An empirical analysis of factors affecting COVID-19 death severity across U.S counties

By: Cong Liu

1006267200

ECO 375: Applied Econometrics

University of Toronto

Department of Economics

Assignment 1

October 2021

Abstract: The outbreak of the COVID-19 pandemic has deeply impacted the human society. If is of great importance for mankind to identify influential factors behind the pandemic death severity. This paper leverages cross-sectional data on COVID-19 situation, demographic and behavioural features on U.S. county level and employs linear regression models to understand how the Republican vote share and full-dose vaccination rate affect the pandemic death rate. The results show there that is a confounding effect between vaccination and partisanship to the death rate trend.

1. Introduction

COVID-19 has brought life threats and economic hardships to the world. As of October 14, 2021, there have been 239,007,759 confirmed cases of COVID-19, including 4,871,841 deaths (WHO, 2021). An article from the *Economist* addressed a concerning phenomenon that U.S. states with a higher Democratic vote share were hit harder by COVID-19 in 2020, demonstrated by an on average, three times higher mortality rate by mid-May 2020 (Economist, 2020). With the implementation of lockdown policy and the introduction of vaccines, the pandemic curve has been flattened during the past year. Yet the emergence of a new delta variant of the COVID-19 virus is leading to another pandemic wave. It is worth studying factors affecting pandemic death rate in the United States, with the focus on Republican vote share and full-dose vaccination rate.

2. The Context and Data

This paper outlines an analysis conducted on a sub-dataset of 3,000 observations sampled from a larger cross-sectional data. The larger dataset records, on the U.S. county level, the magnitude of the latest COVID-19 wave, Republican vote share in the 2020 presidential election, vaccination status, and other factors. In 2,460 out of 3,000 U.S. counties in the dataset, the majority voted for the Republican (Table 1). There is an observable difference in the mean COVID-19 death rate across counties with different political affiliation. Post the 2020 U.S presidential election, in Republican counties, the average pandemic death rate per 100,000 is 13.06 higher. By contrast, the average full-dose vaccination rate is 11.7 percent points lower in the Republican counties. A rough overview of the data reveals a possible positive correlation between Republican vote share and death rate. In addition, Republican counties seem to be more slow on vaccination. Then vaccination status and Republican vote share may be confounded in affecting pandemic death severity.

3. Regression analysis

3.1. Simple Linear Regression

The first look on the relationship between partisanship and pandemic severity across US counties is a simple regression between COVID-19 death rate and the Republican vote that yields the model below,

$$\widehat{death\_rate} = 6.993163 + \ 0.4331098\ per\_gop20. \text{ [1]}$$

where $\widehat{death\_rate}$ is the predicted rate of post-election COVID-19 deaths per 100,000 cases and per_gop20 is the counties' percent Republican vote share in the 2020 election.

According to the regression result, for one percent point increase in the Republican vote share within a U.S. county, the rate of COVID-19 deaths per 100,000 is estimated to increase by around 0.4331 on average. To evaluate whether this effect is statistically significant, a hypothesis test is conducted on hypotheses below

$$H_0: \beta_1 = 0$$

versus

$$H_1: \beta_1 \neq 0.$$

where $\beta_1$is the slope coefficient for COVID-19 death rate being regressed on Republican percent vote share.

The t statistic computed from the simple linear regression in [1] is roughly 13.23, which verifies a very statistically significant and practically large effect of Republican vote share on the COVID-19 death rate across U.S. counties.

Along with the significant t statistic, the scatterplot (Figure 1) between Republican percent vote share and COVID-19 death rate shows an approximately linear relationship. Thus a linear model does yield an appropriate estimate of the population relationship between these two variables. All observations are independently and identically distributed since the regressed dataset is a random sample of the larger cross-sectional U.S. county data and it is reasonable to assume the counties are independent. To observe how change in each county's Republican percent vote share has an effect on the COVID-19 death rate, we need variations in the vote share across counties in the dataset. Such variation does exist, as Table 1 reports that the standard deviation of per_gop20 across counties is non-zero. However, it may not be a tenable assumption that all other county characteristics are independently distributed of the Republican vote share. Clearly, the Republican vote share, an indicator of counties' political affiliation, is not randomly assigned to the U.S. counties. Furthermore, one is able to see a difference in racial makeup between Republican and Democratic counties. Plotting the SLR's residuals against the corresponding fitted values (Figure 2) shows a fanning pattern where residuals gradually become more spread out. The homoskedasticity assumption of SLR is likely to fail as residuals serve as a proxy of the population error term.

