How People Move

The UK's Reaction to Covid-19

Tom Clarke

4/26/2020

Introduction

Covid-19 has been the centre of data science since its inception in the public awareness. Crucial flaws in many economic and social systems are being exposed and analyzed whilst all the while my senses are being bombarded with straight negativity. Every new model from Imperial College London, Ph. D X, or Research Group Y has incorporated cutting edge technology, modeling, and analytics to bring ever-more accurate predictions of the course of this pandemic. And, it's no surprise, they get ever-more depressing. I think that the Covid research scene could do with a little positivity, a little boost of hope, a ray of sunshine from Tom Clarke. But how will he distill an upshot from a pandemic? You may wonder. Well, with colorful graphs and hopeful evidence for policy change, that's how.

The research question I set out to answer in this paper is how the differing levels of education individuals have impacts their response to the pandemic. Specifically, I want to know if the level of education one has impacts how they react to government orders and news on shelter-in-place guidelines. The 'reaction' will come in the form of how individuals movement patterns change following the news of Covid-19 actions in their country. If there exists a difference in reaction to the virus, then not only are some groups disproportionately affected, but the net impact on an area is greater. Overcoming the virus with as few fatalities as possible is a cooperation game; the more that everybody commits to following the guidelines, the fewer people die, and the quicker everybody can start holding hands in public. By identifying the group, or groups, which seem to exhibit the least beneficial-for-society-as-a-whole response, then policy makers can hit the drawing board with a goal. Social media campaigns, news broadcasts, presidential addresses, and any dispersion of Covid-related information can be targeted, with the assistance of individuals from the desired target demographic and Human Centered Designers, to maximize the digestion and understanding of the severity of the virus, to increase trust in their government, and to best encourage people to shelter-in-place.

I believe that sheltering in place can only truly be successful if every individual feels it necessary to shelter in place, they need to understand the benefits of doing so and the ramifications of not. To achieve this, having Ph.Ds and various other experts produce research and findings which are impossible for average individual to decipher is not useful. Experts make up a small proportion of the population. The extremely valid information being produced needs to be made easy to internalize for everybody. This is where knowing the demographics of focus can yield huge benefits for the overall recovery of a society when any national issue strikes.

This leads me to my intuition for why education impacts the response to the pandemic. People generally act on what they understand. Or think they understand. If I read literature on the pandemic by top medical experts and think that I understand that, baed on their research, the best individual action I can take is to shelter-in-place diligently and encourage those I know to do the same. If I, however, read countless news reports and listened to everyone around me proclaim that I must stay inside yet I don't fully understand why I should, or why going outside just once is *that* bad, then I will indeed be unhappy to do so. This is where education comes in. I believe that only with higher education or, age, comes the ability to filter through masses of media and junk. And only with educations comes the ability to locate and understand factual, evidence based resources. Thus being more educated comes with the ability to understand more, to know my strongly the reasons why. Thus with a strong grasp of the situation, acting for the social good is easy because it is understood to be the best thing to do. The converse is more impactful. Having a person on your television screen tell you to “stay inside, do not work, and do not panic buy because, **trust me**, we won’t run out of food and everything will be okay,” is not convincing or comforting for many people. Not when this is the only source of information.

It is no secret that education is a privilege. This creates a new privilege, the privilege of knowing and of being able to know. This is very subtle and can create large divides. It is very easy to observe so many people not sheltering in place and be angry at these people. It is much harder to think about why some people just aren’t sheltering in place. Is it because they are malicious? Because they want more death? Because they don’t care about anyone but themselves? Maybe. But maybe it is because they don’t truly, entirely understand the global situation and how they as individuals are an integral part of the global ecosystem. Having some evidence to support this idea would allow for action to be taken on something that I consider a huge problem and a symptom of something much more ubiquitous; the disconnection of becoming more educated.

One key issue I have had to work around in this paper is the lack of individual level mobility data. To resolve this I resort to county level data on average daily movement. This is not ideal since individuals are precisely that and thus average movement doesn’t paint the most accurate picture. Additionally each individual has their own education level and not having access to the movement of each individual and their corresponding level of education is a downside. That said, obtaining such data was not an expectation and the data source I do have I believe to suffice.

