PSet1 - MACS 30150

Josh Laven

1/10/2020

Question 1

a) Find a theoretical or statistical model from a recently published article (no earlier than 2015) in either the American Economic Review, Econometrica, or the Review of Economic Dynamics.

b) Give a detailed citation of the article.

Title: Family Ruptures, Stress, and the Mental Health of the Next Generation

Authors: Petra Persson, Maya Rossin-Slater

Journal: American Economic Review Vol. 109, No. 4-5, April 2018 (pp. 1214-52)

Summary: This paper explores the effect of maternal stress from familial deaths during preganancy on mental well-being of the child, as measured by intake of prescription medication in adolesence and adulthood. The sample in the experiment includes children whose mother loses a family member in the nine months after a child's conception or the year after the child's birth.

c) Write down the mathematical or statistical model (write the equations).

The general equation to estimate the effect of maternal stress from familial death on the child is given by:

$$y_i = \gamma RelativeDeath + \kappa x_i + u_i$$

where y_i is the outcome of interest, x_i is a vector of observable covariates, and u_i are unobservable characteristics. The γ will thus give the impact of both pre-natal and post-natal consequences for children whose mothers experienced a death.

In order to isolate the impact of the death occuring during gestation, as opposed to after birth, the authors refine their model:

$$y_i = \sigma 1[c \leq RelativeDeath < b]_i + \eta x_i + \epsilon_i$$

where $[c \leq RelativeDeath < b]_i = 1$ if the familial deathoccurs during gestation. The σ in this refined equation captures the effect of the death occuring during gestation relative to the effect of the death occuring after birth.

As the researchers note in the paper, exposure to familial death actually reduces the gestanational age of the child. Therefore, there is a correlation between the likelihood of a familial death and the length of pregnancy. This introduces bias into the model. To correct for this, the researchers introduce e_b, the expected date of birth. This is calculated from $c + 280$, and is therefore predetermined and unaffected by the when the death actually occurs. This is added to the model to give:

$$y_{yimp} = \beta_0 + \beta 1 \times 1[c \leq RelativeDeath < e_b]_{iymp} + \psi_y + \phi_m + \rho_p + \beta_2 x_i + \nu_{iymp}$$

where y_{yimp} is the outcome for child i conceived in year (y) and month (m), with mother residing in municipality (p) the year prior to conception. ψ_y, ϕ_m and ρ_p are fixed effects for year, month and municipality, respectively. x_i is a vector of characertistics including mother's age at conception, education, deceased relative age, etc. Here, $1[c \leq RelativeDeath < e_b]_{iymp}$ takes the value 1 if the death occurred during the child's expected date of birth and 0 otherwise.

d) List which variables are exogenous (determined outside the model, assumed) and which variables are endogenous (determined inside the model, the output of the model)

In the final iteration of the model:

$$y_{yimp} = \beta_0 + \beta 1 \times 1[c \leq RelativeDeath < e_b]_{iymp} + \psi_y + \phi_m + \rho_p + \beta_2 x_i + \nu_{iymp}$$

:

The exogenous variables are y_{yimp}, $1[c \leq RelativeDeath < e_b]_{iymp}$, ψ_y, ϕ_m, ρ_p, ν_{iymp}, and x_i. The endogenous variables are β_0, β_1, and β_2

e) Classify the model as static vs. dynamic, linear vs. nonlinear, deterministic vs. stochastic.

- The model is static- it is time invariant
- The model is linear- the regression is linear in parameters
- The model is deterministic- the endogenous variables are fully determined by the inputs with no randomness

(f) List a variable or feature that you think the model is missing that might be valuable.

The model codes in utero exposure to maternal stress due to a familial death as a binary variable. It would be interesting to add an additional variable that captures gestational age. This would show the effects of familial death at different points in pregnancy. It seems reasonable to assume that the longer the exposure to stress, the more severe the effects.

Question 2

(a) Write down a model of whether someone decides to get married.
(b) At least one of the dependent endogenous variables (output of the model) must be 1=get married or 0=not get married.
(c) Make sure that your model is a complete data generating process. That is, you could simulate data from your model given all the parameters and relationships.

A logistic regression:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 Religious + \beta_2 Age + \beta_3 Coupled + \beta_4 RelLength + \beta_5 MarRate + u_i$$

where p is our predicted outcome (the probability of ever getting married); *Religious* is a binary variable that is coded 1 is an individual is religious and 0 otherwise; *Age* is the age of the individual; *Coupled* is a binary variable for whether or not an individual is in a relationship; *RelLength* is the length of an individual's current relationship; $MarRate$ is the marriage rate in the country where the individual lives; and u_i is the error term.

If we run this regression with sample data, we will get values for β_{0-4}. When can then plug data from new individuals into the model to predict the probability that those individuals ever get married. The logistic regression will return a probabilty between 0 and 1. We can determine the probablity threshold of .5, so that for any output of the model $\geq .5$, the individual will get married and any output $< .5$ the individual will not get married

(d) What do you think are the key factors that influence this outcome?

I think cultural factors heavily influence whether or not someone decides to get married. Marriage rates are much higher in religious communities and also fluctuate between countries.

(e) Why did you decide on those factors and not others?

Anecdotally, these factors seem to influence whether someone gets married or not. I added age because if someone is older and not married, they are probably very unlikely to get married. If someone is currently in

a relationship, they are more likely to get married, and the length of the relationship is most likely positively correlated with likelihood of marriage. *Religious* and $MarRate$ are cultural factors which, as noted in (d), I believe are very important in an individual's decision to get married.

(f) How could you do a preliminary test whether your factors are significant in real life?

We can obtain a sample of data and compare the means of each covariate between married and unmarried individuals. For example, if we have a sample of 1000 indiviudals, we can take the mean of those that identify as religious vs those who do not for both the married and unmarried individuals. If hypothesis is correct, the mean of the religious variable should be higher for married individuals than unmarried individuals. We can then repeat this experiment for all four of the other covariates.