3.2. Multiple linear regression

The pandemic curve is evidently flattened out after the wide use of COVID-19 vaccines. Having witnessed the effect of vaccination on the pandemic severity, not taking vaccination into model consideration may end up with a biased association between the variables of interest. A multiple linear regression (MLR) allows the addition of extra regressors to the linear model.

There is a striking change in the sign of the coefficient on $pergop\_20$ when moving from simple to multiple linear regression. Additionally, the effect of Republican vote share on COVID-19 death rate is no longer statistically significant (t-statistic of $\widehat{\beta_1^{MLR}} \approx 1.39$) as it was in the previous SLR model. On the other hand, the multiple regression result shows a strongly statistically significant (t-statistic ≈ -18.09) and negative effect of full vaccination on the death rate. A one percent point change in full vaccination rate is associated with a 1.247173 decrease in the death rate per 100,000 cases on average, holding the Republican vote share constant. The effect of full-dose COVID-19 vaccinations in reducing the pandemic death rate is also practically large. As Republican vote share and full-dose vaccination rate have a strong negative correlation (Table 3), full-dose vaccination rate plausibly satisfies the two criteria for an omitted variable: it is a determinant of the death rate and is correlated with the Republican vote share. The omission of it in the previous SLR clearly results in a positively biased slope estimate for the Republican vote share.

Theoretically, all economic benefits by nature yield a decreasing return. One may raise concern about the decreasing return of vaccination in combating the pandemic. Since MLR allows a flexibility in the functional form while preserving the linearity in parameters, a quadratic term of the full vaccination rate is included in the model to investigate its non-linear effect. The effect of full vaccination on COVID-19 death rate remains statistically significant and negative. However, it is estimated that for a one unit change in the squared full vaccination rate, there could be an additional 0.0078913 increase in the COVID-19 death rate, holding all other variables constant. This signals that vaccination might start to yield a decreasing return, and the computed t-statistics of 2.08 for the slope estimate supports the statistical significance of the quadratic term. We then could conclude that there is a non-linear effect of full-dose vaccination on COVID-19 death rate. Yet the turning point of vaccine's effect in reducing the death rate is at 118.618732% of full-dose vaccination rate, meaning that more than 100% of the population is fully vaccinated. Thus

practically, the diminishing return of vaccination in reducing the death rate should not be a concern.

To control on demographic and behavioural features of U.S. counties, more regressors are added to the model specification. Surprisingly, Republican vote share regains its statistical significance in this larger model and has a similar positive effect on COVID-19 death rate as shown in the SLR result, details in Table 2. This shift in significance implies that unobserved variables are still not independently distributed of the Republican vote share, despite extending the model. The regression result shows that the full-dose vaccination still has a statistically significant negative effect on the death rate. The prevalence of obesity and diabetes both have a significant effect on COVID-19 death rate at 5% significance level. Obesity has the most significant, and unexpectedly negative effect on the death rate. Aligned with vaccination's negative effect on the death severity, counties that are more hesitant to be vaccinated are estimated to have a higher death rate comparing to those that are probable to get the vaccine. Interestingly, counties with a higher percentage of American Indian and Black people have a higher death rate relative to the Other (non-Hispanic/Asian/Native Hawaiian) racial group. Urbanity of the counties is also shown significantly influence death severity. A non-core county is estimated to have a death rate per 100,000 COVID-19 cases lower by 8.709596, relative to a large-central metropolitan. Counties that are socially vulnerable are also more vulnerable to the COVID-19 hit. For counties categorized as having a very high social vulnerability, they are estimated to have a 14.75054 higher death rate relative to counties with a very low social vulnerability.

4. Limitations of results

Even in the largest model, there is evidence that potential confounders are still not included in the specification. For example, it is likely that counties with a high proportion of older population, even with high vaccination rate, to have higher COVID-19 death rate. That is, the percentage of the elderly population at each U.S. county could plausibly affect the death rate and may be correlated with independent variables in the existing model, resulting in an omitted variable bias. There are many other disease transmission related factors that are truly underpinning the U.S. counties' pandemic death rate, yet omitted in the above model specifications. The true effect of these factors on the death rate could even be non-linear. The lack in consideration of the lurking unobservable leads to my conclusion that the regression results above are purely descriptive instead of causal.

