Nam Nguyen

Do innovation and technology drive economic growth?

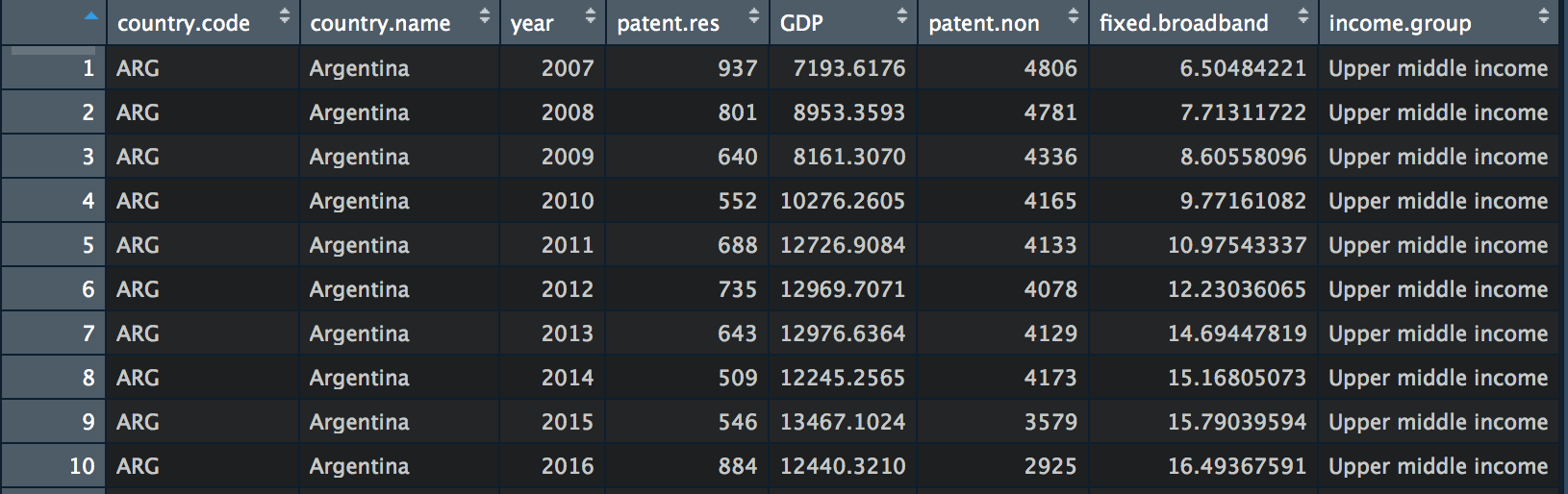
**I Introduction**

In the current fast-changing world, the idea that innovation and technology drive economic growth is undeniable. It is widely known that innovation is creating values by introducing something new. In different contexts and industries, the definitions of innovation might be slightly difference. For instance, innovation in business may be finding a new process to improve performance and increase output; innovation in medical and pharmacy may be inventing new vaccines that save million lives; innovation in technology can be designing new machines to boost productivity. There is an enormous number of researches focusing on the impact of innovation and technology on a particular industry or field. However, the relationship between innovation and the economy as a whole has yet been fully investigated as the term “innovation” itself is broad and continuing to grow. This research will assess the relationship between innovation, and technology and economic growth using data from 84 countries.

**II Data sources**

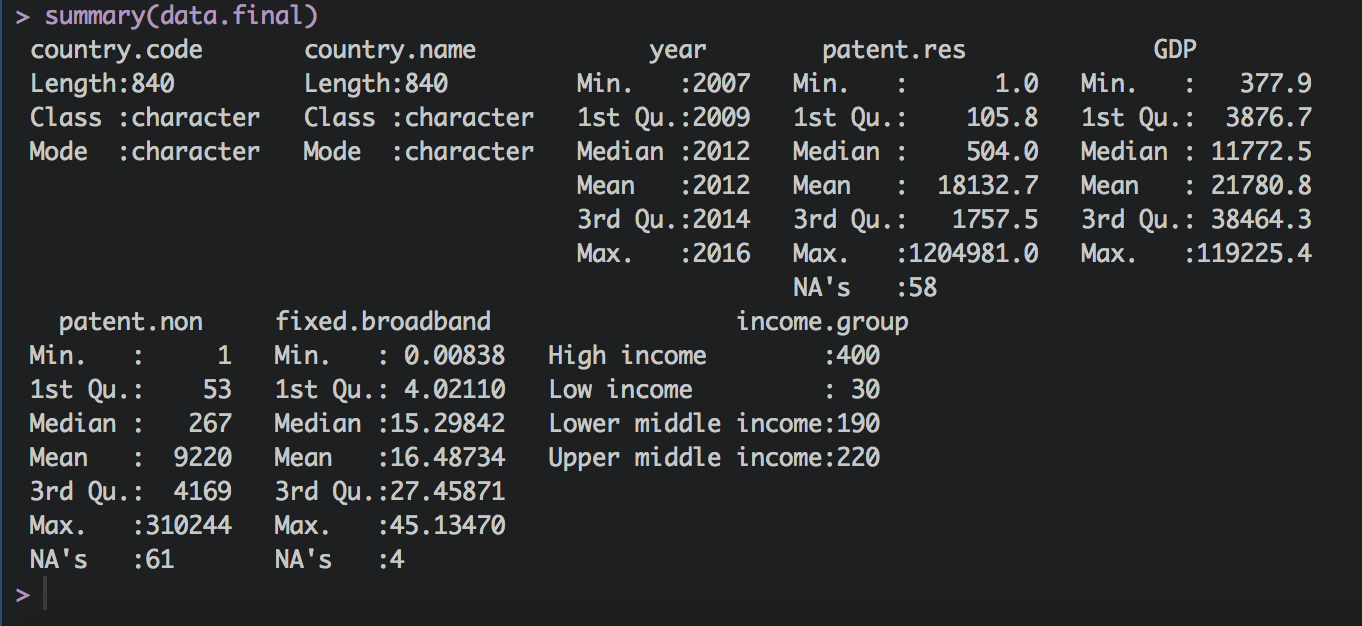
There are three variables selected to indicate the level of innovation and technology namely Patent Applications (Residents), Patent Application (Non-residents), and Fixed Broadband Subscriptions. Patent Applications are the number of patents that are filed in home country by residents or non-resident. Fixed Broadband Subscriptions is the number of Internet Subscription (both high speed and landline). The independent variable is GDP growth per capita, which is Gross Domestic Product divided by midyear population. All the data is retrieved from Word Bank public data set from 2007 to 2016. The original datasets contain over 195 countries. However, due to the missing values within some countries, this research will solely focus on 84 countries with available data for all four variables.

