**[title]**

**By: Yoke Hong Si**

**1005815806**

**Eco 375: Applied Econometrics**

**University of Toronto**

**Department of Economics**

**Assignment 1**

**February 2021**

Abstract:

[concise abstract; best practice: write this last]

1. **Introduction**

[write this second last; Chapter 19, section 19-5 “Writing an Empirical Paper” in the course textbook has some general ideas about style guidelines in economics writing]

1. **The Context and Data**

[briefly describe the context and data, with reference to, at least, a table of descriptive statistic]

Israel is currently a leading country for COVID-19 vaccination. The vaccination campaign allows Israel to collect medical data from their population of 9 million. The data that is collected will be the data that we will analyse in this empirical paper.

Table 1 contains a summary statistic of our variables in order to help analyze the effect of vaccination on COVID-19 infection in Israel. The description of the variables can be found after the references page.

From the table, there is a total of 198 observations of COVID-19 cases occurring on different dates. Additionally, the mean weekly COVID-19 cases is determined to be around 2598.

From Table 1, we determined the mean weekly COVID-19 cases to be around 2598. Hence, a 3% decrease in COVID-19 cases will reflect a decrease of around 78 cases. This is around

1. **Regression analysis**
   1. **Simple Linear Regression**

[Background reading includes Chapter 2-6 and Lecture 4 and 5. You will want to pay special attention to the SLR assumptions in Chapter 2]

(0.0051737) (.0818457)

*n* = 198, *R2* = 0.0039

Firstly, if lagvacc\_per = 0, the predicted case\_log will be the intercept which is 7.323544. If we write the predicted change in case\_log as a function of change in lagvacc\_per:

. This means that if the lagvacc\_per increases by one percentage point, , then case\_log is predicted to change by about 7.323544. In practical terms, this implies that the number of weekly COVID-19 cases (per million people) increased by 0.45166% for every additional unit increase in lagvacc\_per.

The t-statistic gives a value of 0.87 which falls within the interval of [-1.96, 1.96] that corresponds to a significance level of 0.05. Therefore, the effect of vaccination on COVID-19 infections is not statistically significant. Hence, in our sample, we determined the value of lagvacc\_per to be 0.0045166 however, the value of lagvacc\_per is not sufficiently great in order to conclude a statistically significant effect at the population level. Thus, we fail to reject the null hypothesis.

SLR.1 does not hold due to the theoretical mechanism of the external benefits from vaccination. There would be a nonlinear effect of vaccination on the COVID-19 cases due to the positive externality of the vaccine and 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. People with higher vaccination rates tend to be given priority to the vaccination since it consists of the elderly population part of the old age group, leading to a low dispersion across the age groups. On the other hand, people with lower vaccination rates would include ordinary 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 shown in Figure 1.

* 1. **Multiple linear regression**

[Here, pay special attention to the MLR assumptions in Chapter 3, the discussion of omitted variable bias in section 3-3, and the discussion of functional form in section 6-2a and 6-2b (which is also appended to this overview). The regression specification in textbook example 6.2 is particularly close to our specification in the assignment. It is worth a look.]

From table 2 column 1, we determine the coefficient of the lagvacc\_per to be .0045166. This implies that the number of weekly COVID-19 cases (per million people) increased by 0.45166% for every additional unit increase in lagvacc\_per. Similarly, in column 2, the number of weekly COVID-19 cases decreased by 0.57974% for every additional unit increase in lagvacc\_per when controlling for a set of week and age group dummy variables. In column 3, the number of weekly COVID-19 cases decreased by 1.34144% and increased by 0.00967% for every additional unit increase in lagvacc\_per and lagvacc\_per2 respectively when controlling for a set of week and age group dummy variables. Lastly, in column 4, the number of weekly COVID-19 cases decreased by 45.50193%, 54.30955% and 67.61747% for every additional unit increase in lagvacc\_0\_10, lagvacc\_10\_20 and lagvacc\_20\_100 respectively when controlling for a set of week and age group dummy variables.