5. Conclusion

Based on the above analysis, we can conclude that U.S. counties the observed relationship between partisanship and pandemic death severity is confounded with vaccination status of the counties. Linear regression result shows that counties where the majority voted for the Republican Party have a lower vaccination rate on average. Since full-dose vaccination is shown by regression to have statistically and practically significant effect in decreasing the pandemic death, we then observe the trend that counties with a higher Republican vote share have a higher COVID-19 death rate on average. Despite the significant positive linear effect shown by SLR and the largest MLR model, the Republican vote share does not have a meaningful impact on the severity of COVID-19 death.

References:

Covid-19 is hitting democratic states harder than republican ones: Daily chart. 2020. *The Economist (Online)*. May 22, http://myaccess.library.utoronto.ca/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fmagazines%2Fcovid-19-is-hitting-democratic-states-harder-than%2Fdocview%2F2405919824%2Fse-2%3Faccountid%3D14771 (accessed October 9, 2021).

Stack Overflow. "stata – use esttab to generate summary statistics by group with columns for mean difference and significance." https://stackoverflow.com/questions/15556078/use-esttab-to-generate-summary-statistics-by-group-with-columns-for-mean-differe (accessed October 10, 2021)

Wooldridge, J. M. *Introductory Econometrics, A Modern Approach*. South-Western: Cengage Learning, 2020. (accessed October 10, 2021)

WHO. "WHO Coronavirus (COVID-19) Dashboard." https://covid19.who.int/ (accessed October 14, 2021)

Table 1: Summary statistics of variables in the regression analysis

	Counties where majority voted Republican		Counties where majority voted Democratic	
	mean	sd	mean	sd
Deaths per 100, 000 post June 26 2021	37.51	32.31	24.45	24.13
Percent of votes for GOP pres: ER data	70.94	10.00	37.73	9.79
Percent fully vaccinated: VAXS data	34.21	8.69	45.91	10.80
Prevalence of Obesity: UC data	35.51	3.63	32.61	6.52
Prevalence of Heart Disease: UC data	8.89	1.59	7.30	1.89
Prevalence of Diabetes: UC data	13.21	2.31	12.47	3.86
Hesitant: Definitely not: VH data	9.02	3.01	6.93	3.52
Hesitant: Probably not: VH data	4.83	1.48	3.67	1.69
Hesitant: Unsure: VH data	5.97	1.35	5.50	2.08
Hesitant: Probably of Definitely get Vax: VH data	80.18	4.76	83.89	6.32
Percent Hispanic: VH data	8.13	11.53	14.00	17.62
Percent American Indian (non-hispanic): VH data	1.33	4.47	2.23	11.34
Percent Asian (non-hispanic): VH data	0.78	1.04	3.48	4.18
Percent Black (non-hispanic): VH data	6.80	10.63	20.23	22.81
Percent Native Haawiian (non-hispanic): VH data	0.06	0.16	0.10	0.21
Percent White (non-hispanic): VH data	80.96	15.47	57.49	23.63
Percent other: VH data	1.94	1.54	2.49	1.31
Cases with documented covid-19 per 100,000 pre June 26th 2021	10299.28	2687.01	9308.87	3391.11
Deaths per 100,000 pre June 26 2021	212.19	109.09	174.42	123.72
Urban vs Rural Code: UC data	4.85	1.39	3.57	1.72
Social Vulnerability Index: Categories: VH data	2.95	1.38	3.20	1.47
Observations	2460		540	

Table 2: Regression Analysis of COVID-19 Death Rates, Republican Vote Share and Vaccination