*Table 1: Sample panel dataset*

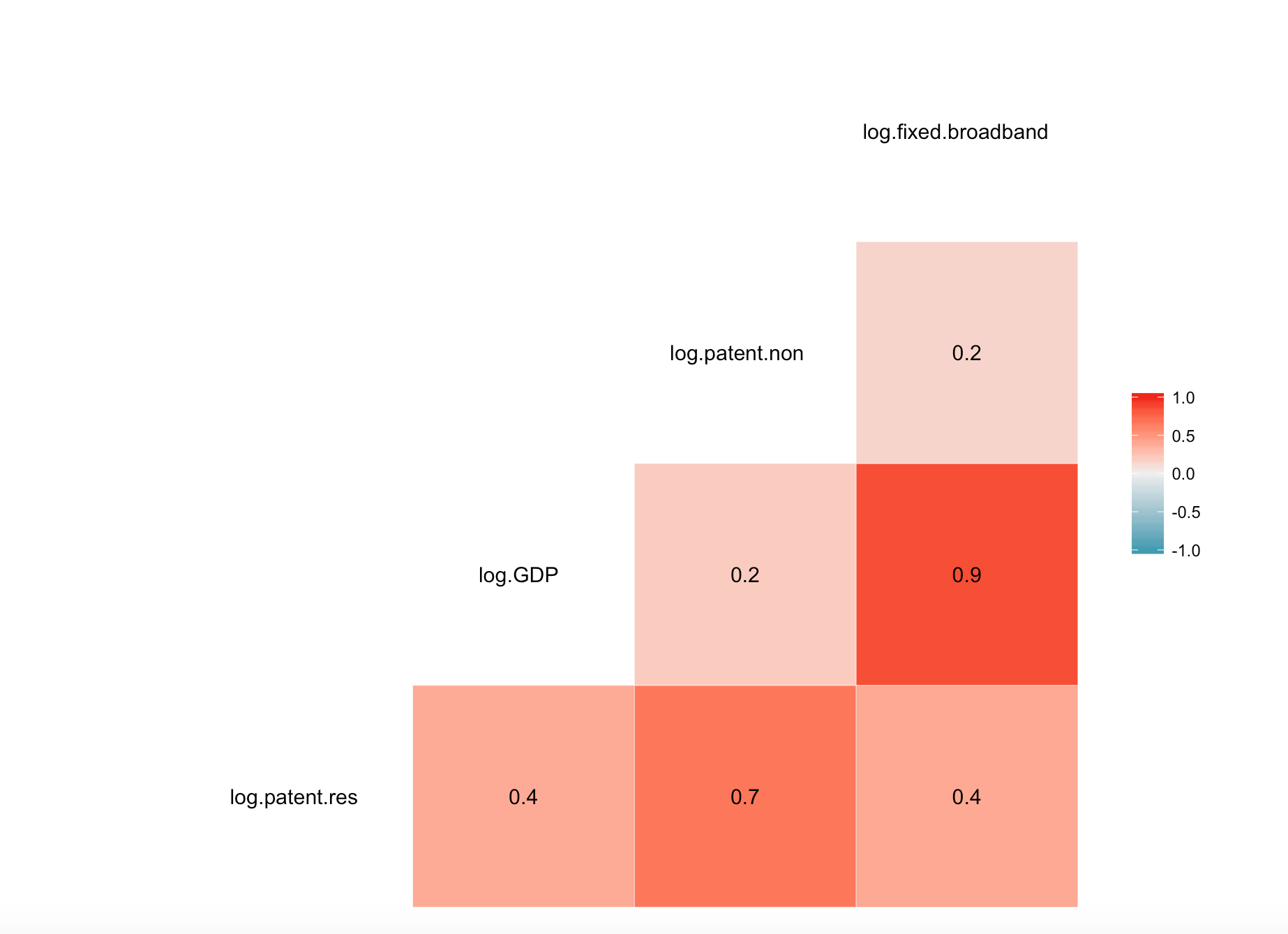
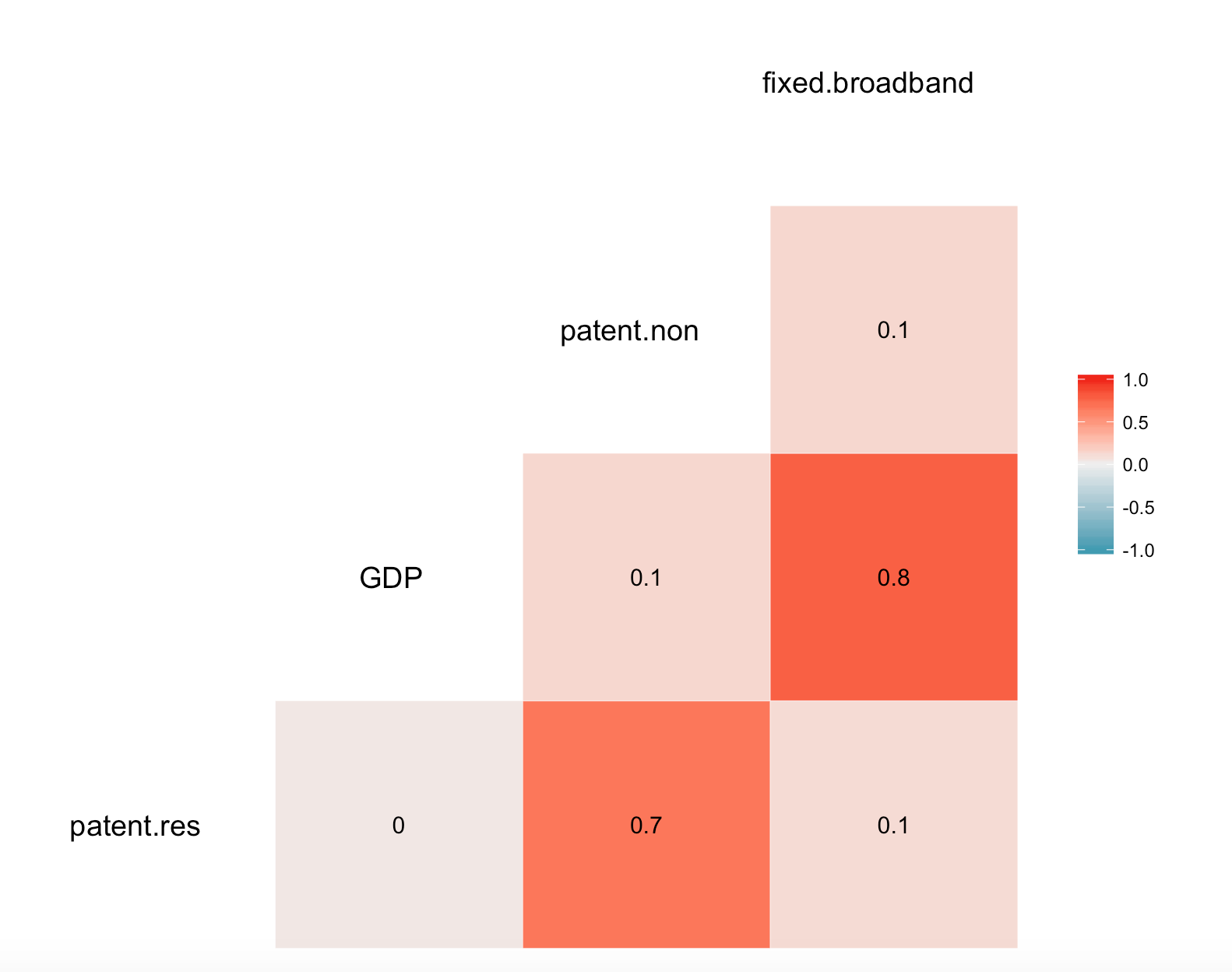


