**By: Yoke Hong Si**

**1005815806**

**Eco 375: Applied Econometrics**

**University of Toronto**

**Department of Economics**

**Assignment 2**

**April 2021**

Abstract:

This paper explores the effect of vaccination on COVID-19 infection in Israel using simple and multi linear regression models in our analysis. From our results, we were unable to determine a statistically significant effect in our simple linear regression analysis. Upon controlling for several confounding variables, we were able to determine a significant effect in our multilinear regression analysis. However, there still exists omitted variables that would lead to bias in our results thus, we cannot conclude that there is purely a causal effect of vaccination on COVID-19 infection.

1. **Introduction**

In this paper, we will be examining whether if citizens demand more democracy and political freedom as their incomes grow. We will be utilising econometric methods such as simple linear regression and multiple linear regression models to analyse this relationship empirically using the data provided.

1. **The Context and Data**

It is known that the relationship between income per capita and political freedom yields a positive correlation as showed in Figure 1. However, we are curious as to whether if the correlation is explained by a causal connection, whereby a higher income per capita indicates that a country becomes more democratic. Thus, we will analyse this relationship through the panel dataset from the paper, “Income and Democracy” written by Acemoglu, et al (2008). To introduce a brief discussion of the summary statistic of our variables, we can look at the size of estimated effects in context. The main democracy index is measured using the Augmented Freedom House Political Rights Index which ranges between 0 and 1, whereby a number closer to 1 would indicate a more democratic country. In other words, a country would have political rights come closest to the ideals suggested by a checklist of questions i.e., whether there are free and fair elections etc. From Table 1, we determined the average of the democracy index (Augmented Freedom House Political Rights Index) between countries to be 0.56830. This indicates that on average, most countries are leaning towards democracy since the average index happened to be greater than 0.5. On the other hand, another democracy index, known as the Polity IV index, provides information for all independent countries starting in 1800 and has an average of 0.55627. As shown in figure 2, summary statistics such as education tend to increase from low income to high income countries. On the other hand, colonial countries tend to have lower education levels compared to non-colonial countries as shown in figure 3.

**Regression analysis**

1. **Base Specifications**

**3.1 Base Specification 1.**

The estimated simple linear regression of the Augmented Freedom House Political Rights Index (AFHPRI democracy index) regressed on the one period lag of Log real GDP per capita (PWT) with robust standard errors is as follows:

(0.0624196) (0.0072955)

*n* = 1023, *R2* = 0.4034

Firstly, if L1.lrgdpch (lagged log real GDP per capita) = 0, the predicted fhpolrigaug (democracy index) will be the intercept which is 7.323544. If we write the predicted change in fhpolrigaug as a function of change in L1.lrgdpch:

, it means that if the L1.lrgdpch increases by one unit, , then fhpolrigaug is predicted to change by about . In practical terms, this implies that the democracy index increases by 0.2254815 for every additional unit increase in one period lag of log real GDP per capita.

The t-statistic gives a value of 30.91 which falls outside the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy is statistically significant. The 95% confidence interval for this estimate is computed to be [0.2111657, 0.2397974]. In this case, zero is not part of this interval so we reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between 0.2111657 and 0.2397974.

**3.2 Base Specification 2.**

In this specification, we estimate an SLR of the AFHPRI regressed on the one period lag of Log real GDP per capita with standard errors clustered at the country level. In this regression, we obtained identical point estimates for both the coefficient of the log real GDP per capita, and the intercept as Base Specification 1. However, the 95% confidence level is determined to be [0.2008087, 0.2501544] which differs since the robust standard errors are different. Similarly, zero is not part of this interval so we reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between 0.2111657 and 0.2397974 for clustered standard errors. Clustered standard errors allow for unrestricted forms of serial correlation and heteroskedasticity in panel data. Since we expect the error term of the regression to be correlated across countries or over time, it is important to control for this factor using clustered standard errors. Therefore, the regression would suffer from omitted variable bias since time-invariant confounders could lurk in the error term that is not accounted for. An example of a time invariant factor that varies with democracy and income systematically which may bias the results is as follows. Consider a comparison between the United States and North Korea where the US is both richer and more democratic thus, a simple cross-country comparison that does not control for fixed country effects would suggest that higher per capita GDP causes democracy. Therefore, we should account for the fixed country effects in order to investigate “within-country variation” instead for instance, if the US is more likely to be relatively democratic as it becomes relatively richer.

