one time. However, we may need to use higher difference levels for some analysis. So it would be good to test for difference 2 as well.

Difference 2 lag order 0:

```
        Augmented Dickey-Fuller Test

data:  diff2_patent_china
Dickey-Fuller = -12.351, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Difference 2 lag order 1:

```
        Augmented Dickey-Fuller Test

data:  diff2_patent_china
Dickey-Fuller = -4.4719, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary
```

Difference 2 lag order 2:

```
        Augmented Dickey-Fuller Test

data:  diff2_patent_china
Dickey-Fuller = -3.0114, Lag order = 2, p-value = 0.1804
alternative hypothesis: stationary
```

Second difference shows highy significant p-values for lag orders 0 and 1 but loses its significancy at lag order 2.

Test for gross_capital_formation (Investment) variable:

Lag order 0:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gross_capital_fromation
Dickey-Fuller = -2.0795, Lag order = 0, p-value = 0.5422
alternative hypothesis: stationary
```

Lag order 1:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gross_capital_fromation
Dickey-Fuller = -2.8332, Lag order = 1, p-value = 0.2483
alternative hypothesis: stationary
```

Lag order 2:

```
        Augmented Dickey-Fuller Test

data:  last_data_for_plm_china$gross_capital_fromation
Dickey-Fuller = -2.1484, Lag order = 2, p-value = 0.5154
alternative hypothesis: stationary
```

The results for investment's p-values are very high to reject the null hypothesis so we see that the series is non-stationary, however, we may convert it to stationary by applying differences. Let's try:

Difference 1 Lag order 0:

```
        Augmented Dickey-Fuller Test

data:  growth_diff_patent_investment_china$diff_gross_capital_formation
Dickey-Fuller = -4.1754, Lag order = 0, p-value = 0.01425
alternative hypothesis: stationary
```

Difference 1 Lag order 1:

```
        Augmented Dickey-Fuller Test

data:  growth_diff_patent_investment_china$diff_gross_capital_formation
Dickey-Fuller = -4.5639, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary
```

Difference 1 Lag order 2:

```
        Augmented Dickey-Fuller Test

data:  growth_diff_patent_investment_china$diff_gross_capital_formation
Dickey-Fuller = -3.5254, Lag order = 2, p-value = 0.05503
alternative hypothesis: stationary
```

The series appears to be converting to stationary under first differences.It is very highly significant at lag order 0 and 1 but loses some significance at lag order=2. It may be appropriate to use integration order of 2 sometimes. So let's check that as well.

Difference 2 Lag order 0:

```
        Augmented Dickey-Fuller Test

data:  diff2_gross_capital_formation
Dickey-Fuller = -6.3707, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Difference 2 Lag order 1:

```
        Augmented Dickey-Fuller Test

data:  diff2_gross_capital_formation
Dickey-Fuller = -6.1236, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary
```

Difference 2 Lag order 2:

```
        Augmented Dickey-Fuller Test

data:  diff2_gross_capital_formation
Dickey-Fuller = -5.1891, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

At differenciation level 2, it is very highly significant and thus, stationary.

In conclusion, at lag orders of 0, GDP growth is stationary already at significance level 0.1 without requiring any difference. This means that series is stationary, however if we may need higher significance levels, it may not be adequate as it is only significant at lowest significance level that is 0.1. Also, in some cases we may need differenced version of gdp_growth. When first difference is taken it is significant at even 0.01 significance level. Highly confirming that the series is stationary at I(1). For total patents, I(0) is way far from significance threshold, making it non-stationary. However, if we check difference of 1, we have high significance at level 0.01 making I(1) of total patents series non-stationary. I f we further check second difference level of total patents we see that it is also highly significant, thus stationary here as well. For gross_capital_formation, the series appears to be non-stationarity with a high p-value failing to reject null hypothesis. At first difference though, it becomes stationary with a very low p-value conforming it's stationarity. Similarly for I(2) as well. For lag orders of different than 0 the images I put could be inspected for appropriate analysis.

Another method to reduce stationarity in the series could be by applying logarithms into it. Let's test for all series in logarithmic format:

ADF test for log(gdp_growth):

