Paige Bodine

Prof. Casanova

Econ 504

22 May 2022

Final Report

‘The project described in this paper relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS.’

The Covid-19 Pandemic affected the entire world. Stay at home orders were put in place in an attempt to lessen the spread of the virus. Many people advocated for the use of masks and vaccines to help lessen the spread of Covid-19 and to lessen the effects that the virus had on those who fell ill with it. However, the pandemic and these mandates for masks, isolation and vaccines became very political, at least in the U.S. Many people have different attitudes toward Covid-19, masks, vaccines and isolating. These different attitudes can be due to individual characteristics such as race, gender, age, political beliefs, education levels, religion, income and the list goes on. One paper explains the story of a woman driving to receive her first Covid vaccine in which she observed that everyone in line to receive their vaccine was a woman (Ungar 2021). This paper also goes on to explain that the split between women and men receiving a Covid vaccine was around 60% to 40% respectively (Ungar 2021). Another article states that men are more likely to get the vaccine as women are more likely to fear adverse effects to the vaccination and that this is consistent with existing literature on vaccine hesitancy (Bellon 2021). However, age also plays a role in this as well. A paper regarding age and gender attitudes toward Covid vaccines states that women aged 50 and older are more likely to get a Covid vaccination than men in the same age category (Keenan, et al. 2021). Lastly, race also plays a role in individuals who are vaccinated. A paper explains that black individuals are less likely to have been vaccinated in comparison to white individuals (Ndugga, et al. 2022). In this paper, I will be looking at individual attitudes, specifically gender, toward Covid-19 vaccinations to see how being male versus female and having other characteristics affects how one views Covid-19 vaccines.

For this paper, I will first be running a multivariate OLS regression and then I will be using the ordered probit model to estimate my equation. This will show me the differences between using a simple model and a model better suited for the variables I am using. I say better suited because the OLS model is used to measure simple linear models, however the data I am using for this model is not linear. The ordered probit model is a different form of the multinomial probit model. The multinomial probit model is used to estimate models in which the dependent variable has more than one category to choose from. My dependent variable in my model is a categorical variable with four different categories. So, in deciding which model to use, I at first thought about using a regular multinomial probit model. But, after some research, I found that the ordered probit model would be a better fit as my dependent variable has categories that are ordered in a specific way. In order to use the ordered probit model, we have to make the assumption that there is a particular probability distribution for the error term when representing the probabilities of each outcome, specifically that there is a normal distribution among the error term. This is because the probit model is estimated by a procedure called the maximum likelihood. The probit model is nonlinear because the response probability is a nonlinear function of the parameters and it cannot be measure by a normal OLS model.

The probit model requires a few steps when running it. The first being the joint probability which gives us the probability of observing the sample data, also known as the likelihood function. The maximum likelihood function chooses the estimates of the unknown parameters that maximize the likelihood of observing the sample. However, we usually maximize the log of the likelihood function which is what is done in Stata. This is done because measuring the log of the likelihood function as opposed to measuring the likelihood function without the log is easier. The reasoning is that the log likelihood function is a sum of a series of terms opposed to the product of a series of terms. After running the ordered probit model, we need to run the marginal effects. This is the correct way to interpret our results because the estimates from the probit model can be difficult to interpret and running the marginal effects gives us a more accurate measurement of our coefficients. The last thing left to do in the ordered probit model is to run a hypothesis test. This is done using the likelihood ratio test. This is used to test whether certain variables are irrelevant in predicting our model and aids in decreasing the log likelihood in the model.

There are some econometric issues when running an ordered probit model. One of these issues is that you must make an assumption that the error term is normally distributed in order to run the model. Another limitation is that the functional form of the response probabilities is complicated. Also, running a model in Stata to maximize the log likelihood function may not converge in models that have more than four options. Lastly, the regular multinomial probit is used when the choice options have no natural order. As I stated previously, my dependent variable is ordered in a specific way, from strongly disagree to strongly agree. So, I ran the ordered probit to get past this limitation. In order to address causality issues, a joint probability test and a hypothesis test need to be run to ensure that the variables being used in the model are correlated. F tests are also needed to make sure that the variables are statistically significant. Also, it is important to run the marginal effects after running the ordered probit model when interpreting the results. In my model, there may be omitted variable bias due to a lack of data in my survey. One endogeneity concern would be political beliefs. I do not have this variable in my dataset, and this is something that is correlated with both gender and vaccine beliefs. I would expect there to be a negative omitted variable bias from not controlling for political beliefs in my model.