Data

My data comes from two sources. The data on mobility comes from Google’s Covid-19 Mobility Report. For 132 different countries, Google provides county-level data on the average time spent in 6 different categories of location. The average time and number of visits are recorded over a two month period, every day, beginning 15/02.. The average for each day is compared to a baseline for that specific day of the week, which is calculated by taking the median value of time spent and visits made from Jan 3 - Feb 6 2020. I chose to focus specifically on counties in England. Using the 2011 Census data I acquired data on number of of individuals within a county with a given education level and a breakdown of counties by social class. The census data for education provides various cateogories, each denoting the highest level of education that an individual has obtained. The categories are; No Academic or Professional Qualifications, Level 1, Level 2, Level 3, and Level 4. These levels translate approximately to; No Qualifications, Completed at Most 9th Grade, Completed at Most 11th Grade, Completed at Most 12th Grade, and Completed a College Degree or Higher. I created a new variable to represent the individuals who responded to the survey but whose qualifications didn’t fall into any of the categories. These are individuals with foreign qualifications or qualifications which cannot be indentified equivalently in the UK education system. Only a small proportion of the population fall into this category. This is large enough to include them in my model.

change_rec_retail	**No.Qualifications**	**Grade.9**	**Grade.11**	**Grade.12**	**College.Grad**
	0.3	0.34	0.21	-0.44	-0.42
		0.4	-0.1	-0.37	-0.83
			-0.22	-0.64	-0.31
				-0.09	-0.21
					0.18

Inspecting the correlation between the different highest qualification one has and the average change in mobility we find the following. $\rho(Change\ In\ Mobility,\ No.Qualifications) = 0.3$, $\rho(Change\ In\ Mobility,\ Grade.9) = 0.34$, $\rho(Change\ In\ Mobility,\ Grade.11) = 0.21$, $\rho(Change\ In\ Mobility,\ Grade.12) =$ -0.44 and $\rho(Change\ In\ Mobility,\ College.Grad) =$ -0.42. These correlations become more negative with change in mobility as the education level increases. This is in aligment with my expectations and thus provides a good basis to build up an emirical model based on this.

Since I have not been living in the UK during the pandemic, I was not sure of the exact date of shelter-in-place advising. I spoke with family and they recall it being around Mid-March. Plotting the change in mobility over a two month period for a random selection of five counties in the UK (see **Figure 1**) consistently produced results that show the average date at which mobility drops significantly below the baseline occurs around March 14. I use this date to group my data to pre-announcement mobility and post-announcement mobility. I calculate the mean for before and the mean for after and use the difference between these two values as being my dependent variable: average change in mobility.