```
        Augmented Dickey-Fuller Test

data:  log(last_data_for_plm_china$gdp_growth)
Dickey-Fuller = -3.7478, Lag order = 0, p-value = 0.03538
alternative hypothesis: stationary
```

ADF test for log(patent_total):

```
	Augmented Dickey-Fuller Test

data:  log(last_data_for_plm_china$patent_total)
Dickey-Fuller = -2.3201, Lag order = 0, p-value = 0.4484
alternative hypothesis: stationary
```

ADF test for log(gross_capital_formation):

```
	Augmented Dickey-Fuller Test

data:  log(last_data_for_plm_china$gross_capital_fromation)
Dickey-Fuller = -2.5162, Lag order = 3, p-value = 0.372
alternative hypothesis: stationary
```

The results give us stationarity at 0.05 level significance for gdp_growth increasing the significancy of 0.1 level without taking logarithm. However, the results for patent_total and gross_capital_formation is still insignificant.

Model Update: $\boldsymbol{model\ 1_{updated}}: gdp_{growth} = \alpha + \beta_1 \Delta patent_{total} + \beta_2 \Delta Investment$

With addition of Δ (delta), we check first order differences which transforms patent_total and gross_capital_formation to stationary. We may also establish a model like:

$\boldsymbol{model\ 1_{alternative}}: \Delta gdp_{growth} = \alpha + \beta_1 \Delta patent_{total} + \beta_2 \Delta Investment$, since gdp_growth is also significant at I(1) and it may be appropriate to check relative levels for both but since gdp growth is already percentaged difference of GDP, I prefer to use $\boldsymbol{model\ 1_{updated}}$. Now we can check for autocorrelations and partial autocorrelations.

AUTOCORRELATION

Autocorrelations happen when a time series have serial correlation between its values in different time intervals.

To check autocorrelation we can use Durbin-Watson test for the first version of $\boldsymbol{model\ 1}$, however using Durbin Watson test for $\boldsymbol{model\ 1_{updated}}$ may not be appropriate as it includes lagged independent variables. Instead for $\boldsymbol{model\ 1_{updated}}$ we can use Breusch-Godfrey test (LM test).

$model1$ Durbin-Watson Test:

```
> durbinWatsonTest(model_patent_investment)
 lag Autocorrelation D-W Statistic p-value
   1       0.2988649      1.282313   0.008
 Alternative hypothesis: rho != 0
```

As for model 1 with n=35, and independent variable=2, dL=1.343 and dU=1.584. Since DW value of the **$model1$** is 1.282 and it is below 1.343, we find that the model has positive autocorrelation.

$model1_{updated}$ Breusch-Godfrey Test:

```
	Breusch-Godfrey test for serial correlation of order up to 1

data:  model_patent_investment_updated
LM test = 11.34, df = 1, p-value = 0.0007584
```

Here, to assess autocorrelation level we check p-value. The hypothesis for this is:

H0: no auto-correlation

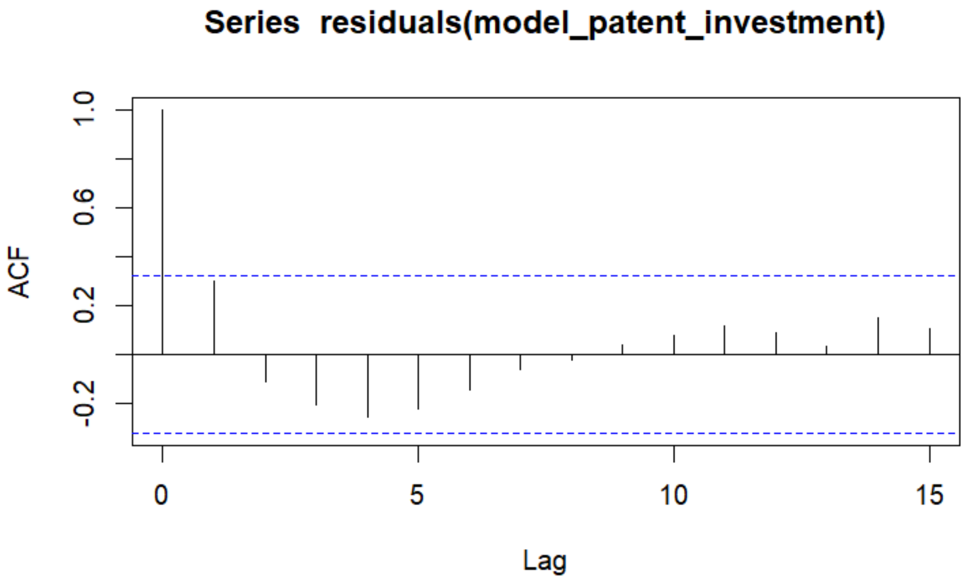
H1: auto-correlation

Since, we reject the null hypothesis here as our p-vallue is below 0.01, we find there is autocorrelation.

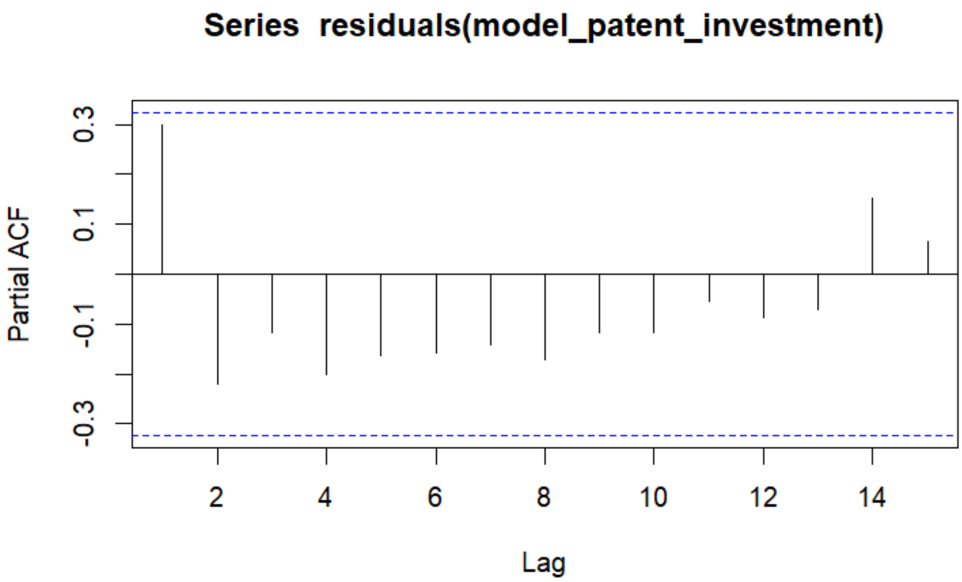
Both models show autocorrelation. To understand autocorrelation structure we can also use ACF and PACF plots. Results can be seen in the next page.

For $model1$ ACF and PACF plots:

ACF:



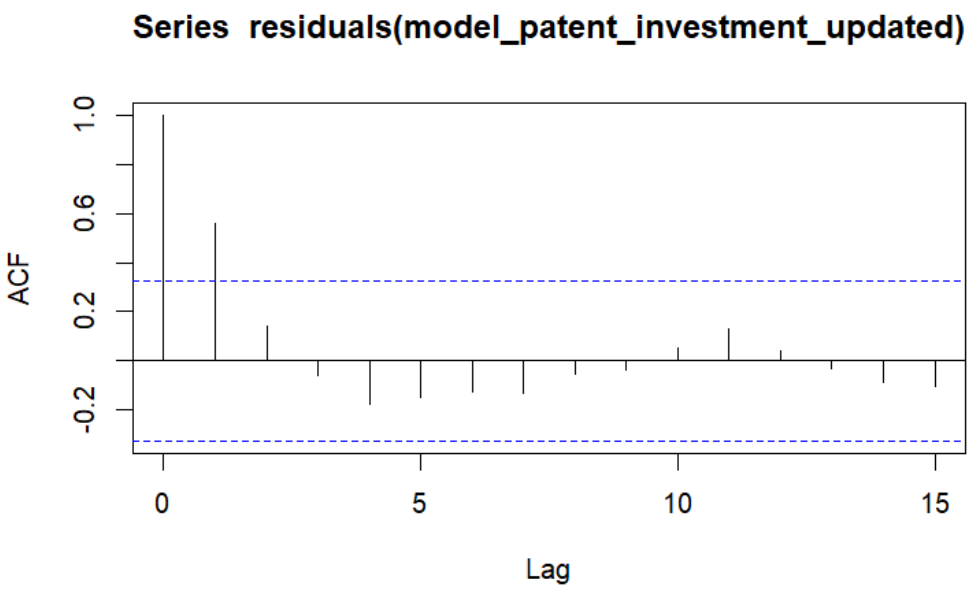
PACF:



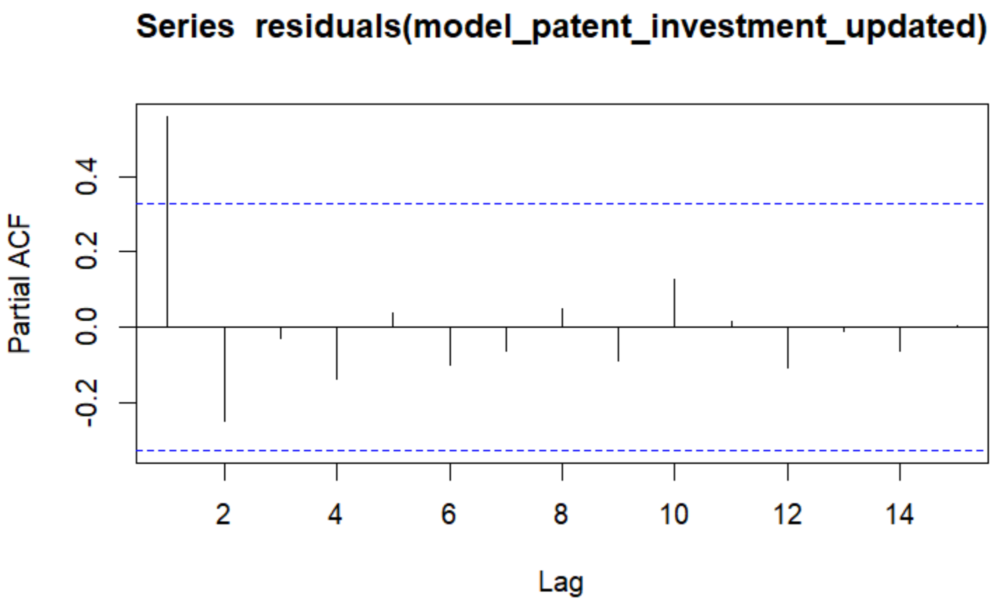
Through ACF and PACF plots here we can understand the aurocorrelation structure in the model. For this $\boldsymbol{model1}$ we can see