By adding a set of control variables, the statistical significance of the coefficient of lagvacc\_per changes from not being statistically significant to being significant at the 1 percent level. Therefore, in practical means, we can infer that there is a statistically significant difference in the number of weekly COVID-19 cases under the effect of vaccination. The data in table 1 indicates that the number of weekly COVID-19 cases decreased by 0.57974% for every additional unit increase in lagvacc\_per when controlling for a set of week and age group dummy variables. Thus, we reject the null hypothesis.

As we move from specification (1) to (2), we are effectively taking the year-week and age group out of the error term and including it explicitly in our equation. Since our coefficient of interest is the coefficient of lagvarr\_per, we would want to fix all other factors that affects it. Year-week and age group are examples of factors that can affect the result of the coefficient because we would expect a positive correlation between the factors and the vaccination rate: vaccination rate tends to be higher as the year-week progresses as well as for elderly individuals. Now that year-week and age group are included in the equation, we can measure the effect of vaccination rate on COVID-19 cases holding the year-week and age group fixed. Hence, the coefficient of lagvarr\_per in (2) yields a more accurate result.

• Discuss what happens when moving to specification (1) to specification (2). For example, why

would including controls for year-week and age group impact results? Is there a confounding

time trend in both infection and vaccination? (hint: see Figure 2 attached). Do different age

groups have different chances of COVID-19 infection? Different rates of vaccination? These

types of examples can be used to anchor the general discussion.

• Is there a non-linear effect of vaccination? From a theoretical perspective? And is there evidence

of a non-linear effect from the estimation? See Section 6-2b in the textbook (also appended to

this overview),

due to the theoretical mechanism of the external benefits from vaccination. There would be a nonlinear effect of vaccination on the COVID-19 cases due to the positive externality of the vaccine and eventually, herd immunity will be achieved and there will be zero effect of further vaccination on case rates.

Yes. If we look at specification (3), we determined the coefficient of lagvacc\_per and lagvacc\_per2 to be -0.0134144 and 0.0000967 respectively. Since the coefficients of lagvacc\_per is negative and lagvacc\_per2 is positive, the quadratic of specification (3) will have a parabolic shape with an absolute minimum. In addition, the coefficients also implies that at low vaccination rates, an additional unit increase in vaccination rate has a negative effect on the log of weekly cases. However, at some point, the effect becomes positive and the shape of the quadratic means that the semi-elasticity of weekly cases with respect to vaccination rate is increasing as vaccination rate increases. We obtain this turning point of the vaccination rate by using the equation . Realistically, we would not expect weekly COVID-19 cases to begin increasing as the vaccination rate increases beyond 138.72 thus, for practical purposes, the quadratic to the right of 138.72 can be ignored. To determine the percentage change of increasing vaccination rate on weekly cases, we can use the equation:

Hence, an increase in vaccination rate from 10 to 11 decreases the weekly COVID-19 cases by about whereas an increase in vaccination rate from 20 to 21 decreases the weekly COVID-19 cases by around

.

• Use these above points to discuss whether the specifications (2) – (4) are able to address any

areas where the initial SLR assumptions from specification (1) fail.-

1. **Limitations of results**

[discuss possible problems with these specifications, especially omitted variables that may still lurk in the residual (i.e. do you interpret your results as causal or are they purely descriptive?). What are the most important remaining threats to the validity of your regression results?]

1. **Conclusion**

[Based on your analysis what conclusions would you draw about the effects of vaccination on COVID-19 infections in Israel.]

