

Problem Set #1

MACS 30100, Dr. Evans

Bobae Kang

1. Classify a model from a journal.

(a), (b) Citation: Holbein, J. B. and Hillygus, D. S. (2016), Making Young Voters: The Impact of Preregistration on Youth Turnout. *American Journal of Political Science*, 60: 364-382. doi:10.1111/ajps.12177

(c) The paper uses two models, one for the upper bound of the preregistration effect (the difference-in-difference model) and the other for the lower bound (the lagged model).

$$(1) Y_{it} = \lambda_0 + \lambda_p P_{st} + \lambda_\alpha \alpha_s + \lambda_\delta \delta_t + \lambda_\gamma \gamma_{st} + \lambda_X X_{it} + \epsilon$$

$$(2) Y_{it} = \lambda_0 + \lambda_p P_{st} + \lambda_Y Y_{s,t-2} + \lambda_\gamma \gamma_{st} + \lambda_X X_{it} + \epsilon$$

(d) Endogenous variable: Y_{it} (whether the individual reported voting).

Exogenous variables: P_{st} (whether the respondents state had a preregistration law in effect), α_s (the state fixed effects; first model only), δ_t (the year fixed effects; first model only), γ_{st} (the full set of interactions between the two for state-specific year effects), X_{it} (a matrix of time-varying controls), and $Y_{s,t-2}$ (lagged turnout aggregated to the state level).

(e) Both models (1) and (2) are static, linear and stochastic.

(f) Perhaps it is important to account for the neighbor effect among states, i.e. the known voting rates of the neighboring states. An individual may evaluate the importance of voting based on not only the factors already included in the current model but also social pressure originating from the known voting behavior of the nearby states. This is based on an observation that, first, Americans often identify themselves with the state of their residence and, second, US states are often considered to be in a competitive relationship with one another. In addition, voter participating is generally presented as socially desirable. If these observations are true, it is reasonable to assume that individuals may have a sense of competition which tells it is desirable for her state to achieve a higher voter turnout than other states.

2. Make your own model.

$$(a) Y_i = \beta_0 + \beta_P P_i + \beta_G G_i + \beta_I I_i + \beta_F F_i + \beta_M M_i + \beta_H H_i + \epsilon$$
$$H_i = \gamma_0 + \gamma_Y Y_p + \gamma_C C_i + \gamma_D D_i + \epsilon_H$$

Y_i is the predicted lifespan in years for an individual (endogenous); P_i is the popularity indicator of the individual; G is the musical genre of the individual; I_i is the income level of the individual; F_i is whether the individual is biologically female; M is the marital status of the individual; H_i is the health indicator of the individual (endogenous); Y_p is the average age at which the individuals' parents passed away;

C_i is whether the individual has been a life-long cigarette smoker; D_i is whether the individual has been an drug addict.

(d), (e) The key factors for this model are I_i , F_i , M_i , and H_i as well as other variables required to calculate H_i , namely, Y_p , C_i , and D_i . I_i is important because higher income allows for the individual to take better care of the self. F_i is important because in general the life expectancy of females is known to be higher than that of males. M_i is important because one's committed romantic partner, which generally coincides with a marital partner, is known to offer health benefits to the individual. H_i and other variables required to calculate H_i are important because one's life expectancy is closely related to one's health and these variables collectively provide a measurement for health. The relevance of all variables used to calculate H_i on one's life expectancy, i.e. the effect of the longevity of the individual's parents, smoking and drug addiction on the individual's life expectancy, is supported by the empirical evidence from a number of scientific studies.

On the other hand, I do not consider P_i and G_i as key factors to this model because they have little known effect on musicians' lifespan. Even if they are found to have significant effect, they are most likely intermediary factors that mediate the effect of other factors. For example, a subculture of a specific musical genre may be correlated with the chance that an individual is a heavy smoker. Therefore, I suspect that these two factors will appear insignificant after controlling for other variables.

(f) In order to test the significance of these factors, I can use a random sample of musicians who have passed away to fit the model. Since it is a straightforward linear regression model, testing the statistical significance of each parameter will be a relatively easy task, with an assumption that the true data generating process is linear and each variable is normally distributed as well as i.i.d.