that at ACF there is one spike and then fast cutoff. This cutoff at ACF at level 1 suggest there might be MA(1). PACF series don’t have any spikes within confidence band, thus, there might not be AR() element in the best specification.

$model1_{updated}$ ACF and PACF Plots:

ACF:



PACF:



Similar to ACF and PACF results of $\boldsymbol{model1}$, $\boldsymbol{model1_{updated}}$ results here suggests there could be ARMA specifications. For this series the best specification could be ARMA(1,2) as there are 2 spikes at ACF and 1 spike at PACF.

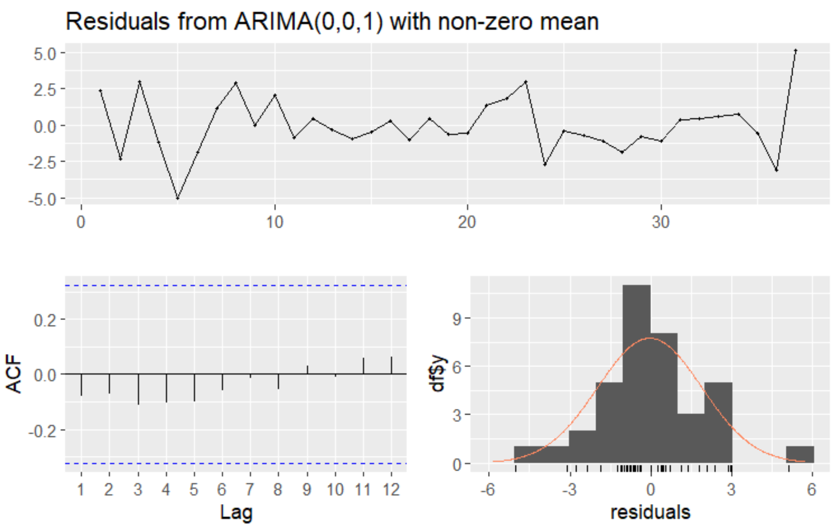
Now I can apply ARIMA models I have specified above to eliminate autocorrelation. I will also check for both mean and drift and I will include only if there are significant mean or drift. For $\boldsymbol{model1}$ I have specified MA(1) model, so let's try that first.

MA(1) for $\boldsymbol{model1}$:

```
Series: residuals(model_patent_investment)
ARIMA(0,0,1) with zero mean

Coefficients:
         ma1
      0.6194
s.e.  0.1764

sigma^2 = 3.747:  log likelihood = -76.67
AIC=157.35   AICc=157.7   BIC=160.57
```

The results on the left is with zero mean and no drift. They were insignificant and thus, not included. For ma1 variable we see coefficient is 0.61 and standard error is 0.17 indicating there is a high t-value. We can check AIC and BIC values for its goodness of fit. AIC and BIC values are often used in comparisons, and more than two unit differences are considered significant improvements for most of the time. We can use them to compare this ARMA specification with other specifications if we want.

Residuals Graphs and Ljung-Box Test for MA(1):



Residuals should be randomly distributed with no trend or pattern. Here no pattern is observed. For ACF, there is no significant spikes ans residual distribution looks normal.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,0,1) with zero mean
Q* = 2.1181, df = 6, p-value = 0.9085

Model df: 1.   Total lags used: 7
```

Ljung-Box test's null hypothesis is that there is no autocorrelation so, p-value being higher than significant levels show that we fail to reject the null hypothesis thus, no aoutocorrelation exists.

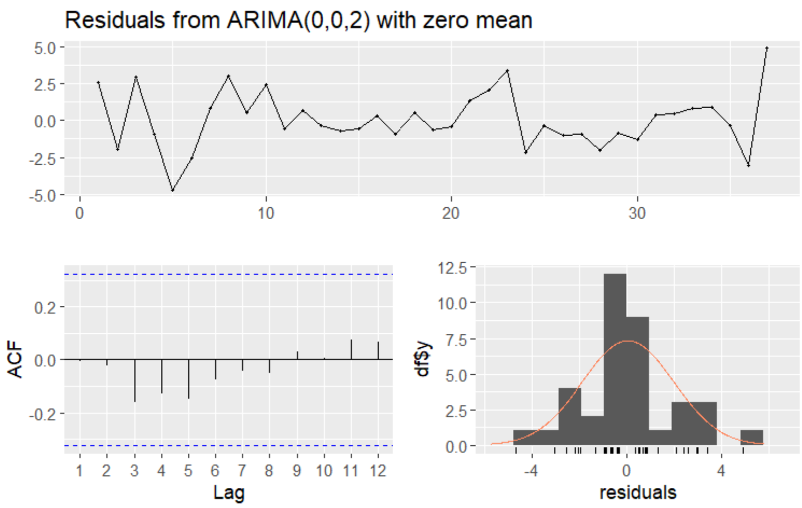
The model MA(1) is successful with eliminating autocorrelation but we can have other ARMA models and compare their AIC and BIC values. "auto.arima()" function in R program gives best AIC, BIC models, and when I check it is MA(1) as well, so we can say MA(1) is the best model however to test let's try ARMA(1,1) for example.

ARMA(1,1) for $model1$:

```
Series: residuals(model_patent_investment)
ARIMA(1,0,1) with zero mean

Coefficients:
          ar1     ma1
      -0.1411  0.7013
s.e.   0.2821  0.2051

sigma^2 = 3.827:  log likelihood = -76.55
AIC=159.09   AICc=159.82   BIC=163.92
```

In this model, we see ar1 additional to ma1. However ar1 is not significant. I also checked for mean and drift and they were not significant as well. The AIC and BIC values are approximately 2 units high, indicating this is a significantly worse model.