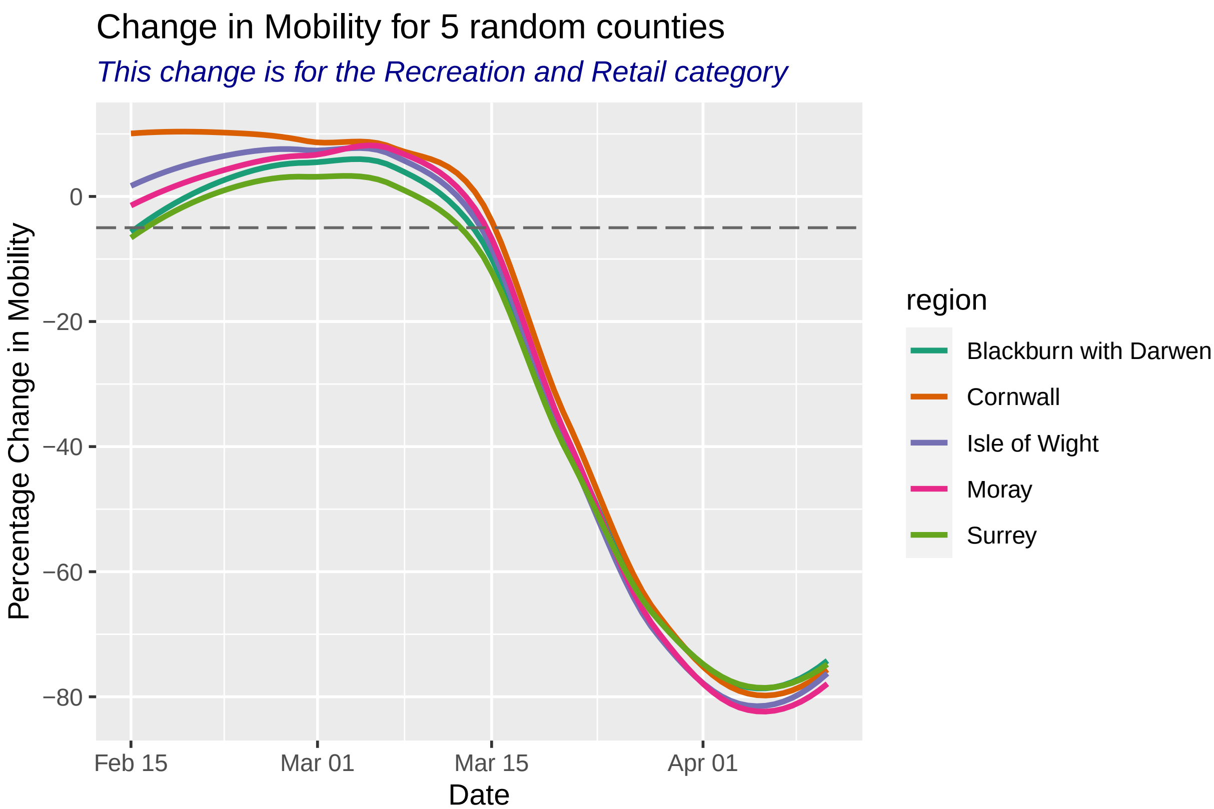

Figure 1

Empirical Framework

Model 1:

$\Delta Mobility_i = \beta_0 + \beta_1 Not.Identified + \beta_2 Grade9 + \beta_3 Grade11 + \beta_4 Grade12 + \beta_5 College.Grad + u_i$

I estimate this model using regular OLS. The coefficients I would expext would be most positive for Not.Identified and/or Grade.9 and become more negative as we move from Grade.11 to Grade.12 to College.Grad. Since I choose to not include the education category "No.Qualifications" the coefficients should be interpeted as follows. Take β_4. β_4 says that if an additional 10% of the population of a county have High-School diplomas relative to having No Qualifications then, ceteris paribus, the average percentage change of mobility for that county will be $= \beta_4$.

The results of the regression are that the two highest levels of education, Grade.12 and College.Grad, are the only regressors with negative coefficients. This aligns, generally, with my prediction that the more educated somebody is, the more their mobility will decrease (more obedience of shelter-in-place guidelines). However, it does not support the monotonic relationship that I supposed existed, particularly that those with college degrees would respond most cooperatively with government policy.

I suspect that College.Grad, and then subsequently the other qualification regressors, are endogenous. A source of this endogeneity is income. Specifically, I believe that income is biasing College.Grad. Those who

	Model 1
(Intercept)	-68.41 ***
	(p = 0.00, se = 11.27, CI = [-90.89, -45.93])
Not.Identified	66.58 **
	(p = 0.01, se = 23.61, CI = [19.51, 113.65])
Grade.9	58.46 *
	(p = 0.02, se = 23.51, CI = [11.57, 105.34])
Grade.11	28.29
	(p = 0.40, se = 33.59, CI = [-38.69, 95.28])
Grade.12	-58.23 *
	(p = 0.04, se = 27.59, CI = [-113.25, -3.21])
College.Grad	-13.96
	(p = 0.17, se = 10.18, CI = [-34.26, 6.35])
N	77
R2	0.40

*** p < 0.001; ** p < 0.01; * p < 0.05.

have a college education earn more, on average, than those who don't have a college degree. In 2013 in the UK, those with the equivalent of College.Grad earned 66% more than those with the equivalent of Grade.12. Suspecting that those with college degrees have higher income is not unreasonable. Then we can tell a story. As income rises, aversity to many changes decreases since more wealth can solve more problems. Thus, when panic spreads across a nation due to a pandemic, it would be likely that wealthier individuals are more comfortable and less concerned about outcomes. It is somewhat as though wealth privelege provides a sheet of ignorance to the external world. Life tends to be easier when one has money and consequently there is a more carefree attitude regarding the rules of the pandemic.