For this paper, I will be using the UAS 350 dataset. This dataset asked survey respondents to answer Covid-19 related questions. I used this dataset for all of the variables I ran in my model. The UAS 350 dataset is a panel survey titled “Coronavirus tracking survey wave 30”. This survey asked questions to the respondents about the impact that the Coronavirus pandemic had on their lives. However, this survey was a standalone survey and was not part of the previous Covid surveys conducted from the Understanding America Study. It included normal questions about the individual’s income, race, gender, etc. and also included more specific questions regarding Covid-19 such as mask and vaccination mandates, booster shots, etc. This survey was offered to 9,736 UAS participants and 7,172 respondents completed the survey. 107 respondents out of the 7,172 did not fully complete the survey and therefore, are not counted as respondents. The overall response rate was 73.66% for the UAS 350 survey.

I adjusted a few variables in the dataset by recoding them. For example, I recoded education to have less categories. These new categories were less than high school, high school diploma, some college, bachelor's degree, and more than a bachelor's degree. I also recoded the household income variable with new categories. These categories included less than $5,000, $5,000 to $49,999, and $50,000 and above. The last variable I recoded was marital status. I coded this to categories including married, divorced/separated, and single. The last change I made to the data was renaming the variable cr072d to vacc\_useful to make it easier to find the correct variable when running my models. This variable is a question asking respondents if they believe that Covid vaccinations are useful and effective. I tabulated each of the variables in my model and found that many of them were not missing a lot of responses. The vaccine variable was the variable that was missing the most responses and in total, 119 individuals did not answer the question while 7,160 individuals did. This is most likely due to the fact that 107 respondents did not complete the survey. However, 98 individuals did not see the question at all. I decided to use the vaccine variable as I wanted to look at attitudes toward vaccinations instead of whether an individual was vaccinated or not. The reason for this is that individuals may have been required to receive a vaccination for work or some other reason even if they did not desire to. I also chose education, race, income, age, and marital status as these can affect how an individual views Covid-19 and vaccines. I decided to recode variables to create simplicity within the model. For example, with education, I believe that there will not be much of a difference in the way an individual views Covid vaccines whether they have a first grade level of education compared to a second grade level of education. Therefore, I bunched this group up into individuals with less than a high school education. The same reasoning goes for marital status and household income. One shortcoming in my data is the gender variable has an uneven number of men and women. There are around 60% women and 40% men in the respondents in this survey.

When running my model in Stata, I used the regression equation:

Vacc\_useful = β0 + β1 gender + β2 marital status + β3 age + β4 age^2 + β5 education + β6 race + β7 household income +

I first ran this equation using an OLS regression. For simplicity, I will only be interpreting the coefficient value on my main explanatory variable for the OLS model. To interpret the value in the OLS regression I ran, being male leads to a 0.015 unit increase in the probability of believing that Covid-19 vaccines are useful and effective, holding all other variables constant. However, this value is not statistically significant because the p-value is 0.472, suggesting the gender is not correlated with Covid-19 vaccination attitudes. When looking at the F test measured in the OLS model, we get a value of 0 which means that the variables are jointly significant in explaining the model at the 0.5% significance level.

In order to accurately test the model, I ran an ordered probit due to the dependent variable being in a specific order. With the ordered choice model, we received a positive coefficient estimate for gender, which means that being male leads to an increase in the probability that they strongly agree that covid vaccines are useful and effective in comparison to women. However, this value is still not statistically significant, suggesting that gender and Covid beliefs are not correlated.

In order to accurately interpret the results of the ordered probit model, I ran the marginal effects to receive the correct coefficient values. I will be interpreting all variables in the equation to see the effects they have on vaccination beliefs. The coefficient for gender on the marginal effects estimation is 0.0078 for the fourth category, therefore being male increases the probability of strongly agreeing that Covid-19 vaccines are useful and effective by 0.0078 units when compared to women. Looking at the first category we have a value of -0.0027, therefore, being male decreases the probability of strongly disagreeing that Covid-19 vaccines are useful and effective by 0.0027 units. However, all values estimated with the marginal effects on gender are not statistically significant with p-values of approximately 0.449 and above. When looking at marital status, individuals who are divorced have a coefficient value that is essentially 0 and a high p-value, suggesting that being divorced does not affect Covid vaccination attitudes. However, looking at individuals who are single in the fourth category, we see that the coefficient is 0.082 and is highly statistically significant. This means that an individual being single increases their probability of believing that Covid vaccines are useful and effective by 0.082 units in comparison to individuals who are married. Looking at age^2, we see that the coefficient is nearly 0, therefore meaning that age is not correlated with Covid beliefs. With education, we see that Bachelor’s degree and above levels of education lead to an increase in the probability that these individuals believe that Covid vaccines are useful when compared to individuals with less than high school education. With race, being Black leads to a 0.039 unit decrease in believing that Covid vaccines are effective while being Asian leads to a 0.073 unit increase in believing that Covid vaccines are effective when compared to white individuals. Lastly, with income, we see that higher levels of income lead to individuals believing that Covid vaccines are useful. Individuals with income $50,000 and above have a 0.14 unit increase in the probability of strongly agreeing that Covid vaccines are useful.