Residuals Graphs and Ljung-Box Test for ARMA(1,1):



This ARMA(1,1) model also eliminates autocorrelation as seen by no spikes at ACF or high p-value in Ljung-Box test below, however since the first specification with MA(1) has greater AIC, BIC we can say it is more appropriate to choose that MA(1) model.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,0,2) with zero mean
Q* = 3.2126, df = 5, p-value = 0.6672

Model df: 2.   Total lags used: 7
```

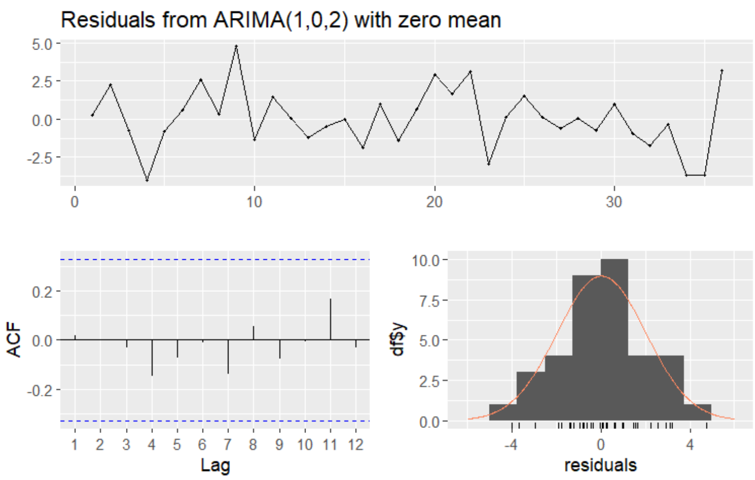
Now we can move on to find the best ARIMA specification for our updated model. As I have argued above $\boldsymbol{model1_{updated}}$ could have ARMA(1,2) model. Let's try that:

ARMA(1,2) for $\boldsymbol{model1_{updated}}$:

```
Series: residuals(model_patent_investment_updated)
ARIMA(1,0,2) with zero mean

Coefficients:
         ar1     ma1     ma2
      0.1362  0.5952  0.0719
s.e.  0.7354  0.7463  0.4666

sigma^2 = 4.232:  log likelihood = -75.76
AIC=159.52   AICc=160.81   BIC=165.85
```

In the ACF and PACF plot I have thought ARMA(1,2) might be the best fit however, there is not even one significant variable in this specification including mean and drift as well. There might be better possibilities. However before checking that first let's see the residuals structure.

Residuals Graphs and Ljung-Box Test for ARMA(1,2):



These results show that series with ARMA(1,2) specification doesn't have autocorrelation even if the coefficients for ARMA(1,2) values are not significant.

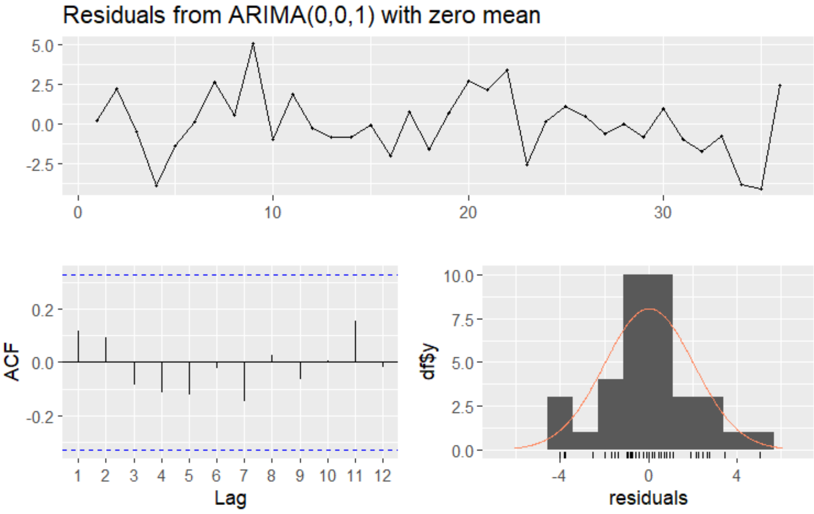
```
        Ljung-Box test

data:  Residuals from ARIMA(1,0,2) with zero mean
Q* = 2.1384, df = 4, p-value = 0.7103