**III Data Summary**

*Table 2: Summary of the data set*



This is a panel data of 84 countries in a period of 10 years. There are some missing values in both Patents Applications and Fixed Broadband. Therefore, the panel is unbalanced. However, we do not omit missing observations if those are not problematic that is, in the middle of the period. Most of the missing data are either in the first two years (2007 and 2009) or the last year (2017). In Patent Applications (Residents), China filed the largest number of patents in 2017 (1,204,981 patents) while some other Low Income group countries filed only one patent a year.



*Figure 1: Correlation plot on raw data Figure 2: Correlation plot (log transformation)*

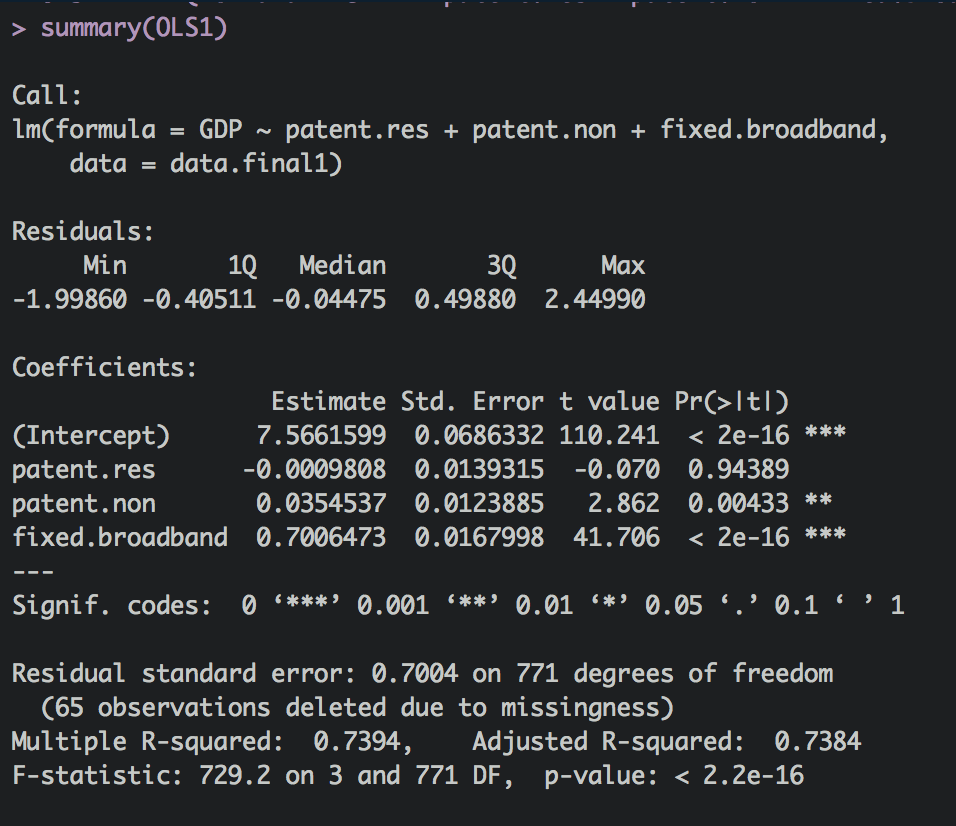
Based on Figure 1, there are clearly a strong, positive correlation between Fixed Broadband Subscription and GDP per capita (0.8). Nevertheless, the correlation between Patents Applications (both Residents and Non-Residents) is almost zero. As GDP per capita usually grows exponentially so taking log of GDP per capita may help smoothen the data. Moreover, log transformation for other variables also lends a hand to model explanation as the coefficients are basically elasticity. Figure 2 shows that there is positive relationship between GDP per capita and other independent variables.