	(1)	(2)	(3)	(4)
Republican vote share (in percent)	.4331098***	-.0583225	-.0568547	.4537798***
	(.0327252)	(.0420863)	(.0418664)	(.0590526)
Vaccination Rate (in percent)		-1.247173***	-1.872112***	-.6749273**
		(.0689597)	(.3302102)	(.2882136)
Vaccination Rate Squared			.0078913**	.006094*
			(.0037888)	(.0032518)
Obesity				-1.545116***
				(.1660958)
Heart Disease				2.349424*
				(1.261948)
Diabetes				1.759495**
				(.8582376)
SVI: Very Low Vulnerability				0
				(omitted)
SVI: Low Vulnerability				6.182104***
				(1.262034)
SVI: Moderate Vulnerability				11.76136***
				(1.534529)
SVI: High Vulnerability				13.62652***
				(2.000428)
SVI: Very High Vulnerability				14.75054***
				(2.637309)
Hesitancy (definitely not)				1.222268***
				(.2469946)
Hesitancy (probably not)				1.673465***
				(.4956858)
Hesitancy (unsure)				.6927978*
				(.4014685)
Covid survival (pre period)				-.0012184***
				(.0002059)
Covid death (pre period)				.028919***
				(.0064267)
Large central metro				0
				(omitted)
Large fringe metro				1.74764
				(2.638447)
Medium metro				-1.58948
				(2.677665)
Small metro				.5594404
				(2.910955)
Micropolitan				-4.681079*
				(2.788406)
Noncore				-8.709596***
				(2.866047)
Controls for racial make-up	No	No	No	Yes
Adjusted R-Squared	0.0496	0.148	0.149	0.363
N	2,998	2,998	2,998	2,998

Notes: Robust standard errors in parentheses. Pre period means records pre June 26 2021. Some variables are omitted to avoid perfect multicollinearity. Vote share from ER data. Vaccination from VAXS data. Underlying condition prevalence and urbanity from UC data. Hesitancy and SVI from VH data.

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

Figure 1: Scatterplot between Republican percent vote share in 2020 U.S. presidential election and COVID-19 death rate by U.S county