Results

Validating Endogeneity

Following from this we have correlation between income and the type of qualification one has *and* between income and the change in mobility. I will use social class as a metric for income.

```
## `geom_smooth()` using formula 'y ~ x'
```

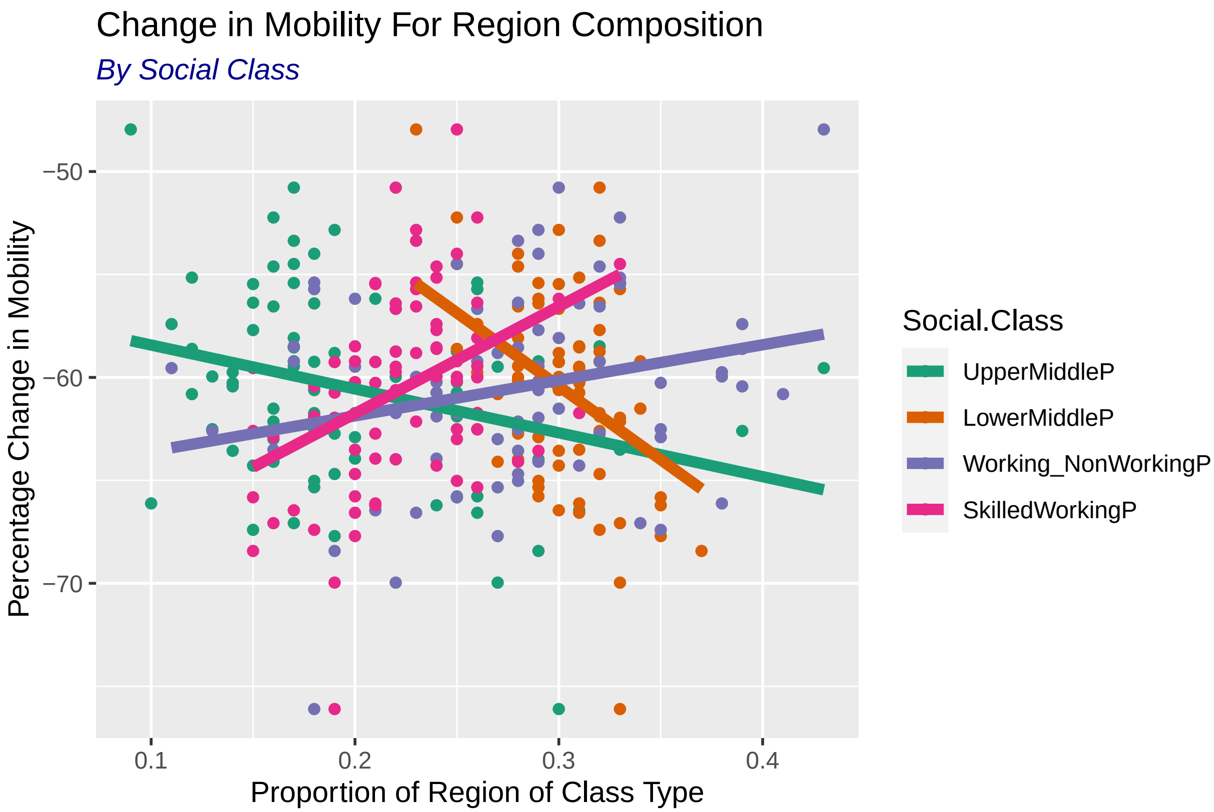

Figure 2

From this plot we can see increasing the proportion of UpperMiddle and LowerMiddle class individuals within a region is associated with decreasing levels of mobility. In addition to this, increasing the proportion of Skilled Working, Working, and Non Working Class individuals in a region leads to a smaller decline in mobility.