**IV Discussion**

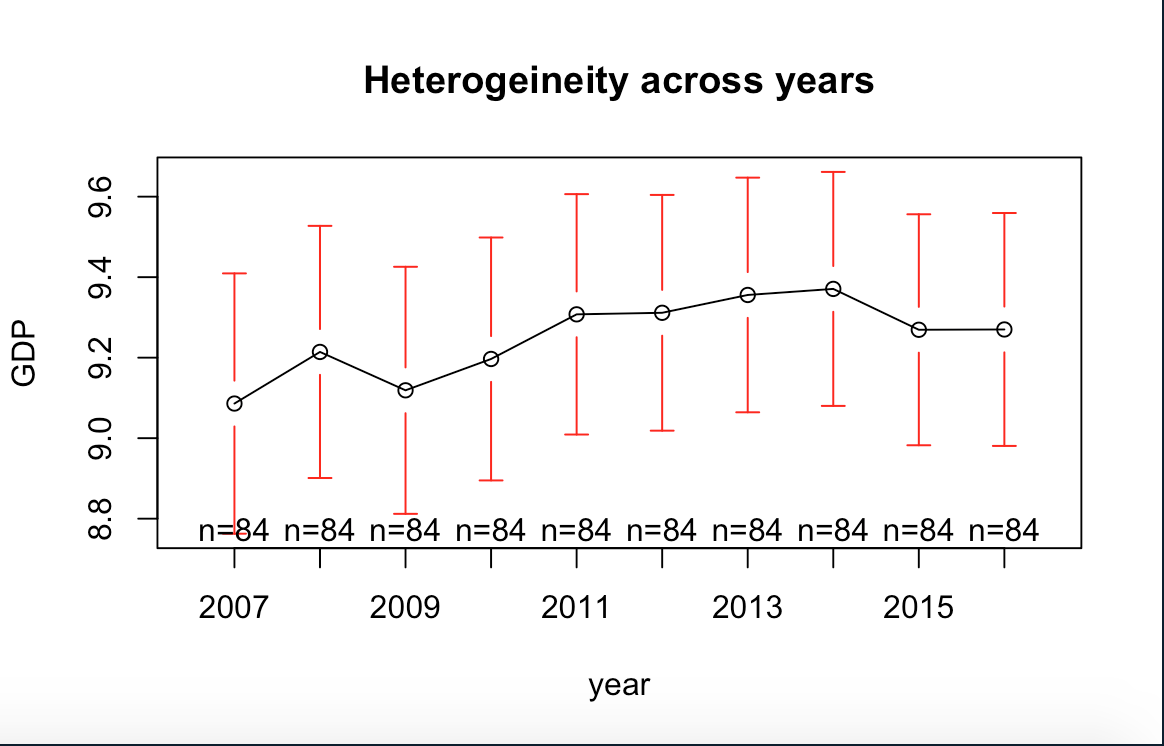
There are four economics models that could be appropriate for the panel dataset including Ordinary Least Squares, Pooled Model (OLS), OLS with Dummy Variable, Fixed Effects within group (Demeaned), and Random Effects.

1. **OLS, Pooled Model:**  = + \*log + \*log + \*log +

*Table 3: OLS summary*



Surprisingly, while the Fixed Broadband Subscriptions and Patent Applications (Non-residents) have highly statistical significant coefficients, Patent Applications (Residents) has an insignificant coefficient. However, regular OLS regression does not consider heterogeneity across group of countries or years of time period. Therefore, this model is not only biased but also inconsistent.



*Figure 3: Heterogeneity across years of GDP per capita*

Looking at Figure 3, there is a lot of volatility and fluctuation in GDP per capita throughout 10-year period, which indicates there might be heterogeneity in the data set. By lumping together different countries at different times, Pool OLS camouflages the heterogeneity (individuality or uniqueness) that may exist among the countries.