**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. ]

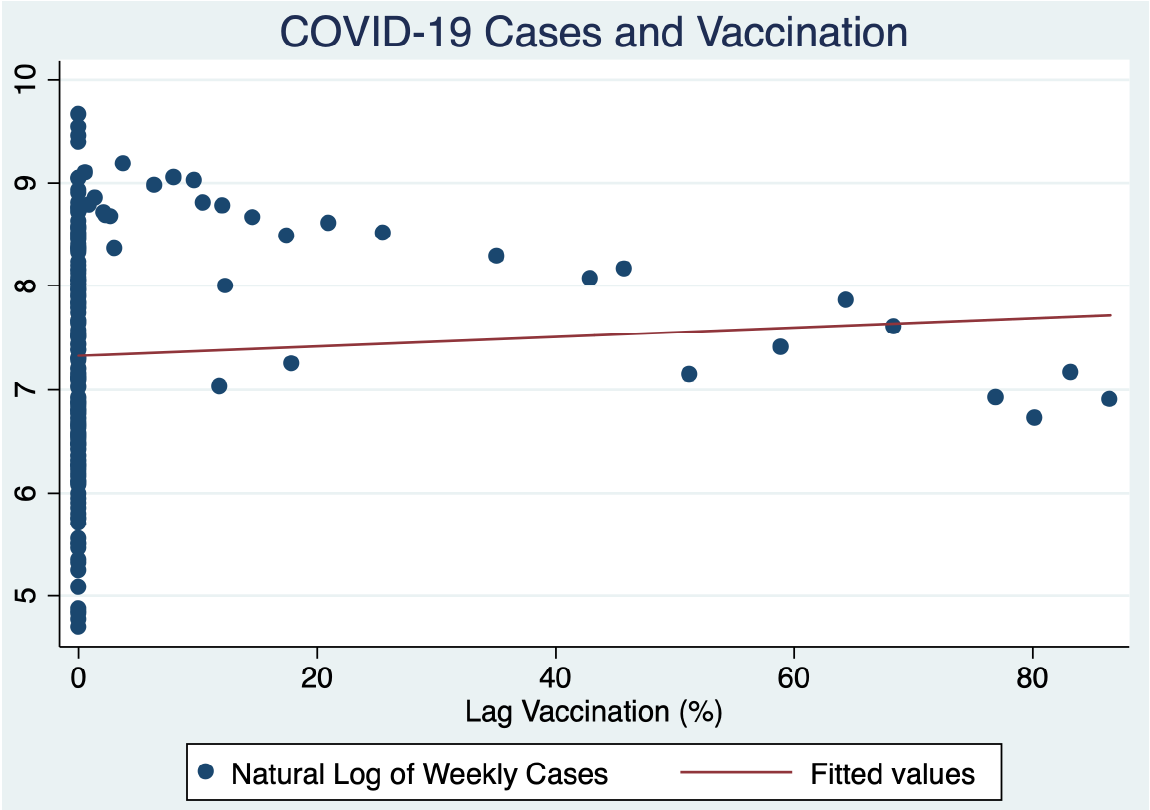
Table 1: Table of descriptive statistic

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Observations** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| Start Date of Week | 22 |  |  | 30th Aug 2020 | 24th Jan 2021 |
| Year-Week | 22 |  |  | 2020 Week 35 | 2021 Week 4 |
| Age Groups | 9 | 40-49 |  | 0-14 | 80+ |
| Weekly COVID-19 Cases  (per million people) | 198 | 2598.147 | 2065.757 | 262.2702 | 9049.648 |
| Natural Log of Weekly COVID-19 Cases | 198 | 7.343567 | 1.104857 | 4.691348 | 9.657331 |
| Two Week Lag of 1st Dose Vaccination (in percent) | 198 | 4.433254 | 15.22428 | 0 | 86.4881 |
| Two Week Lag of 1st Dose Vaccination (in percent) Squared | 198 | 250.2618 | 1094.51 | 0 | 7480.191 |
| Dummy: Two Week Lag of 1st Dose Vaccination |  |  |  |  |  |
| 0% of Population | 166 | .8383838 | .3690314 | 0 | 1 |
| (0%-10%) of Population | 12 | .0606061 | .2392111 | 0 | 1 |
| [10%-20%) of Population | 7 | .0353535 | .1851399 | 0 | 1 |
| [20%-100%] of Population | 13 | .0656566 | .2483086 | 0 | 1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2: Regression Analysis of log COVID-19 infection cases and vaccination | | | | |
|  | (1) | (2) | (3) | (4) |
|  |  |  |  |  |
| Lag Vaccination (in percent) | .0045166 | -.0057974\*\*\* | -.0134144\*\*\* |  |
|  | (.0051737) | (.0010479) | (.0036669) |  |
| Lag Vaccination (in percent) squared |  |  | .0000967\*\* |  |
|  |  |  | (.0000446) |  |
|  |  |  |  |  |
| Lag Vaccination: (0%-10%) |  |  |  | -.4550193\*\*\* |
|  |  |  |  | (.0870872) |
| Lag Vaccination: [10%-20%) |  |  |  | -.5430955\*\*\* |
|  |  |  |  | (.0926803) |
| Lag Vaccination: [20%-100%] |  |  |  | -.6761747\*\*\* |
|  |  |  |  | (.087064) |
|  |  |  |  |  |
| Age: |  |  |  |  |
| 15-19 |  | -.4400148\*\*\* | -.437979\*\*\* | -.3588792\*\*\* |
|  |  | (.043834) | (.0433679) | (.0438842) |
| 20-29 |  | -.0302479 | -.0233398 | .0466429 |
|  |  | (.0438458) | (.0434865) | (.0438842) |
| 30-39 |  | -.3023669\*\*\* | -.2922038\*\*\* | -.2207126\*\*\* |
|  |  | (.0438638) | (.0436402) | (.0436532) |
| 40-49 |  | -.3980066\*\*\* | -.3828524\*\*\* | -.3067294\*\*\* |
