Nam Nguyen

Instructor: Professor Esin Sile

Course name and number: Econometrics – ADEC7320.01

Date: 05/05/2018

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 different. 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 a 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

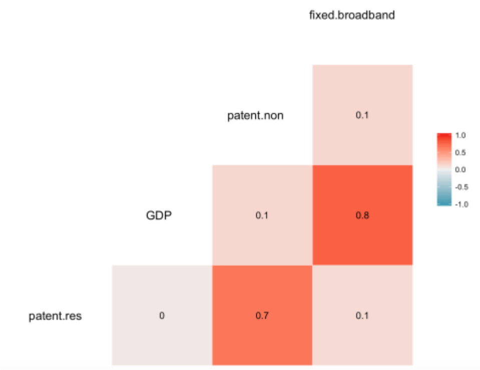
	country.code	country.name	year	patent.res	GDP	patent.non	fixed.broadband	income.group
1	ARG	Argentina	2007	937	7193.6176	4806	6.50484221	Upper middle income
2	ARG	Argentina	2008	801	8953.3593	4781	7.71311722	Upper middle income
3	ARG	Argentina	2009	640	8161.3070	4336	8.60558096	Upper middle income
4	ARG	Argentina	2010	552	10276.2605	4165	9.77161082	Upper middle income
5	ARG	Argentina	2011	688	12726.9084	4133	10.97543337	Upper middle income
6	ARG	Argentina	2012	735	12969.7071	4078	12.23036065	Upper middle income
7	ARG	Argentina	2013	643	12976.6364	4129	14.69447819	Upper middle income
8	ARG	Argentina	2014	509	12245.2565	4173	15.16805073	Upper middle income
9	ARG	Argentina	2015	546	13467.1024	3579	15.79039594	Upper middle income
10	ARG	Argentina	2016	884	12440.3210	2925	16.49367591	Upper middle income

III Data Summary

Table 2: Summary of the data set

```
> summary(data.final)
 country.code       country.name           year        patent.res            GDP
 Length:840         Length:840         Min.   :2007   Min.   :      1.0   Min.   :   377.9
 Class :character   Class :character   1st Qu.:2009   1st Qu.:    105.8   1st Qu.:  3876.7
 Mode  :character   Mode  :character   Median :2012   Median :    504.0   Median : 11772.5
                                       Mean   :2012   Mean   :  18132.7   Mean   : 21780.8
                                       3rd Qu.:2014   3rd Qu.:   1757.5   3rd Qu.: 38464.3
                                       Max.   :2016   Max.   :1204981.0   Max.   :119225.4
                                                      NA's   :58
   patent.non      fixed.broadband                 income.group
 Min.   :     1   Min.   : 0.00838   High income        :400
 1st Qu.:    53   1st Qu.: 4.02110   Low income         : 30
 Median :   267   Median :15.29842   Lower middle income:190
 Mean   :  9220   Mean   :16.48734   Upper middle income:220
 3rd Qu.:  4169   3rd Qu.:27.45871
 Max.   :310244   Max.   :45.13470
 NA's   :61       NA's   :4
>
```

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.



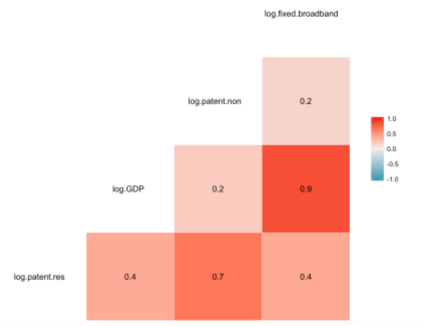


Figure 1: Correlation plot on raw data *Figure 2: Correlation plot (log transformation)*

Based on Figure 1, there is 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 the log of GDP per capita may help smoothen the data. Moreover, log transformation for other variables also lends a hand to model interpretation as the coefficients are simply elasticity. Figure 2 shows that there is a 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:** $\log(\mathrm{PC.GDP})_{it} = \beta_0 + \beta_1 * \log(\mathrm{patent.res})_{it} + \beta_2 * \log(\mathrm{patent.non})_{it} + \beta_3 * \log(\mathrm{fixed.broadband})_{it} + \mu_{it}$