For the next step, I used the likelihood ratio test and included only my explanatory variable for gender and got a value of 0.450. With this value, I can conclude that the gender variable is not statistically significant, which means that the prediction of gender having no significance in the model is strongly not rejected. Therefore, gender is not correlated with vaccination attitudes.

Next, I ran the average marginal effects on the OLS regression. The coefficient for gender was 0.154, which means that being male leads to a 0.154 unit increase into believing that Covid-19 vaccines are useful and effective. This p-value is 0.472 which is still not statistically significant. The OLS regression is treating the vacc\_useful variable as a continuous variable instead of a categorical variable, so I am not surprised that the results from the ordered probit model are different.

After running the average marginal effects on the OLS model, I realized how important using the ordered probit was regarding interpretations of the model. This created more accuracy when looking at individual variables and their interactions with the different categories in the dependent variable. When comparing the coefficient for gender after estimating the marginal effects between the OLS and the ordered probit model, I found that the coefficient values were very different in magnitude, but they were both positive. The coefficient for the OLS regression was nearly double the magnitude of the coefficient in the ordered probit marginal effects, which makes sense because the OLS is a linear while the ordered probit is nonlinear. It also makes sense because the OLS only takes on values of 0 and 1. Lastly, the OLS marginal effects is treating the dependent variable as a continuous variable. This shows that the OLS may have been biasing the estimates and that the ordered probit model addressed those issues. This also shows that there was a positive bias in the OLS regression. However, it makes sense that the p-values in both marginal effects were very high meaning that the gender variable is not statistically significant as this has been consistent with all my findings and that the coefficient values were both positive. This is because both the OLS and the ordered probit model are measuring the probability scale of gender and the effects are nonlinear.

I decided to merge data from the UAS 15 dataset to test whether the variable for political views would bias my estimates as I suspected and would be useful for my model. To do this I merged the S006 variable which is a categorical variable measuring political views from most conservative to most liberal into my UAS 350 dataset. I changed the S006 variable to the name political. Then I added this to my ordered probit model to see the effects this would have on my estimates. When I reran the probit model with the added variable for political views, I found that it created a higher value for the coefficient on gender and the variable is statistically significant. This shows that the variable for political views created a negative omitted variable bias. When I compared the Pseudo R^2 in both models, I saw that it increased from 0.05 to 0.09 showing that adding the political variable better explained my model. However, I did not keep this in my model as the observations for the political variable were very low compared to the rest of the observations in my model.

In this paper, I used the ordered choice model to look at how gender and other individual characteristics affects how one views Covid vaccinations. I found that gender is not correlated with vaccination beliefs which is inconsistent with previous studies done. This can be due to a lack of data and observations. However, I did find that men were more likely to have positive beliefs toward Covid vaccines than women, despite it being statistically insignificant. I also found that individuals who are Asian are more likely to have positive beliefs toward Covid vaccines and Black individuals were the opposite when compared to white individuals. This is consistent with previous papers I read regarding race and vaccination beliefs. Individuals who received postsecondary education were also more likely to have positive beliefs toward Covid vaccinations and so were individuals with higher income. Education appeared to have the highest effect on Covid vaccination beliefs along with high income. An individual who is single is more likely to have positive beliefs toward Covid vaccine compared to individuals who are married. Lastly, age did not appear to make much of a difference on Covid beliefs. Therefore, marital status, race, income and education can impact an individual’s views toward Covid vaccines. However, gender and age do not play a role in Covid vaccine beliefs, which is inconsistent with previous articles I read.

References

Ndugga Follow @nambinjn on Twitter, N. F. @hill\_latoya on T., Hill, L., Artiga, S., & Haldar, S. (2022, April 11). *Latest data on covid-19 vaccinations by race/ethnicity*. KFF. Retrieved May 7, 2022, from https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/

Keenan, T. A., Lampkin, C., & Rainville, G. C. (2021, March 1). *Views on vaccines: Gender makes a difference*. AARP. Retrieved May 7, 2022, from https://www.aarp.org/research/topics/health/info-2021/views-on-vaccines-gender-differences.html

Bellon, M. (2021, September 23). *Gender differences in covid-19 vaccine hesitancy*. The Clayman Institute for Gender Research. Retrieved May 7, 2022, from https://gender.stanford.edu/news-publications/gender-news/gender-differences-covid-19-vaccine-hesitancy

Ungar, L. (2021, August 13). *The gender vaccine gap: More women than men are getting Covid shots*. Kaiser Health News. Retrieved May 7, 2022, from https://khn.org/news/article/gender-vaccine-gap-more-women-than-men-vaccinated-against-covid/