**2. OLS with Dummy Variables for countries:**  = + +…… + \*log + \*log + \*log +

*Table 4: OLS Dummy summary*

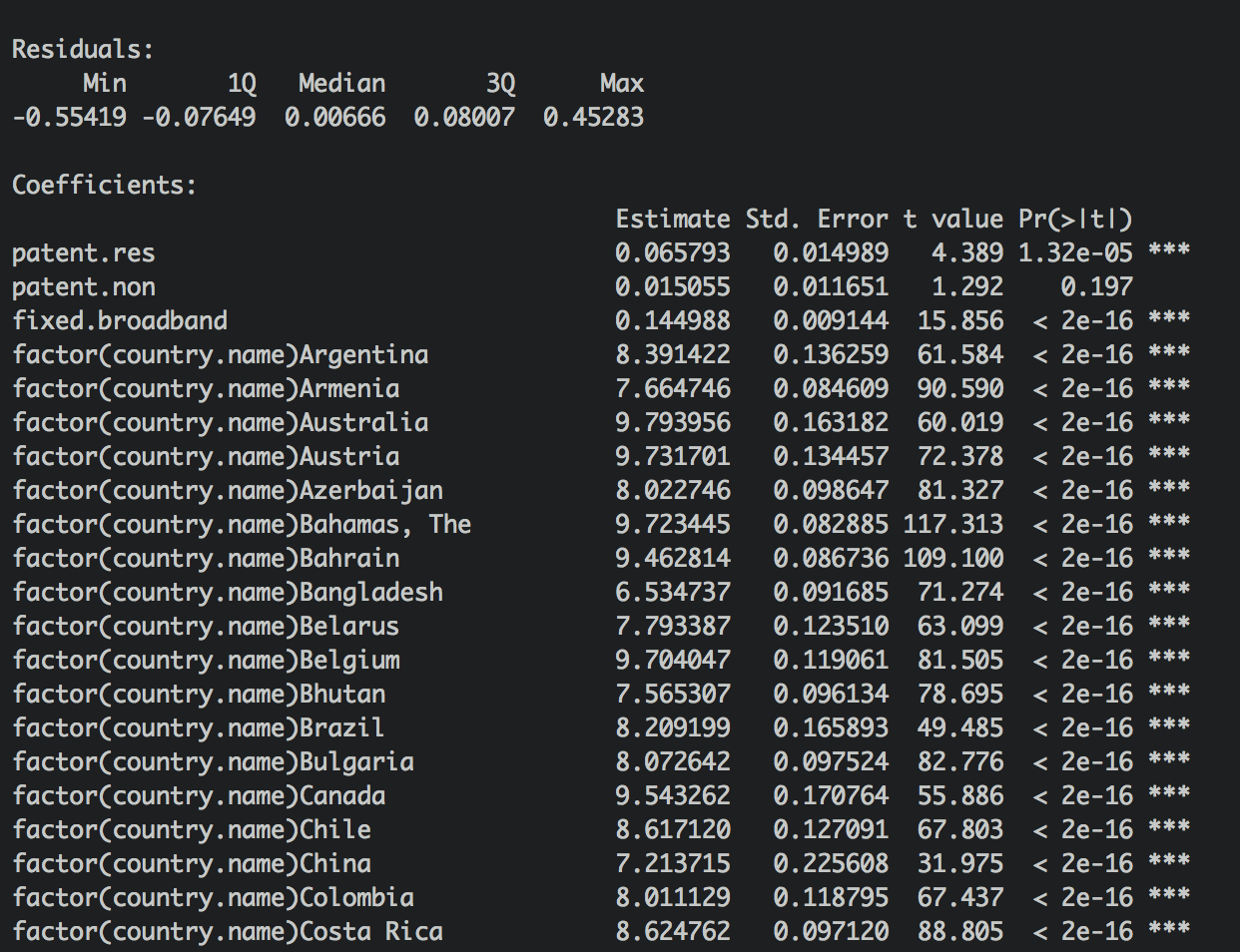
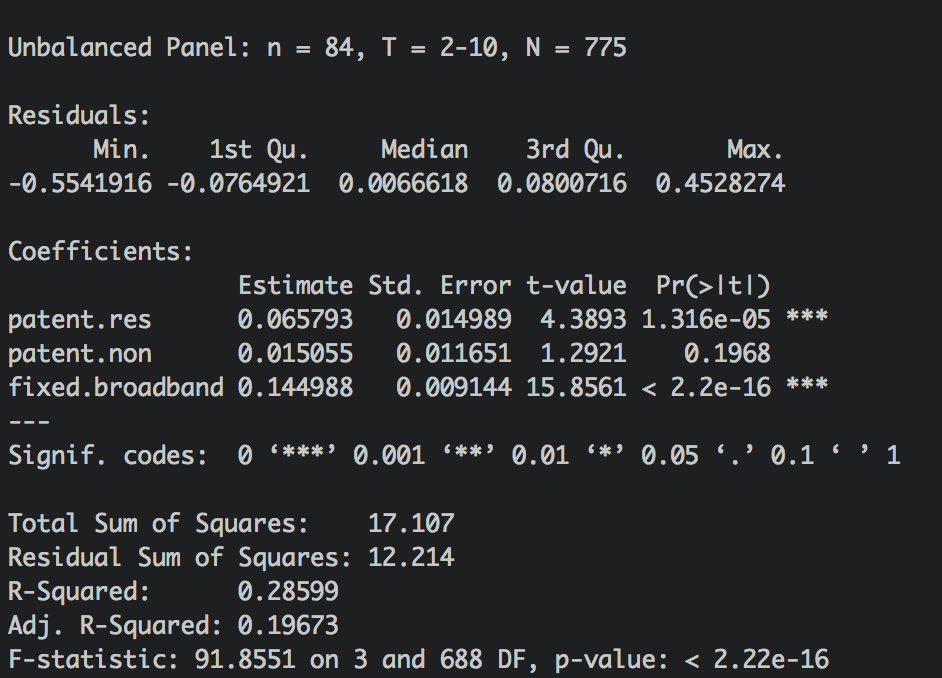


Table 4 shows that the coefficient of each country is significant, which affirms the fact that there is heterogeneity among countries. The question, therefore, is how we account for the unobservable, or heterogeneity effects so that we can obtain consistent and/or efficient estimates of the parameters of the variables of interest, which are Patent Applications and Internet Subscriptions. However, OLS Dummy model has several shortcomings. Firstly, as the OLS Dummy introduces too many dummy variables, it loses the degrees of freedom. Secondly, with many dummy variables in the model, there is always the possibility of multicollinearity.