To illustrate this more clearly with a generalized split of higher-class vs lower-class:

```
## `geom_smooth()` using formula 'y ~ x'
```

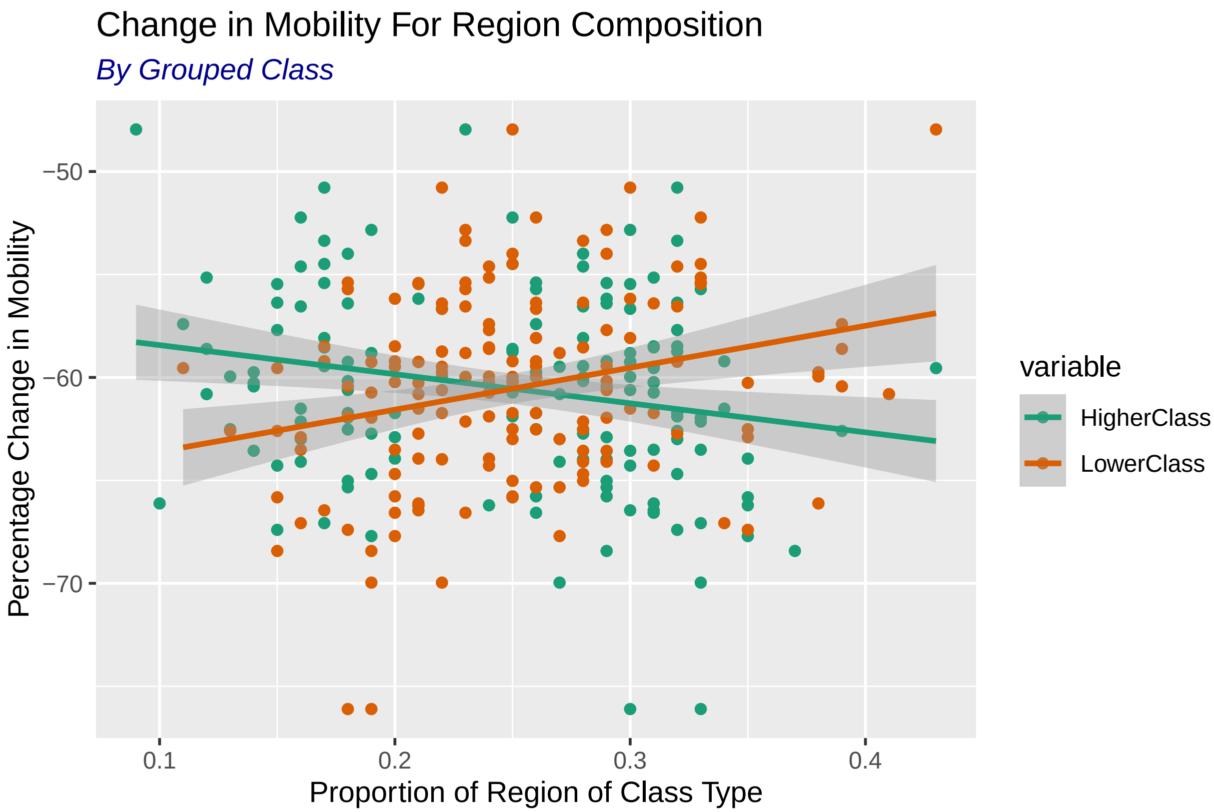

Figure 3

With this illustration we can clearly see that higher proportions of higher income in a region leads to a bigger decrease in mobility and higher proportions of lower income in a region leads to a smaller decrease in mobility.

To support this even further, $\rho(Higher\ Income,\ Change\ in\ Mobility) =$ -0.213, a negative correlation and $\rho(Lower\ Income,\ Change\ in\ Mobility) = 0.257$, a positive correlation.