|  |  | (.0439233) | (.0440059) | (.0434309) |
| 50-59 |  | -.6554999\*\*\* | -.6363954\*\*\* | -.5753782\*\*\* |
|  |  | (.0440983) | (.0445023) | (.0434309) |
| 60-69 |  | -1.015151\*\*\* | -.9975742\*\*\* | -.9478091\*\*\* |
|  |  | (.0447388) | (.0449908) | (.0436201) |
| 70-79 |  | -1.65868\*\*\* | -1.653272\*\*\* | -1.606745\*\*\* |
|  |  | (.0453784) | (.0449547) | (.0436201) |
| 80+ |  | -2.042018\*\*\* | -2.032193\*\*\* | -1.983133\*\*\* |
|  |  | (.0450697) | (.0448102) | (.0436201) |
|  |  |  |  |  |
| Year-Week Dummies | No | Yes | Yes | Yes |
| Adjusted R-Squared | -0.0012 | 0.9827 | 0.9831 | 0.9849 |
| N | 198 | 198 | 198 | 198 |
| Notes: pertinent details; source data; details on variable definitions; time period; robust standard errors? etc.  The values in Table 2 are computed using Stata. The dataset used to compute the values is obtained from an Israeli dataset on COVID-19 infection and vaccination by age and week. The time period of this dataset is between 30th of August 2020 to the 24th of January 2021.  Lag Vaccination refers to the vaccination rate (the cumulative percent of people in that age group vaccinated with the first dose of the Pfizer vaccine) from two weeks before the current week. Lag Vaccination (0%-10%) refers to a dummy variable that indicates whether if the value of Lag Vaccination is between 0%-10% (excluding 0% and 10%). Lag Vaccination [10%-20%) refers to a dummy variable that indicates whether if the value of Lag Vaccination is between 10%-20% (including 10% and excluding 20%). Lag Vaccination [20%-100%] refers to a dummy variable that indicates whether if the value of Lag Vaccination is between 20%-100% (including 20% and 100%).  The quantities in parentheses below the estimates are the standard errors. | | | | |
| \*\*\*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



To determine the percentage change of increasing vaccination rate on weekly cases, we can use the equation:

Hence, an increase in vaccination rate from 10 to 11 decreases the weekly COVID-19 cases by about whereas an increase in vaccination rate from 20 to 21 decreases the weekly COVID-19 cases by around

.