**3. Fixed Effects:**  = \*log + \*log + \*log +

*Table 5: Fixed Effects (demeaned) results*

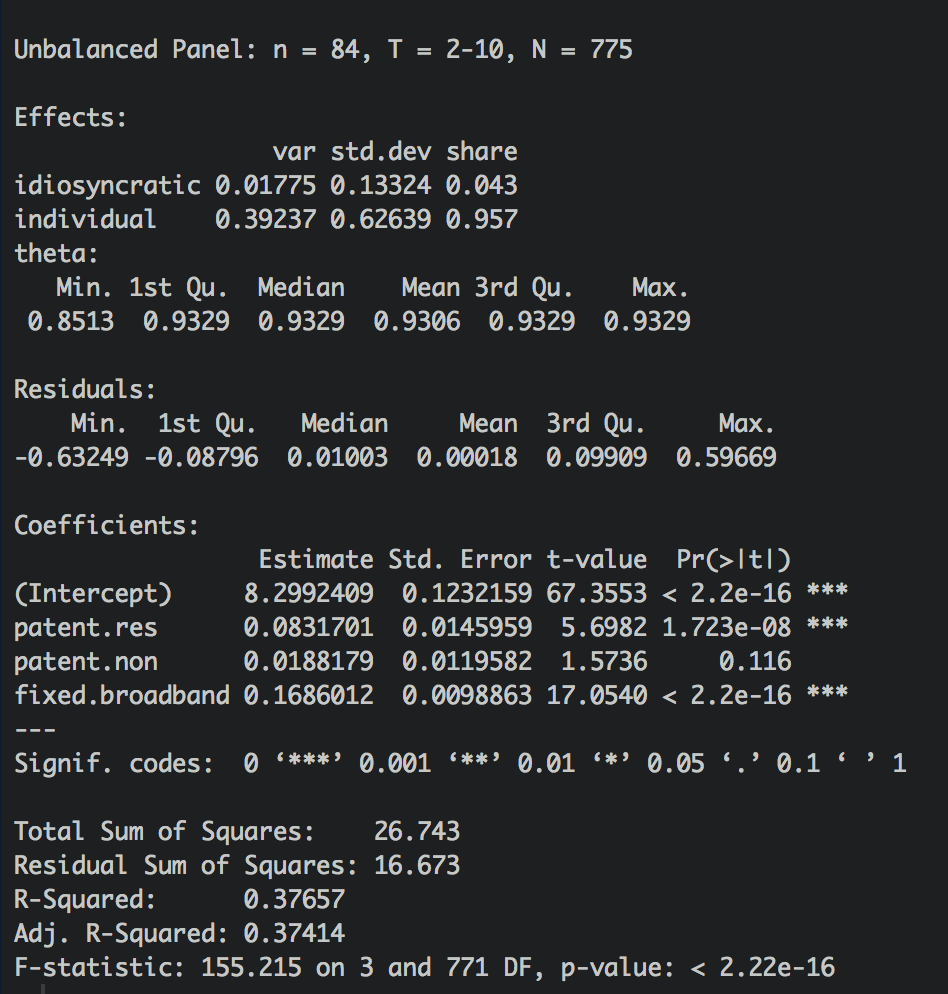


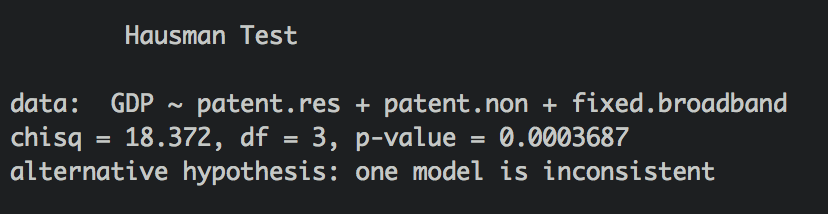
Based on Table 5, there is obviously a change in the coefficients of Patent Applications (Residents) compared to Pool OLS model. The coefficient of Patent Applications (Residents) now is statistically significant. This model takes the heterogeneity among 84 countries into account, not by the dummy variable method, but by eliminating it by differencing sample observations around their sample means. The benefit of this method is that it returns consistent estimates of the slope coefficients.

Based on the regression result, keeping every other variable constant, on average, when the number of Patent Applications (Residents) increases by 1%, GDP per capita increases by 0.0657%. Similarly, when the number of Fixed Broadband Subscriptions increases by 1%, GDP per capita increases by 0.1449%. Both of these coefficients are significant. However, the Patent Applications variable (Non-Residents) is not significant. The R-squared is 0.28, which is relatively low. However, for panel data, R-Squared is usually not high.