UpperMiddleP	LowerMiddleP	No.Qualifications	Grade.9	Grade.11	Grade.12	College.Grad
	0.41	-0.87	-0.34	-0.02	0.17	0.92
		-0.47	0.04	-0.19	0.33	0.41
			0.4	-0.1	-0.37	-0.83
				-0.22	-0.64	-0.31
					-0.09	-0.21
						0.18

Finally, $\rho(Higher\ Income,\ College.Grad) = 0.17$ and $\rho(Higher\ Income,\ No.Qualifications) = 0.41$. We thus have an issue of endogeneity, though not what I originally anticipated. I proposed that the higher income was leading to carelessness around the national situation. What we actually see is that higher incomes are correlated with bigger mobility decreases and lower incomes are associated with smaller decreases in mobility. Moreover, taking note of the orange and green lines in **Figure 2**. The orange line represents the second highest income group in a region and the green line represents the highest income group. Observing that the orange line is steeper than the green, we suspect that the absencce of some metric for income in our model caused an overestimate of Grade.12, whose correlation is highest with LowerMiddleP. This overestimate likely extends to all variables.

Including the social grade in our model may ease the endogeneity problem but won't serve as an instrumental variable, since I believe that income does directly impact change in mobility. I simply hope to reduce the omitted variable bias. I will remove the "NonWorking_Working" variable since the sum of the four class variables is the total respondents to the census survey or, since I use their proportion relative to all responses,

1. By omitting one of the variables, we still satisfy MLR.3, no perfect collinearity.

Model 2:

$\Delta Mobility_i = \beta_0 + \beta_1 Not.Identified_i + \beta_2 Grade9_i + \beta_3 Grade11_i + \beta_4 Grade12_i + \beta_5 College.Grad_i + \beta_6 SkilledWorking_i + \beta_7 LowerMiddle_i + \beta_8 UpperMiddle_i + u_i$

	Model 1	Model 2
(Intercept)	-68.41 ***	-50.92 ***
	(p = 0.00)	(p = 0.00)
Not.Identified	66.58 **	49.99 *
	(p = 0.01)	(p = 0.05)
Grade.9	58.46 *	50.87
	(p = 0.02)	(p = 0.12)
Grade.11	28.29	-30.99
	(p = 0.40)	(p = 0.53)
Grade.12	-58.23 *	-34.94
	(p = 0.04)	(p = 0.33)
College.Grad	-13.96	-71.22 *
	(p = 0.17)	(p = 0.03)
SkilledWorkingP		24.67
		(p = 0.28)
LowerMiddleP		-38.82
		(p = 0.16)
UpperMiddleP		58.88 *
		(p = 0.02)
N	77	77
R2	0.40	0.49

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

After running this regression we see that the initial monotonic relationship I predicted is supported in the second regression. The coefficients from Grade.9 through College.Grad get more negative stepwise. College.Grad is now a statistically significant variable.

