# Problem Set 1

Anqi Zhou
anqiz1@uchicago.edu

## 1. Classify a model from a journal.

a. I choose the article "Family Health Behavior" from American Economics Review (Sept. 2019).

b. Fadlon, I. & Nielsen T. H. (2019). Family Health Behaviors. *American Economic Review 2019, 109(9): 3162–3191.*

c. The equation is: $y_{i,t} = \alpha + \beta treat_i + \sum \gamma_r \times I_r + \sum \delta_r \times I_r \times treat_i + + \lambda X_{i,t} + \varepsilon_{i,t}$
Basically, this article tends to investigate the dynamic effects of severe health shocks on family's spend on preventive care and other health-related behavior. (Fadlon & Nielsen, 2019) So in the equation, y is the outcome of particular family at time t; treat identifies whether the family is in treatment group or not; $I_r$ are time indicators relative to the shock year; $X_{i,t}$ are a series of control variables, including age fixed effects, calendar year fixed effects, gender and education.

d. In the above equation, $y_{i,t}$, $treat_i$, $I_r$, $X_{i,t}$ are all exogenous variables, and $\alpha$, $\beta$, $\gamma_r$, $\delta_r$, $\lambda$, $\varepsilon_{i,t}$ are all endogenous variables, including which $\delta_r$ are the researchers' parameters of interest, indicating the treatment effects in period $r$ relative to period -1.

e. This model is a dynamic difference-in-difference model. According to Fadlon & Nielsen (2019), they estimate the treatment effect from change in the outcome difference between the two groups over time.

f. As far as I'm concerned, The missing factor that may be valuable is the wealth level of the families. As we can guess, if the family is rich, the change in outcome difference across time will be larger. But if the family is relative poor, they can't put too much money on preventive care, even though they now become very risk-adverse and are very willing to invest.

## 2. Make your own model

a. I try to use logit model to analyze this topic.
My equation is:

$$y_t = \exp(X_t'\beta) / (1+\exp(X_t'\beta)),$$

$X_t$ is a vector of all my explanatory variables and $y_t$ denotes whether the individual decides to marry or not at time t. More precisely, $X_t$ includes individual's income at time t, age at time t, whether individual's parent are divorced or not, whether individual is at school at time t, whether individual have a boy/girlfriend at time t.

b. We can first gather some data to run the regression of logit model, in order to estimate all the parameters (all the $\beta s$). Given the values of all the parameters, we can generate the probability of

entering into marriage by applying the logit model. And we assume the output for marriage is "yes" when the estimated probability is larger than 50%, otherwise the output should be "no".

c. As we explained above, this model is a complete data generating process. We can simulate the output whether individual decides to marry or not, given all the relationships and parameters.

d. As we can see, the increase in residents' income and the change in consumption pattern has made people pay more and more attention to economic foundation of marriage. Nowadays, reasonable income has increasingly become a necessary condition for the foundation of marriage. (Mu & Xie, 2011) Obviously, age is also an important consideration. When people are getting closer to around 30 years old, they tend to think about marriage, because it seems that now it is the proper time to get married. This even becomes a tradition that people around you will remind you of marriage when you're getting 30. (Yu & Xie, 2013) I innovatively add the dummy variable reflecting whether individual's parents are divorced or not. It's from my observation. People whose parents are divorced tend to feel insecure and many of them have a feeling of homeless, so they are more likely to make the decision of getting married and forming their own family. I also add the dummy variable reflecting whether individual is at school or not, because compared with graduate marriage, students who marry at school often need to overcome more academic and family difficulties. (Thornton et al., 1995) The last variable indicates whether the individual have a boy/girlfriend at time t. Most of people consider getting married only when they have boy/girlfriends. This variable should be very significant, and it's meaningful. Some people have strong marriage willingness, but they don't get married because there's no potential object to marry with. Some people have boy/girlfriends but they still won't get married.

e. I select variables, mainly referring to the models in Mu & Xie (2011) and Yu & Xie (2013). I also review other related literatures and summarizing most of similar and significant variables in these literatures. I also consider about adding innovative variables based on my personal experience, and try to find out whether it makes sense.

f. As for preliminary test, we can find a related database, even it is not large enough. We just use this database to run the regression simply, and thus have a broad understanding of the variables —- whether they are significant or not, how much they can explain the marriage output in all ($R^2$), etc. Then according to our preliminary test, we can adjust the selection of variables. The preliminary test database for this problem, I think, can be the CGSS (Chinese General Social Survey) data.[1]

[1] For more details, please refer to http://cgss.ruc.edu.cn.

## References

Fadlon, I. & Nielsen T. H. (2019). Family Health Behaviors. *American Economic Review 2019, 109(9): 3162–3191.*

Mu, Z. & Xie, Y. 2011. "Marital Age Homogamy in China: A Reversal of Trend in the Reform Era?" *PSC Research Report 11-742, http://www.psc.isr.umich.edu/pubs/abs/7324.*

Yu, J. & Xie, Y. (2013). Social Changes and Changes in Factors Influencing First Marriage. *Sociology Research, 4, 1-25.*

Thornton, A., W. Axinn & J. Teachman 1995, "The Influence of School Enrollment and Accumulation on Cohabitation and Marriage in Early Adulthood." *American Sociological Review 60(5).*