Model df: 3.   Total lags used: 7
```

"auto.arima()" for $model1_{updated}$: MA(1)

```
Series: residuals(model_patent_investment_updated)
ARIMA(0,0,1) with zero mean

Coefficients:
         ma1
      0.6329
s.e.  0.1162

sigma^2 = 4.098:  log likelihood = -76.22
AIC=156.44   AICc=156.8   BIC=159.6
```

autor.arima() found the smallest AIC and BIC is **MA(1) model**. In the model MA(1), ma1 is highly significant. If we check AIC and BIC, we see there is more than 2 unit differences in both criteria, thus we can say that MA(1) model is significantly better than ARMA(1,2) model.

Residuals Graphs and Ljung-Box Test for MA(1):



The residuals plot is not very different from ARMA(1,2) and is also very good at eliminating autocorrelation. The Ljung-Box test also shows very high p-value (higher than ARMA(1,2)) signifying its strength on eliminating autocorrelation.

```
        Ljung-Box test

data:  Residuals from ARIMA(0,0,1) with zero mean
Q* = 3.4502, df = 6, p-value = 0.7506

Model df: 1.   Total lags used: 7
```

Now we have foud best ARIMA specifications for both $model1_{updated}$ and $model1$. Turned out both models' best ARMA specifications are MA(1). Further on, we can test for heteroskadasticity. Heteroskadasticity happens when the variance of the errors in a regression model is not constant across all observations of the independent variables. We can apply the Breusch-Pagan test, the White test, and the Goldfeld-Quandt (GQ) test. Let's look for those in our models $model1_{updated}$ and $model1$.

HETEROSKADASTICITY

Breusch-Pagan Test Hypothesis:

H0: homoskedastic

H1: heteroskadastic

White Test Hypothesis: (It is non-studentized Breusch-Pagan test)

H0: homoskedastic

H1: heteroskadastic

Goldfeld-Quandt (GQ) Test Hypothesis:

H0: homoskedastic

H1: heteroskadastic

Breusch-Pagan test for $model1$:

```
        studentized Breusch-Pagan test

data:  model_patent_investment
BP = 0.44839, df = 2, p-value = 0.7992
```

Since p-value is higher than significance thresholds, we fail to reject to null hypothesis that leads us to find that the series is homoskedastic.

White test for $model1$:

```
        Breusch-Pagan test

data:  model_patent_investment
BP = 0.55565, df = 2, p-value = 0.7574
```

Since p-value is higher than significance thresholds, we fail to reject to null hypothesis that leads us to find that the series is homoskedastic.

Goldfeld-Quandt test for $model1$:

```
        Goldfeld-Quandt test

data:  model_patent_investment
GQ = 0.62525, df1 = 16, df2 = 15, p-value = 0.8193
alternative hypothesis: variance increases from segment 1 to 2
```

Since p-value is higher than significance thresholds, we fail to reject to null hypothesis that leads us to find that the series is homoskedastic.

Breusch-Pagan test for $model1_{updated}$:

```
        studentized Breusch-Pagan test

data:  model_patent_investment_updated
BP = 3.9853, df = 2, p-value = 0.1363
```

Since p-value is higher than significance thresholds, we fail to reject to null hypothesis that leads us to find that the series is homoskedastic.

White test for $model1_{updated}$:

```
        Breusch-Pagan test

data:  model_patent_investment_updated
BP = 4.6712, df = 2, p-value = 0.09675
```

p-value is slighlty lower than 0.1 significance level, thus, despite it's slightness, we reject the null hypothesis at 0.1 significance level and find homoskedasticity in this situation, however at other significance levels we still fail to reject the null hypothesis.

Goldfeld-Quandt test for $model1_{updated}$:

```
        Goldfeld-Quandt test

data:  model_patent_investment_updated
GQ = 1.3827, df1 = 15, df2 = 15, p-value = 0.269
alternative hypothesis: variance increases from segment 1 to 2
```

Since p-value is higher than significance thresholds, we fail to reject to null hypothesis that leads us to find that the series is homoskedastic.

Heteroskadasticity Conclusion:

There is only one rejection in all 6 test applications and it is only slightly with p-value at 0.09675 at White Test for $\boldsymbol{model1_{updated}}$. Since heteroskadasticity in other test for $\boldsymbol{model1_{updated}}$ is also rejected, I will accept that it is homoskedastic, however if we had found homoskedasticity we would have to do one of the followings to address the problem: using heteroskedasticity-consistent standard errors (HCSE), using weighted least squares (WLS), using generalized least squares (GLS). Now we can move on to test for multicollinearity in our models.

MULTICOLLINEARITY

Now, we have concluded heteroskadasticity as well, we can also look for multicollinearity with variance inflation factor (VIF) and correlattion coefficients. Multicollinearity occurs when two or more independent variables in a regression model are highly correlated. In our case, since we have 2 independent variables, for multicollinearity we need to specificly look for those variables. For VIF, values greater than 10 indicate high multicollinearity, and for correlation coefficients, values above 0.8 or below -0.8 indicate high correlation, thus a possible multicollinearity.

VIF for $\boldsymbol{model1}$:

```
vif(model_patent_investment)
        patent_total  gross_capital_fromation
            1.647227                 1.647227
```

The value is very small and away from 10, indicating no or small multicollinearity.

VIF for $\boldsymbol{model1_{updated}}$:

```
vif(model_patent_investment_updated)
        diff_patent_china diff_gross_capital_formation
                 1.003332                     1.003332
```

The value is very small and close to 1, indicating no multicollinearity.

Alternatively, if we can check correlation matrix's:

For $model1$:

```
> cor_matrix_2
                        patent_total gross_capital_fromation
patent_total               1.0000000               0.6268327
gross_capital_fromation    0.6268327               1.0000000
```

0.62 is a bit high but common threshold for high correlation is 0.8, thus being lower than that indicates positive sign for no multicollinearity.

For $model1_{updated}$:

```
> cor_matrix_1
                             diff_patent_china diff_gross_capital_formation
diff_patent_china                   1.00000000                  -0.05762393
diff_gross_capital_formation       -0.05762393                   1.00000000
```

Correlation is so low with -0.05, indicating no multicollinearity.

In conclusion, we have found no multicollinearity in both of the models.

MODEL SPECIFICATION

Ramsey RESET test:

Ramsey RESET test is used to test for omitted variable bias (OVB) and model misspecifications. For example, if we test for our $\boldsymbol{model1}$ here we find the result as:

```
        RESET test

data:  model_patent_investment
RESET = 2.2161, df1 = 2, df2 = 32, p-value = 0.1255
```

In here the hypothesis is:

H0: model is not misspecified.

H1: model is misspecified

By the result we find that the model is not misspecifed as p-value is higher than significance levels and thus, we fail to reject the null hypothesis.

4. CONCLUSION

The Augmented Dickey-Fuller (ADF) test revealed that while GDP growth was stationary, patents and gross capital formation required differencing to achieve stationarity. Applying logarithms didn't work for patents and investment. Both the initial and updated models exhibited autocorrelation, as indicated by the Durbin-Watson and Breusch-Godfrey tests. ACF and PACF plots suggested the presence of MA(1) and ARMA(1,2) specifications. However, for both models MA(1) model was found to be the best fit for eliminating autocorrelation. as confirmed by the AIC/BIC values. Heteroskadasticity results gave the results that the models are homoskedastic without a problem. Variance Inflation Factor (VIF) values and correlation matrices showed no significant multicollinearity in either model, confirming the robustness of the regression coefficients. The Ramsey RESET test indicated that the model 1 was not misspecified, suggesting that the chosen variables adequately capture the relationship between investment, innovation, and GDP growth.

5. BIBLIOGRAPHY

- Solow, R. M. (1956). *A contribution to the theory of economic growth*. Quarterly Journal of Economics, 70(1), 65-94. https://doi.org/10.2307/1884513
- Romer, P. M. (1990). *Endogenous technological change*. Journal of Political Economy, 98(5, Part 2), S71-S102. https://doi.org/10.1086/261725
- Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical, and statistical analysis of the capitalist process*. McGraw-Hill.
- Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Business.
- Penrose, E. T. (1951). *The economics of the international patent system*. Johns Hopkins Press.
- Lin, J. Y. (2012). *Demystifying the Chinese economy*. Cambridge University Press.