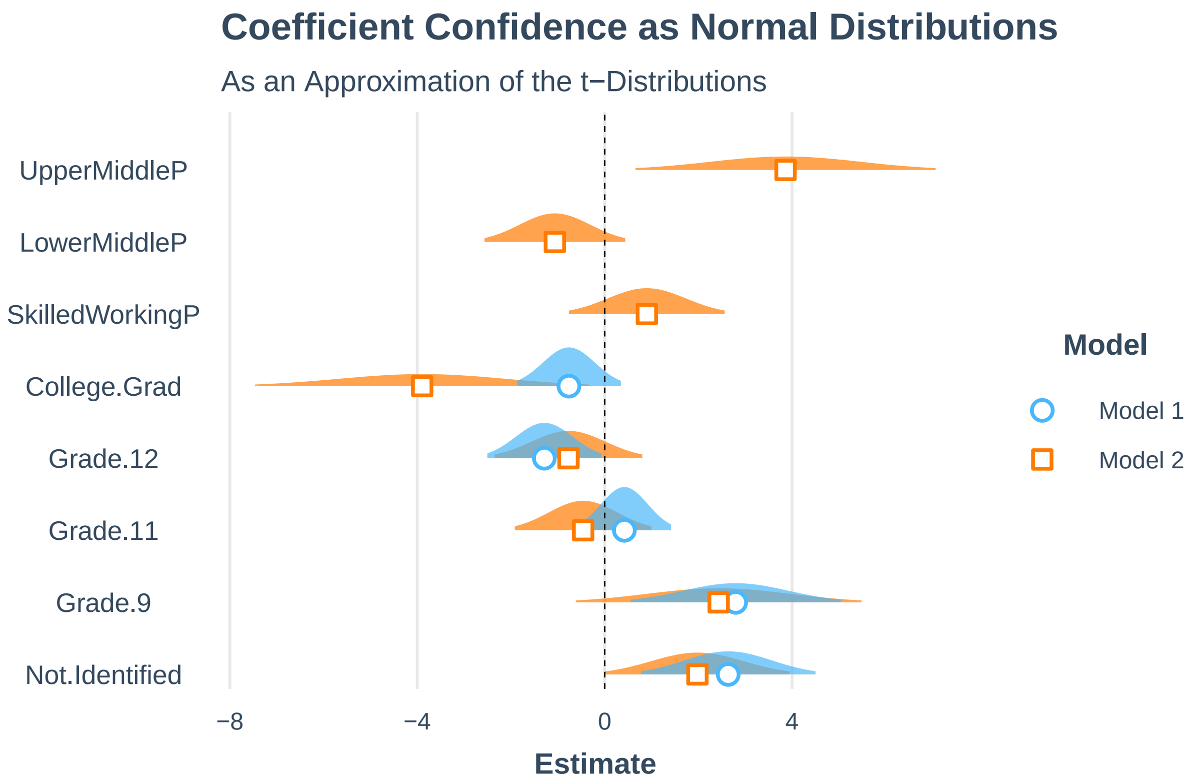

Figure 4

This illustration of the confidence intervals demonstrates the magnitude of significance of our regressorrs. The esimate values are scaled to allow coefficients, particularly those with wide intervals, to have similar heights for easier comparison. For example, we already noted that both College.Grad and Grade.11 variables are not statistically significant. Now we can see that the insignificance on College.Grad is more slight and occupies less of the distribution. For Grade.11, though, zero is far past the thin left tail and much more central to the coefficient uncertainty distribution. This also highlights the impact on mobility, from bottom to top we have increasing education level and a general decrease in mobility. With the second model we can notice different distributions and also see that we lose significance on a variable.

With the plot below we can see the difference in the distributions of mobility change for the two highest education levels. Level 3 is statistically significant at the 5% level whilst Level 4 is not, even at the 10% level. The plot illustrates the higher standard error in $\hat{\beta}_4$ (green) than in $\hat{\beta}_5$. Also illustrated is the greater effect on mobility change that increasing the proportion of high school graduates within a county has compared to the proportion of college graduates.

