

Question 1

The article I chose for this assignment was Eric Bettinger, Lindsay Fox, Susanna Loeb and Eric Taylor's article *Virtual Classrooms: How On line College Courses Affect Student Success* from *American Economic Review* [1]. The article models the impact of taking a course on line as opposed to a traditional in person classroom style has on the student's final outcome. In other words, does taking an online class result in a higher or lower grade for students.

The mathematical model that the authors use is

$$y_{ict} = \delta Online_{ict} + y_{ict\tau > t}\alpha + X_{it}\beta + \pi_c + \phi_t + \psi_{b(it)} + \rho_{p(it)} + \varepsilon_{ict}$$

The variables used in the model are:

Exogenous Variables

Variable	Description
$\delta Online_{ict}$	Indicator Variable who's value is 1 if the course taken on line and 0 if taken in person
$y_{ict\tau > t}\alpha$	Control for Students grades prior to term (t). This includes two primary variables: First,i is the grade point average in all courses taken online. Second is the GPA for the person in all courses taken in person.
$X_{it}\beta$	Students Gender and Age
π_c	Course fixed effects for each of the 750 courses
$\rho_{p(it)}$	for each of the 22 different degree programs
ϕ_t	nonparametric time trned over the 27 terms of data (4.5 years for an term of 8 weeks)
$\psi_{b(it)}$	fixed effects for student i home campus

Endogenous Variables

Variable	Description
y_{ict}	student outcome in terms of grade received. Grade received by student i in course c during term t

This model would by classified as a static model, as well as a linear model. We can classify this as a linear model since the regression for the model does not include any quadratic or polynomial terms. Further more, the model can be classified as deterministic model since the outcome (i.e. student's grade) is purely dependent on the inputs (endogenous variables) that are in the regression can not due to some inherent randomness that is inherent to stochastic models. This is a static model because the regression controls for time in several of its terms.

One variable that may be valuable to understanding the student outcomes of on line and in-person classes is the employment status of the student. Employment status could be a dummy variable that indicates whether or not the person is working full time or not, or it could be the number of total hours that a person works in a week. This variable may provide valuable insight into the number of hours an individual can dedicate to the class. For example, someone one working 40 hours a week will have far less time to dedicate to the class, and therefore master it's content, than someone that is not employed and has more time to dedicate to the subject.

References

- [1] Susanna Loeb Bettinger Eric P., Lindsay Fox and Eric S.Taylor. Virtual classrooms: How online college courses affect student success. *American Economic Review*, 107(9):2855–2875, 2017.

Question 2:

Write down a model of whether someone decides to get married

Model:

$$married = \beta_0 + \beta_1 age_i + \beta_2 ed_i + \beta_3 income_i + \beta_4 Divorced_i + \beta_5 NumChild_i + \beta_6 Partner_i + (\beta_7 Divorce_i NumChild_i)$$

In this model, our endogenous variable, *married*, is equal to 1 if person *i* gets married and equal to 0 if person *i* does not get married. In this model, we can see that all the exogenous variables can be generated or represented either by some data collection scheme or through simulation, including the dummy variables.

The key factors that influence the outcome are:

- *age* - the age of the individual getting married
- *ed* - the level of education attained by individual *i*, which in the model is represented by the number of years of education
- *income* - the income of individual *i*
- *Divorced* - If individual *i* has already been married (*Divorce* = 1) than it may influence their decision to remarry in the future
- *NumChild* - the number of children that individual *i* may have
- *Partner* - whether or not the individual is currently in a relationship (if individual *i* is in a relationship *Partner* = 1)

I chose these variables because first and foremost, they are measurable. In other words, there are data sets that exist or can be simulated to reflect current population trends to satisfy, and therefore test, the model. If the model were to use something that attempts to measure attitude or feeling the model would have less impact because such things are very difficult to measure let alone understand. Furthermore, I chose these variables because in many ways they reflect the preferences of individuals. For example, people with higher income may show more preferences to be married since they can afford a wedding and to support their spouse partner, where as age and education may go against the probability of getting married since younger people who have less education (i.e. a young person with a bachelor degree versus a thirty something with a masters). Finally, two of the most important factors that I brought into my model are *Divorced* and *NumChild*. While both variables exist on their own, I also included an interaction term in the regression. I inserted the interaction term to more accurately reflect a person's preferences and therefore likelihood to get married. For example, while a Divorced individual with no kids may still get remarried, if that divorced individual had kids that probability would be even stronger. In many ways, these variables reflect the preference (i.e. utility) of an individual that cannot be directly measured. To demonstrate, a divorced individual with kids would be more likely to get married because that individual would want to provide shelter and security for their children, and getting married is a means of accomplishing that.

There are many ways you could do a preliminary test to see if the factors are significant. First, one could use the General Social Survey or Census Record (e.g. IPUMS) to simulate the model to test which factors are and are not significant in the probability of getting married. Second, if the data is not attainable through the GSS or census records, it could easily be collected through a survey. Since the information is not (overly) sensitive people may be more willing to respond to the questions that collect these data points. Once the data is collected you could conduct statistical analysis to test which of the variables are the most influential on the outcome variable, *married*.