**3.3 Base Specification 3.**

Here, we investigate a regression of the AFHPRI on the one period lag of log real GDP per capita with the inclusion of country fixed effects (code\_numeric) in the regression. The estimated coefficient on the lag of lrgdpch is determined to be 0.0550704 which is a lot smaller than in specification (2). It means that if the L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0550704. In practical terms, this implies that the democracy index increases by 0.0550704 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country fixed effects. The point estimate has changed from specification (2) and it could be interpreted as a small effect by using a t-test. The t-statistic gives a value of 1.67 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with country fixed effects is not statistically significant. The 95% confidence interval for this estimate is computed to be [-0.0100479, 0.1201887]. In this case, zero part of this interval so we do not reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between -0.0100479 and 0.1201887. However, the regression would still suffer from omitted variable bias since time varying confounders could be present, affecting the point estimate. An example of a time varying factor that varies by country which may also vary with democracy and income systematically is as follows. Consider a comparison between the United Kingdom and Vietnam during the Vietnam War. Thus, we would expect the UK to be both richer and more democratic during this war period which would invalidate any causal interpretation from the regression.

**3.4 Base Specification 4**

In the fourth base specification, we regress the AFHPRI on the one period lag of log real GDP per capita with the inclusion of both country fixed effects (code\_numeric) and year fixed effects (year\_numeric) in the regression. The estimated coefficient on the lag of lrgdpch is determined to be 0.0587276 which is slightly larger than 0.0550704 in specification (3) but still relatively small in terms of overall magnitude. The coefficient implies that if the L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0587276. In practical terms, this implies that the democracy index increases by 0.0587276 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country and year fixed effects. The point estimate has changed from specification (2) and it could be interpreted as a small effect by using a t-test. The t-statistic gives a value of 1.29 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with country and year fixed effects is not statistically significant. The 95% confidence interval for this estimate is computed to be [-0.0310453, 0.1485006]. In this case, zero part of this interval so we do not reject the null where there is no effect of income on democracy. In other words, we are also 95% confident that the effect of one period lag of log real GDP per capita on the democracy index will fall between -0.0310453 and 0.1485006.

By testing the exclusion of the year fixed effects, we obtain an F-statistic of 106.39854 and p-value of 1.313e-23.

**3.5 Base Specification 5**

In this base specification, we first regressed the democracy index on the one period lag income per capita, including country and year fixed effects as well as lagged demographic controls (age, education and log population). We then defined a subsample using this regression. Finally, we ran specification (4) on the subsample generated and obtain the following regression model.

(0.4895312) (0.053769)

*n* = 685, *R2* = 0.7259

**3.6 Base Specification 6**

For our last base specification, we first regressed the democracy index on the one period lag income per capita, including country and year fixed effects as well as lagged demographic controls (age, education and log population). The estimated coefficient on the lag of lrgdpch is determined to be 0.0159635 which is smaller than 0.0587276 in specification (4) and still relatively small in terms of overall magnitude. The coefficient implies that if the L1.lrgdpch increases by one unit, then fhpolrigaug is predicted to change by about 0.0159635. In practical terms, this implies that the democracy index increases by 0.0159635 for every additional unit increase in one period lag of log real GDP per capita, whilst controlling for country and year fixed effects as well as demographics. The point estimate has changed from specification (4). This could be due to the change of sample between specifications (6) and (4) where specification (6) has 685 observations whereas (4) have 1023 observations. By conducting a t-test, we can interpret that the coefficient is of a small value. The t-statistic gives a value of 0.30 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of income on democracy with the controls in specification (6) is not statistically significant.

By testing the exclusion of the year fixed effects, we obtain an F-statistic of 50.928385 and p-value of 2.889e-12. Testing the exclusion of the age variables, we obtain an F-statistic of 4.3295226 and p-value of 0.00185436. Lastly, testing the exclusion of the year fixed effects, we obtain an F-statistic of 42.127508 and p-value of 1.199e-42.

The result is also statistically insignificant as its p-value for the t-statistic of the income variable failed to be rejected at the 10% level. Moreover, education, log population, and age groups are mostly statistically insignificant except for the 30-45 age group, which is significant at the 10% level.[[1]](#footnote-2) The coefficient for the 30-45 group is -2.502, implying that for every 1% increase in the 30-45 group composition as part of a country’s total population, the democracy index decreases by -0.02502. The F-statistics for the year (fixed effect), age, and demographic controls are all statistically significant since the null hypotheses that the means between the restricted and unrestricted models are equal are all rejected at 1% level (Table 2), implying that all those controls can provide a better fit for the regression model.