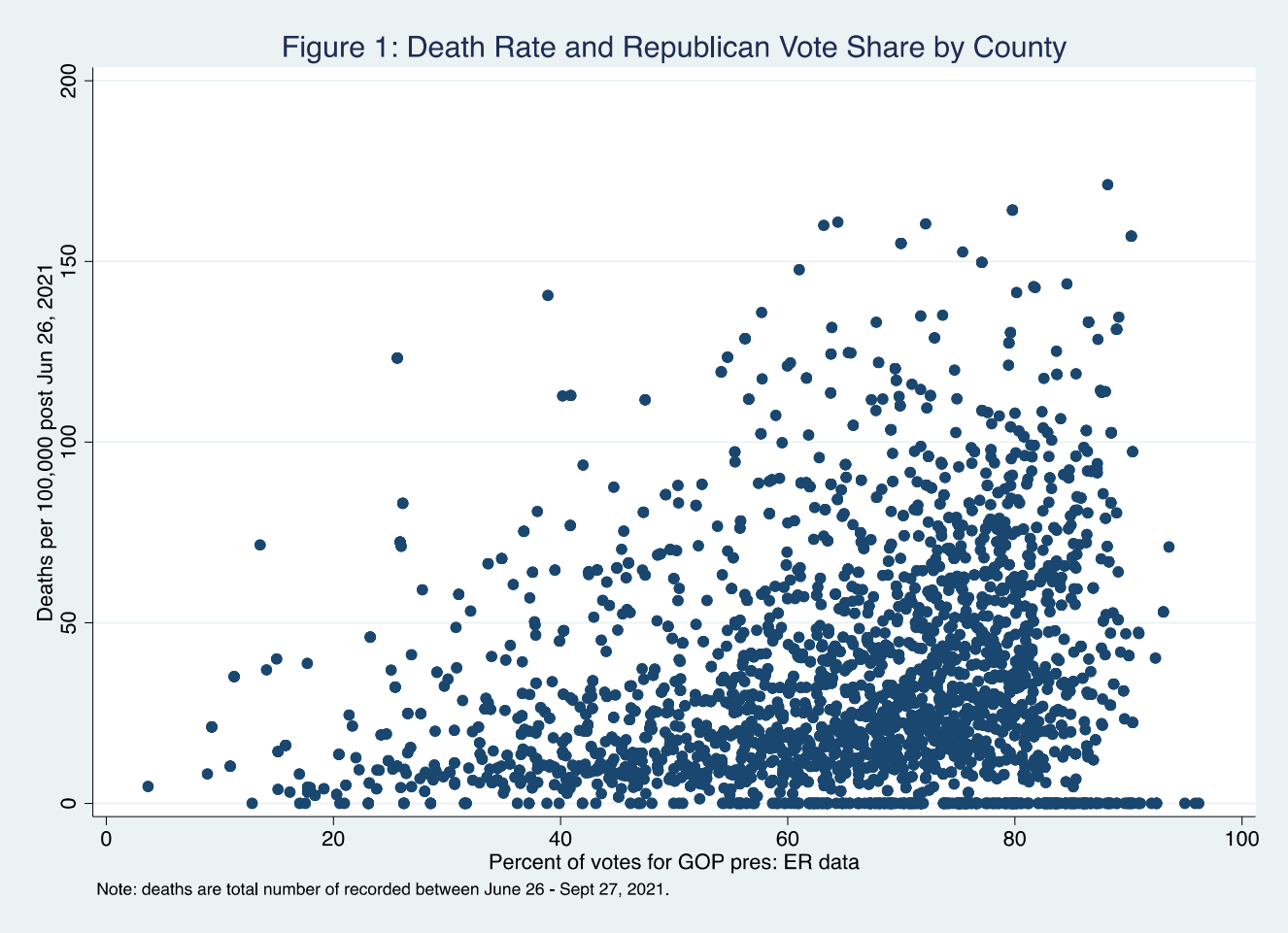

Figure 2: Fitted value versus residual of SLR model

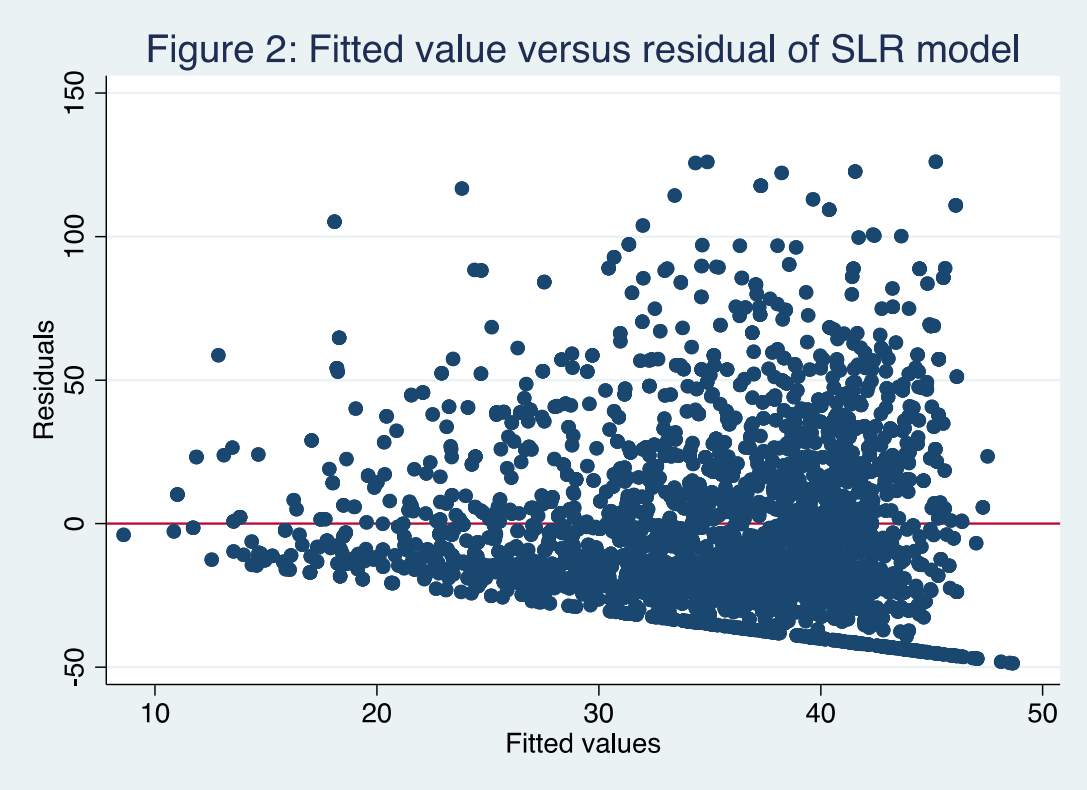

Table 3: Regression Analysis of Full-dose Vaccination and Republican Vote Share

	(1) Percent fully vaccinated: VAXS data
Percent of votes for GOP pres: ER data	-0.394***
	(0.0115)
Constant	61.92***
	(0.812)
Adjusted R-Squared	0.395
Observations	2998

Robust standard errors in parentheses.
***Significant at the 1 percent level.