Table 3: OLS summary

```
> summary(OLS1)

Call:
lm(formula = GDP ~ patent.res + patent.non + fixed.broadband,
    data = data.final1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.99860 -0.40511 -0.04475  0.49880  2.44990

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.5661599  0.0686332 110.241  < 2e-16 ***
patent.res      -0.0009808  0.0139315  -0.070  0.94389
patent.non       0.0354537  0.0123885   2.862  0.00433 **
fixed.broadband  0.7006473  0.0167998  41.706  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7004 on 771 degrees of freedom
  (65 observations deleted due to missingness)
Multiple R-squared:  0.7394,    Adjusted R-squared:  0.7384
F-statistic: 729.2 on 3 and 771 DF,  p-value: < 2.2e-16
```

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 a group of countries or years of time period. Therefore, this model is not only biased but also inconsistent.

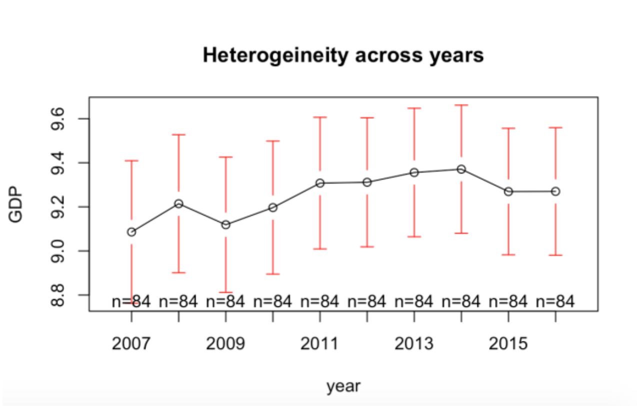


Figure 3: Heterogeneity across years of GDP per capita

Looking at Figure 3, there is a lot of volatility and fluctuation in GDP per capita throughout the 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: $\log(\mathrm{PC.GDP})_{it} = \alpha_1 + \alpha_2 C_{2i} + \ldots\ldots$ $\alpha_{84} C_{84i} + \beta_1 * \log(\mathrm{patent.res})_{it} + \beta_2 * \log(\mathrm{patent.non})_{it} + \beta_3 * \log(\mathrm{fixed.broadband})_{it} + \mu_{it}$

Table 4: OLS Dummy summary

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.55419 -0.07649  0.00666  0.08007  0.45283

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
patent.res                          0.065793   0.014989   4.389 1.32e-05 ***
patent.non                          0.015055   0.011651   1.292    0.197
fixed.broadband                     0.144988   0.009144  15.856  < 2e-16 ***
factor(country.name)Argentina       8.391422   0.136259  61.584  < 2e-16 ***
factor(country.name)Armenia         7.664746   0.084609  90.590  < 2e-16 ***
factor(country.name)Australia       9.793956   0.163182  60.019  < 2e-16 ***
factor(country.name)Austria         9.731701   0.134457  72.378  < 2e-16 ***
factor(country.name)Azerbaijan      8.022746   0.098647  81.327  < 2e-16 ***
factor(country.name)Bahamas, The    9.723445   0.082885 117.313  < 2e-16 ***
factor(country.name)Bahrain         9.462814   0.086736 109.100  < 2e-16 ***
```

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: $\log[\mathrm{PC.GDP}]_{it} = \beta_1 * \log[\mathrm{patent.res}]_{it} + \beta_2 * \log[\mathrm{patent.non}]_{it} + \beta_3 * \log[\mathrm{fixed.broadband}]_{it} + \mu_{it}$

Table 5: Fixed Effects (demeaned) results

```
Unbalanced Panel: n = 84, T = 2-10, N = 775

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.5541916 -0.0764921  0.0066618  0.0800716  0.4528274

Coefficients:
                Estimate Std. Error t-value  Pr(>|t|)
patent.res      0.065793   0.014989  4.3893 1.316e-05 ***
patent.non      0.015055   0.011651  1.2921    0.1968
fixed.broadband 0.144988   0.009144 15.8561 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    17.107
Residual Sum of Squares: 12.214
R-Squared:      0.28599
Adj. R-Squared: 0.19673
F-statistic: 91.8551 on 3 and 688 DF, p-value: < 2.22e-16
```

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: $\log(\mathrm{PC.GDP})_{it} = \beta_0 + \beta_1 * \log(\mathrm{patent.res})_{it} + \beta_2 * \log(\mathrm{patent.non})_{it} + \beta_3 * \log(\mathrm{fixed.broadband})_{it} + \omega_{it} \quad (\omega_{it} = \mu_{it} + \varepsilon_i)$

Table 6: Random Effects