**4. Random Effects:**  = + \*log + \*log + \*log + ( = + )

*Table 6: Random Effects*



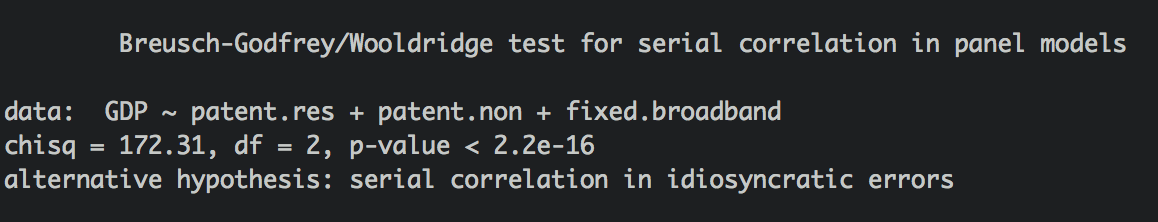
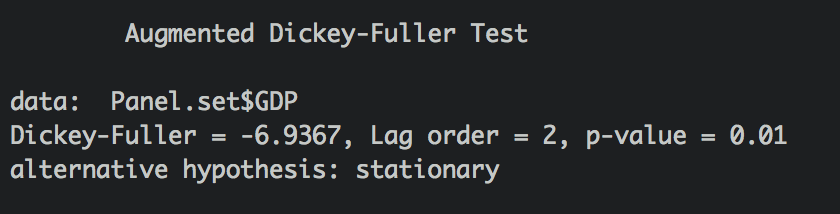


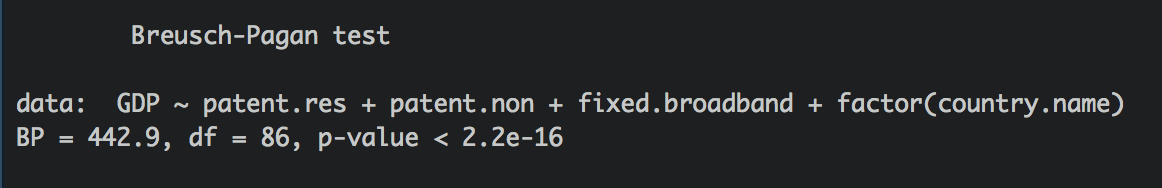
*Figure 4: Hausman Test*

Random Effects can be a good model if the add-on error term is not correlated with any of the explanatory variables included in the model. Since εi is a component

of wi t, it is possible that the error term is correlated with the explanatory variables. According to the Hausman test, which will tell us in a given application if the add-on error term in Random Effects is correlated with the explanatory variables, Random Effects model is inconsistent. Therefore, Fixed Effect would be the better model in this case.

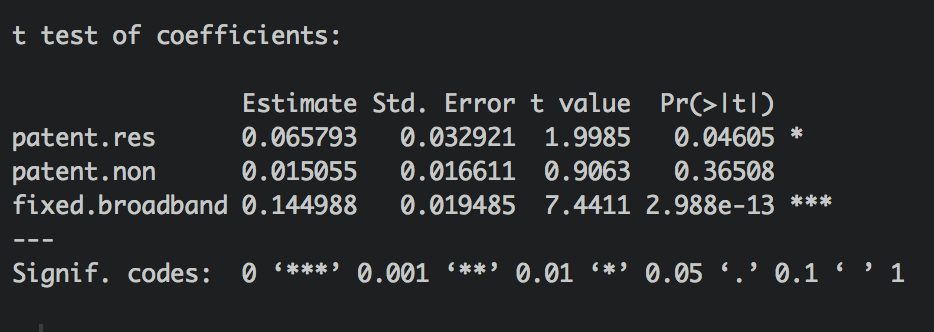
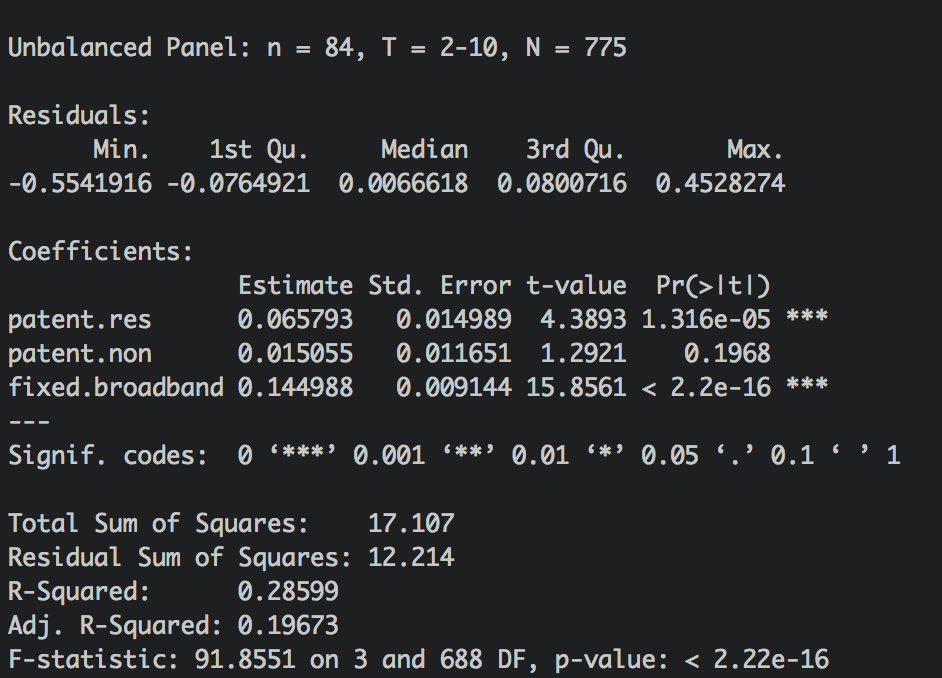
**5. Model diagnostic**





*Figure 5: Data Diagnostics*

As Panel data is the combination of Cross-sectional data and Time Series data, there are several assumptions when building economics model on this type of data. Firstly, The Dickey-Fuller test checks if data is stationary, p-value of this test is 0.01 so we can reject the null that the data is nonstationary. Secondly, the Breusch Godfrey/Wooldridge test points out that there is actually autocorrelation or serial correlation in the data. Fixed Effects model basically an extended version of linear regression, so it is important to make sure that the data is homoscedasticity. The Breusch-Pagan test check the homoscedasticity (linear model assumptions), it turns out the data is heteroscedasticity. Autocorrelation and heteroscedasticity are two things needed taking care of, so we can move on with the Fixed Effects Model (demeaned).