We are uncertain if SLR.1 holds since according to the Simple Linear Regression model, the equation follows SLR.1 in the form of . However, due to the theoretical mechanism, SLR.1 does not hold since there would be a nonlinear effect of vaccination rate on the COVID-19 cases due to the positive externality of the vaccine. Eventually, herd immunity will be achieved and there will be zero effect of further vaccination on case rates. SLR.2 does not hold since the dataset is obtained from different age groups and weeks in Israel for a particular time period. Therefore, since there is correlation across time, there would be a positive correlation across observations which causes an underestimation in our standard errors as we are not accounting for the correlation in the data. SLR.3 holds since from the dataset provided, we know that the values of lagvacc\_per are not all the same value thus, there is variation amongst the values of lagvacc\_per. SLR.4 indicates that we assume . However, given the current context, this will not hold. For example, occupation (that is part of u) correlates with lagvacc\_per (x) since front-line workers (medical staffs) will tend to have a higher rate of vaccination compared to ordinary people. SLR.5 indicates that we assume However, given the current context, this will not hold. For instance, consider where we look at the variability of age groups given the vaccination rate. The older age group tend to have a higher vaccination rate since they are given priority to the vaccine thus, this leads to a low dispersion across the age groups. On the other hand, people with lower vaccination rates would include people such as infants or middle-aged adults, leading to a high variability across the age groups. The graph displaying heteroscedasticity of COVID-19 Cases on Vaccination can be found in Figure 2.

1. **Extension**
2. **Limitations of results**

Despite controlling for year-week and age group in the previous section, our MLR econometric models in Table 1 may be underspecified as it cannot capture all omitted variables that may lurk in the residuals. This could include variables such as income that could affect the vaccination rates as we would expect people with higher income to be able to afford the vaccine compared to lower income individuals. Thus, omitted variables could underestimate or overestimate the coefficient of our variable of interest, leading to omitted variable bias. The bias could take forms of positive or negative bias and it depends on the coefficient of the omitted variable as well as the sign of the correlation between the vaccination rate and omitted variable. Due to this bias, we are unable to interpret the results of COVID cases as purely a causal effect by vaccination.

Additionally, there are several other factors that would affect the validity of our regression results. One being that we are unable to conduct randomised sampling to obtain the data. Due to the lack of random sampling, it would lead to open backdoor paths which fails the conditional independence assumption leading to selection bias in the experiment. Furthermore, due to the nature of the benefits from vaccination, an accurate econometric model representing the effect of vaccination rate on COVID-19 cases will be nonlinear. These will in turn lead to biasness of the OLS estimators for the population parameters.

1. **Conclusion**

In conclusion, through SLR, we are unable to determine a statistically significant effect of vaccination on COVID-19 infections. However, with MLR, we controlled the year-week and age groups thus, a statistically significant effect was achieved, illustrating a decrease in infections as we increase vaccination rate. Ultimately, correlation does not prove causation thus, we are unable to determine a causal effect of the decrease in infections due to the vaccine since there could be bias from any omitted variables.

**References:**

[Roser, Max, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell (2020) - "Coronavirus Pandemic (COVID-19)". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

Israeli Ministry of Health. (2020) REAL-WORLD EPIDEMIOLOGICAL EVIDENCE COLLABORATION AGREEMENT. Accessed February 3, 2021. https://govextra.gov.il/media/30806/11221-moh-pfizer-collaboration-agreement-redacted.pdf. ]