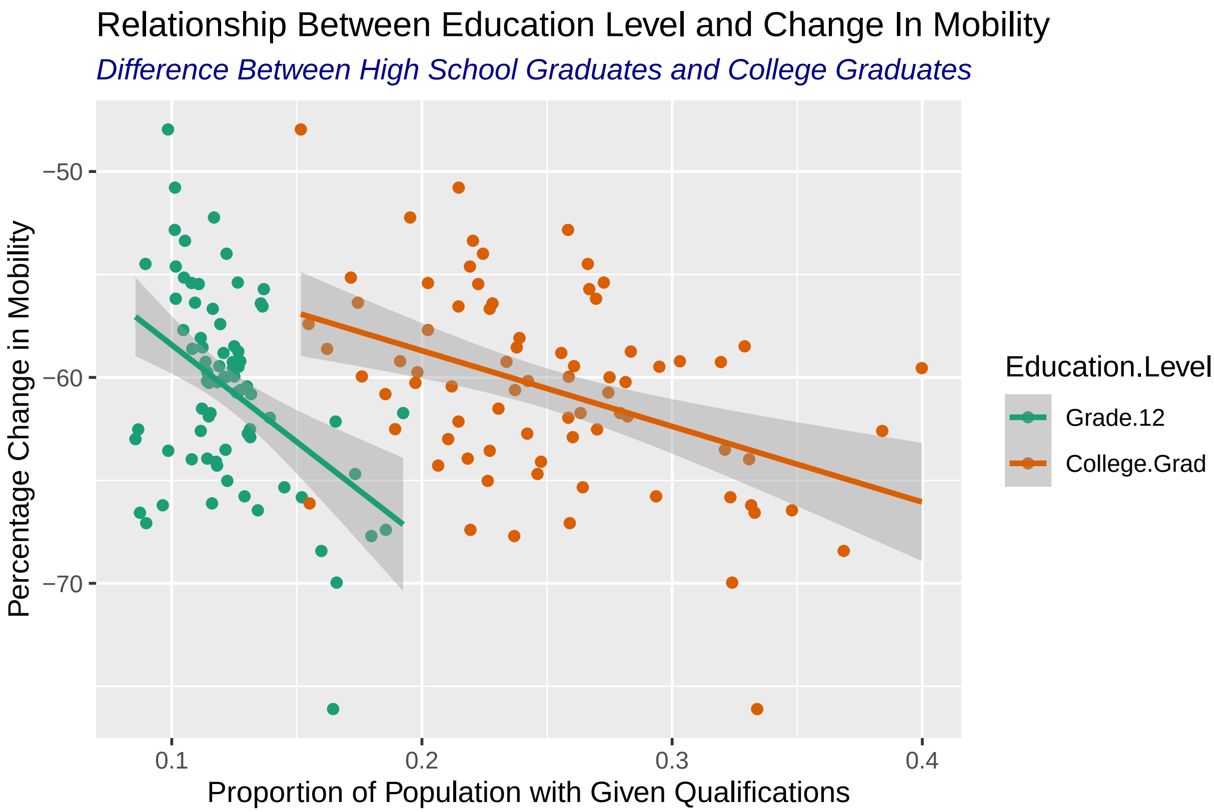

Figure 5

Another important thing to note here is the ranges of proportions. High school graduates sit between 10%-20% of the population whereas college graduates occupy 15%-35% of the population, on average. In the second model p-value for College.Grad = 0.03. This is statisitically significant and due to the large negative coefficient, -71.22, coupled with the average population share, I would argue that the impact of having a college degree on response to Covid-19 is economically significant. For Grade.12, a High-School graduate, it is not only statistiically insignificant (p = 0.33) but it is likely just economically insignificant also due to it's low population share. This said, in the second model the $R^2 = .49$ as opposed to $R^2 = .40$ in the first model. This says that the combination of regressor variables in the second model specification accont for 49% of the variance in the average change in mobility.

Discussion and Conclusion

I would argue that the esimates obtained in this paper are not BLUE. I believe that the model is linear in parametrs and MLR.1 is satisfied. Removing one of the categories of educataion and one of the social classes allowed to hold true MLR.3, No perfect multicollinearity. I doubt that MLR.2 holds true here. The data obtained from google on mobility is stronger in areas with better phone service asnd in areas where more people use their phones and in particular, use smartphones with google tracking enabled. Thus some areas become more heavily overrepresneted. Individuals also become overrepresented. For example, those with low income or no qualificaitons are less likely to have smartphones with google tracking and so these invidiualds become represented withina county. MLR.4, the Zero Conditional Mean assumption also doesn't seem like it is true here. I can think of Population, Average Age, House Size, Average Number of Friends, etc which would have correlation with both independent variables (income, education level) and change in mobility.

Finally, if nothing else, heteroskedasticity seems likely. Re-estimating using robust standard errors.

Figure 6

We see that the confidence interval for each independent becomes narrower.

One last issue is the small number of observations. If the google moility tracker could narrow down to the town level, my degrees of freedom would blow up from around 65 to over 500.

Regardless of lack of statistical / economic significance for *all* variables, I think that the results are nonetheless important and reflect a true situation occuring in the real world right now. These results are serious and imporant. The ideas here should be used to think about how to best connect with all members of a society to achieve the universal best outcome.

SOURCES

Office for National Statistics ; National Records of Scotland ; Northern Ireland Statistics and Research Agency. 2011 Census aggregate data. DOI: http://dx.doi.org/10.5257/census/aggregate-2011-1

Google LLC. 05/2020. "Google COVID-19 Community Mobility Reports". https://www.google.com/covid19/mobility/