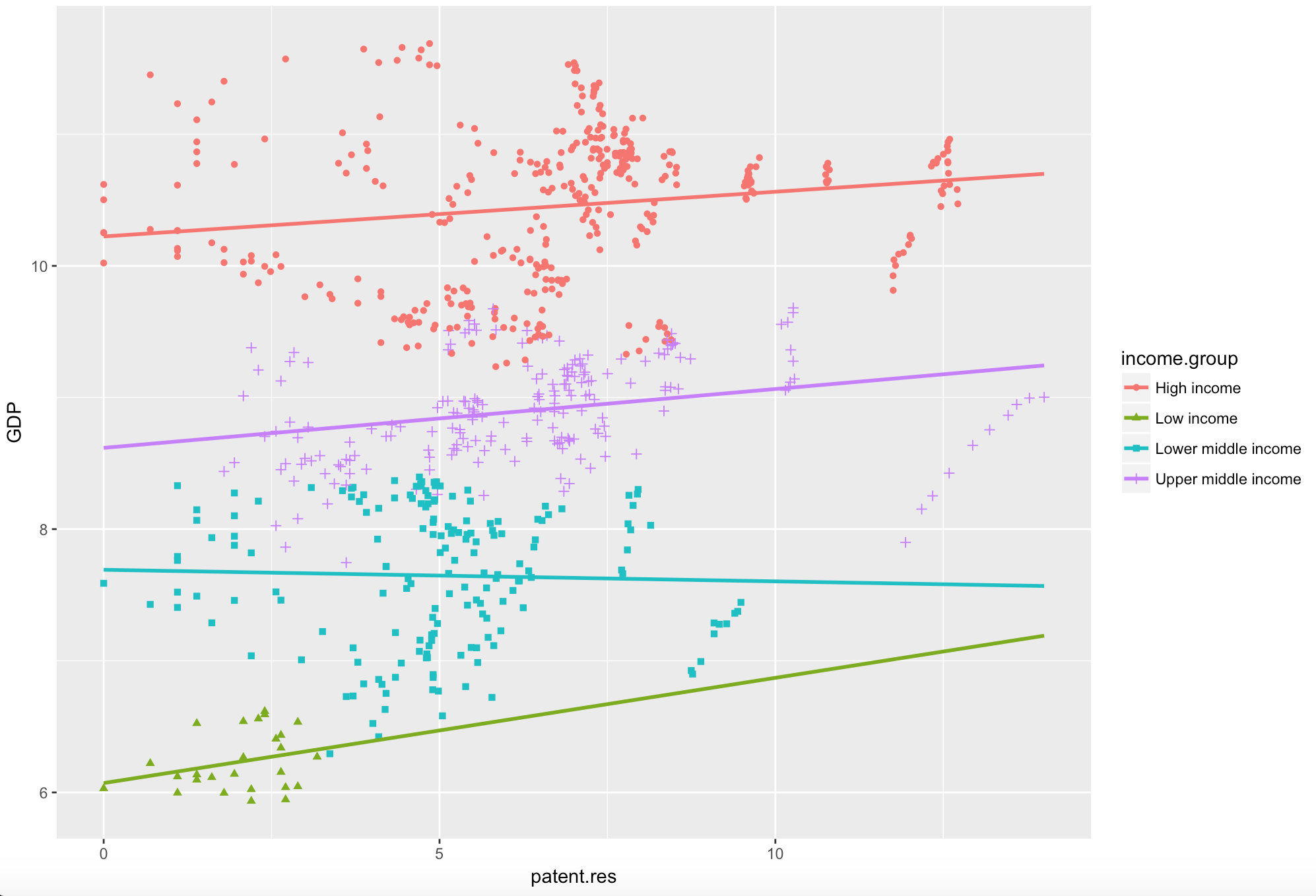


*Figure 6: Before and after Robust covariance matrix estimation (Sandwich estimator)*

We can still perform our regression analysis to correct the issue of Autocorrelation and heteroscedasticity so that our interval estimates, and hypothesis tests are valid. We do this by using Robust covariance matrix estimation (Sandwich estimator) or simply robust standard errors. It is worth noting here that the coefficients do not change before and after the process, but the standard errors and p-value change significantly.

**V Conclusion**

In conclusion, Fixed Effects (demeaned) model is the appropriate model for this panel data set. Based on the regression result, there is positive relationship between economic growth (GDP per capita) and Innovation (Patent Applications of Residents), and Technology (Fixed Broadband Subscriptions). However, the model itself and the model setting can be improved in certain ways. Firstly, this data is not complete due to missing values and limited resources. There are better ways to handle missing values other than dropping the observations such as imputing median or mean. Secondly, the independent variables selection can also be improved. Indicators such as R&D Expenditure and Number of Scientific Journals can be included in this model setting.



*Figure 7: Magnitude of innovation among income groups*

Thirdly, the magnitude of innovation among income groups may be different. Innovation may help to boost the economy in High Income countries but may not help in Low Income countries. Figure 7 indicates that the coefficients and intercepts of four Income Groups are in fact different. Lastly, more careful investigation in Patent Applications system should be done. Some countries have a “Patent Box” policy, which deducts tax on Intellectual Property such as Patents or License, so that big corporations in High Income countries file a huge number of patents in the “Patent Box” countries.

**Reference:**

Data retrieved from: [https://data.worldbank.org/](https://data.worldbank.org/indicator/IT.NET.BBND.P2)