Perspective HW1

Ruxin Chen

1 Classify a model from a journal

a) Paper from *American Economic Review*

b) He, Guojun, and Shaoda Wang. "Do College Graduates Serving as Village Officials Help Rural China?." American Economic Journal: Applied Economics 9, no. 4 (2017): 186-215.

c)

$$y_{it} = \alpha \times CGVO_{it} + X'_{it} \times \beta + \rho_t + \mu_i + \epsilon_{it} \tag{1}$$

where y_{it} is an outcome of interest for village i in year t, $CGVO_{it}$ is a dummy indicator, which equals 1 if village i in year t has a $CGVO$ (College Graduate Village Officials), and 0 otherwise. X_{it} is a set of time-varying control variables, including precipitation and temperature in each village-year pair. ρ_t is a time effect common to all villages in period t, μ_i is a time-invariant effect unique to village i, and ϵ_{it} is a village time-varying error distributed independently of μ_i and ρ_t.

d) exogenous variables:

- $CGVO_{it}$: dummy variable indicating whether the village i in year t has a $CGVO$
- X'_{it}: time-varying control variables for village i in year t including precipitation and temperature
- ρ_t: time fixed effect for year t
- μ_i: individual fixed effect for village i

endogenous variable: y_{it}: outcome of interest for village i in year t.
The paper assesses the contribution made by $CGVOs$ from multiple indices under the form of model (1). For different models, the endogenous variables y_{it} represent the different indices including the number of formally registered poor households, the number of formally registered residents with disabilities, the number of subsidized residents and the proportion of poor-quality rural housing.

e) The model is dynamic, linear and deterministic.

f) One particular important factor missing from the model is the economic growth. As all the outcomes of interest are closely related to economic growth, leaving it out of the model will cause significant bias. Since the data includes 255 villages from 19 provinces with different economical environment, the economic growth is considered to be varying with time and villages, hence, not captured by the original model. Though it is unlikely to include the village-level economic growth (It is dubious that economic data will be collected and released at village level), it is not hard to include the economic growth rate of the city that each village belongs to.

2 Make your own model

a) - c)
$$P(Y_i = 1 | \mathbf{X}_i) = \frac{1}{1 + e^{-\mathbf{X}_i \boldsymbol{\beta}}} \tag{2}$$

where Y_i is a dummy variable which equals 1 if individual i decides to get married, equals 0 if otherwise. $\mathbf{X}_i = \{1, \ Male_i, \ Race_i, \ income_i, \ height_i, \ education_i, \ age_i\}$ is a set of independent/exogenous variables: $Male_i$ is a dummy variable indicating whether individual i is a male, $Race_i$ is a categorical variable defined as follows:

$$Race_i = \begin{cases} 0 & \text{if individual } i \text{ is White} \\ 1 & \text{if individual } i \text{ is Black or African American} \\ 2 & \text{if individual } i \text{ is American Indian or Alaska Native} \\ 3 & \text{if individual } i \text{ is Asian} \\ 4 & \text{if individual } i \text{ is Native Hawaiian or Other Pacific Islander} \end{cases}$$

$income_i$ is the annual log income in US dollars, $height_{it}$ represents the height of individual i in feet, $education_i$ represents the years of education for individual i and age_i stands for the age.

d) The major factors are considered to be $income_i$, age_i and $education_i$.

e) I decide to include these factors to my model based on the paper "A theory of marriage: Part II" by Gary Becker, where he states that "each person tries to find a mate who maximizes his or her well-being, with well-being measured by the consumption of household-produced commodities." He further argues that "the gain from marriage compared to remaining single for any two persons is positively related to their incomes, the relative difference in their wage rates, and the level of non-market-productivity variables, such as education or beauty". Though his paper attempts to study marriage from a macro-level and our model targets on

individual marital decision, the variables he list still provide some insights for determining whether a person wants to get married. A person is more likely to get married if he/she is old, since he/she might need to be taken care of. Unbalanced gender ratio among the population might increase difficulties for a particular gender to get married. For example, if the marriageable age female is much less than the male, then the probability for male to get married is low. It is presumable that high income/education level individuals are less likely to get married in seek of financial dependence. Height is also an important factor of finding a mate, hence affecting the marriage decision.

f) A preliminary test can be applied on a small (but adequate for inference) subset of the data that we are going to use for the main results or some census data available online. We can run a regression of the form (2) and estimate the parameters β in the model. If the parameters are statistically significant, we have some evidences that the factors we specify in the model are relevant. Meanwhile, we can also get some insights by looking at the summary statistics for these variables for the married group and single group.