Table 1: Table of descriptive statistic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Observations** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| Year of observation | 11 | 1975 | 15.8 | 1950 | 2000 |
| Augmented Freedom House Political Rights Index | 1,194 | 0.56830 | 0.35775 | 0 | 1 |
| Log real GDP per capita (PWT) | 1,199 | 8.1650 | 1.0372 | 5.77393 | 10.692 |
| Nominal savings rate | 1,206 | 0.16519 | 0.13644 | -0.76600 | 0.73962 |
| % labor share of gross value added | 504 | 0.37040 | 0.13089 | .07628 | 0.73844 |
| Percent population age 0-15 | 1,551 | 0.36340 | 0.09475 | 0.14292 | 0.51610 |
| Percent population age 15-30 | 1,551 | 0.25460 | 0.02565 | 0.16855 | 0.34658 |
| Percent population age 30-45 | 1,551 | 0.17534 | 0.02983 | 0.08966 | 0.29793 |
| Percent population age 45-60 | 1,551 | 0.11838 | 0.03744 | 0.05601 | 0.22149 |
| Percent population age 60- | 1,551 | 0.08829 | 0.04930 | 0.03165 | 0.24063 |
| Median age in population | 1,529 | 23.121 | 6.6637 | 14.4 | 41.3 |
| Average schooling years | 722 | 4.4584 | 2.8739 | 0.042 | 12.179 |
| Log (total population in thousands) | 1,308 | 8.6819 | 1.9073 | 3.7136 | 14.049 |
| Polity IV index | 1,092 | 0.55627 | 0.37931 | 0 | 1 |
| Dummy for Soviet Block, including iron curtain | 1,617 | 0.14966 | 0.35685 | 0 | 1 |
| Dummy former colony vs non-colony | 1,617 | 0.70068 | 0.45810 | 0 | 1 |

Table 2: Base Specification Regression Analysis Results Summary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
| Lag of log real GDP pc | 0.225\*\*\* | 0.225\*\*\* | 0.0551\* | 0.0587 | -0.0228 | 0.0160 |
|  | (0.0073) | (0.0125) | (0.0329) | (0.0454) | (0.0538) | (0.0532) |
|  | [0.2112, 0.2398] | [0.2008, 0.2502] | [-0.0100, 0.1202] | [-0.0310, 0.1485] | [-0.1296, 0.0840] | [-0.0896, 0.1216] |
| Age: |  |  |  |  |  |  |
| Age group: 15-30 |  |  |  |  |  | -0.509 |
|  |  |  |  |  |  | (0.678) |
| Age group: 30-45 |  |  |  |  |  | -2.618 |
|  |  |  |  |  |  | (0.949) |
| Age group: 45-60 |  |  |  |  |  | 0.452 |
|  |  |  |  |  |  | (1.413) |
| Age group: 60- |  |  |  |  |  | 0.848 |
|  |  |  |  |  |  | (1.390) |
| Education |  |  |  |  | -0.00508 | -0.00351 |
|  |  |  |  |  | (0.0244) | (0.0237) |
| Log population |  |  |  |  | -0.170 | -0.104 |
|  |  |  |  |  | (0.121) | (0.133) |
| F-statistics & P-values: |  |  |  |  |  |  |
| Year fixed Effects |  |  |  | 106.4\*\*\* | 60.99\*\*\* | 50.93\*\*\* |
|  |  |  |  | (1.3e-23) | (2.7e-14) | (2.9e-12) |
| Age Controls |  |  |  |  |  | 4.33\*\*\* |
|  |  |  |  |  |  | (.00185) |
| Demographic Controls |  |  |  |  |  | 42.13\*\*\* |
|  |  |  |  |  |  | (1.2e-42) |
|  |  |  |  |  |  |  |
| Year Fixed Effects | No | No | No | Yes | Yes | Yes |
| Country Fixed Effects | No | No | Yes | Yes | Yes | Yes |
| Clustered Standard Err. | No | Yes | Yes | Yes | Yes | Yes |
| Demographic Sample | No | No | No | No | Yes | Yes |
| Number of Obs. | 1023 | 1023 | 1023 | 1023 | 685 | 685 |
| Number of Countries | 147 | 147 | 147 | 147 | 94 | 94 |
| R-Squared | 0.4034 | 0.4034 | 0.7272 | 0.7570 | 0.6973 | 0.7354 |

Notes: The dependent variable is the Augmented Freedom House Political Rights Index. All variables with estimates in this table are one period lag, where the independent variables such as age and education are from one interval, which is 5 years, after that of the dependent variable (Freedom House democracy index). All estimates are computed and replicated using Stata. The dataset between 1950 and 2000 was collected from Acemoglu et al.’s 2008 “Income and Democracy” paper dataset on March 28, 2021. Standard errors are in parentheses. The standard errors of some of the P-values are in exponential notations (ex, 1.3e-23 = 1.3 x 10-23). Confidence Intervals are in square brackets.

\*\*\*Significant at the 1 percent level. \*\*Significant at the 5 percent level. \*Significant at the 10 percent level.

Other tables and figures if needed.

Figure 1

Figure 2

1. Please refer to Table 2 for the full list of estimates reported. [↑](#footnote-ref-2)