```
Unbalanced Panel: n = 84, T = 2-10, N = 775

Effects:
                  var std.dev share
idiosyncratic 0.01775 0.13324 0.043
individual    0.39237 0.62639 0.957
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.8513  0.9329  0.9329  0.9306  0.9329  0.9329

Residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.63249 -0.08796  0.01003  0.00018  0.09909  0.59669

Coefficients:
                 Estimate Std. Error t-value  Pr(>|t|)
(Intercept)     8.2992409  0.1232159 67.3553 < 2.2e-16 ***
patent.res      0.0831701  0.0145959  5.6982 1.723e-08 ***
patent.non      0.0188179  0.0119582  1.5736     0.116
fixed.broadband 0.1686012  0.0098863 17.0540 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    26.743
Residual Sum of Squares: 16.673
R-Squared:      0.37657
Adj. R-Squared: 0.37414
F-statistic: 155.215 on 3 and 771 DF, p-value: < 2.22e-16
```

```
        Hausman Test

data:  GDP ~ patent.res + patent.non + fixed.broadband
chisq = 18.372, df = 3, p-value = 0.0003687
alternative hypothesis: one model is inconsistent
```

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

```
        Augmented Dickey-Fuller Test

data:  Panel.set$GDP
Dickey-Fuller = -6.9367, Lag order = 2, p-value = 0.01
alternative hypothesis: stationary
```

```
        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  GDP ~ patent.res + patent.non + fixed.broadband
chisq = 172.31, df = 2, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

```
        Breusch-Pagan test

data:  GDP ~ patent.res + patent.non + fixed.broadband + factor(country.name)
BP = 442.9, df = 86, p-value < 2.2e-16
```

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 checks 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).

```
Coefficients:
                Estimate Std. Error t-value  Pr(>|t|)
patent.res      0.065793   0.014989  4.3893 1.316e-05 ***
patent.non      0.015055   0.011651  1.2921    0.1968
fixed.broadband 0.144988   0.009144 15.8561 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
patent.res      0.065793   0.032921  1.9985   0.04605 *
patent.non      0.015055   0.016611  0.9063   0.36508
fixed.broadband 0.144988   0.019485  7.4411 2.988e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

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 a 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.

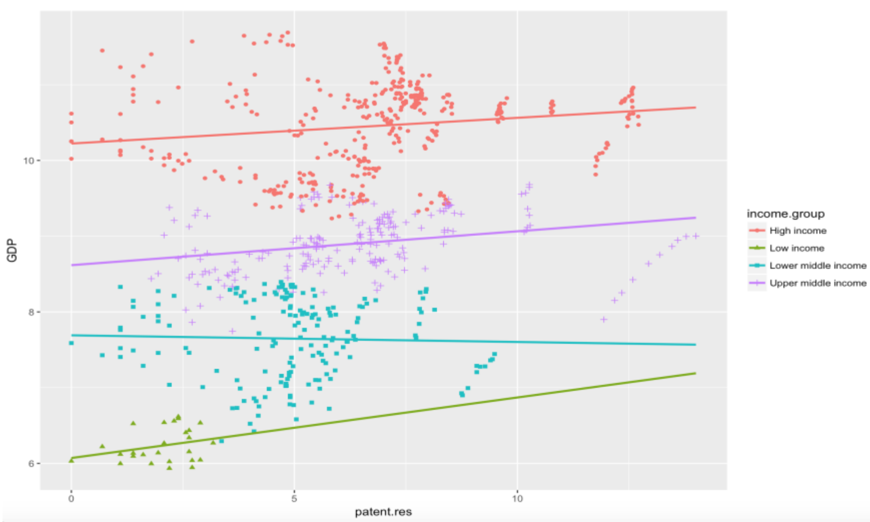


Figure 7: Magnitude of innovation among income groups

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/