PS1

Andi Liao

1. Classify a model from a journal

a)

I found a statistical model about the influence of demands from high-income consumers on wage structure of United States in the American Journal of Sociology.

b)

The citation is as follows:

Wilmers, N. (2017). Does Consumer Demand Reproduce Inequality? High-Income Consumers, Vertical Differentiation, and the Wage Structure. *American Journal of Sociology, 123*(1), 178-231.

c)

The model in the article is the conditional variance of wages:

- $log\left(\sigma^2_{irpt}\right) = \lambda_1 d_{rpt} + \boldsymbol{\lambda_2} \boldsymbol{x'_{rpt}} + \boldsymbol{\lambda_3} \boldsymbol{x'_{irpt}} + \zeta_{1r} + \zeta_{2t}$

  d_{rpt}: the worker's industry-region-year dependence on high-income consumers

  x'_{rpt}: other industry-region-year-level covariates

  x'_{irpt}: individual industry-region-year-level covariates

  ζ_{1r}: region fixed effects

  ζ_{2t}: year fixed effects

  σ^2_{irpt}: estimated residual variances in wages, which is built on the conditional mean regression of wages:

  - $ln\left(y_{irpt}\right) = \beta_1 d_{rpt} + \boldsymbol{\beta_2} \boldsymbol{x'_{rpt}} + \boldsymbol{\beta_3} \boldsymbol{x'_{irpt}} + \alpha_{1r} + \alpha_{2t} + \alpha_{3p} + \sigma_{irpt}$

    y_{irpt}: the individual worker's industry-region-year hourly wages

    α_{1r}: region fixed effects

    α_{2t}: year fixed effects

    α_{3p}: industry fixed effects

d)

For this model, σ^2_{irpt} is an endogenous variable as estimated residual variances in wages is the output of the model. d_{rpt}, x'_{rpt}, x'_{irpt}, ζ_{1r} and ζ_{2t} (*i.e.*, the worker's industry-region-year dependence on high-income consumers, other/individual industry-region-year-level covariates, region/year fixed effects) are exogenous variables because they are determined by the dataset, rather than by the model.

e)

The model is static, given that the time is irrelevant in this model. The model is also linear, as the log-transformed σ^2_{irpt} has a linear relationship to the independent variables. The model is stochastic as well, because it needs an error term, which is a random variable, to make the equation stands.

f)

The model has taken industry region and year into consideration. However, I think that it would be better to add a variable describing the relevant policy factor.

Policy has great impact on the wage structure.

On the one hand, workers receiving different levels of wages pay various levels of tax rates. For instance, wages of group A have higher mean and larger variance than group B. After paying taxes, wages of group A still has higher mean, but the variance becomes smaller than group B. If different tax rates is ignored, we might draw the conclusion that group A has higher wage inequality, which will be changed after calculating taxes. Therefore, by considering different tax rates, the measurement of wages will be more accurate, and it will be easier to have a clear picture of wages structure.

On the other hand, the corporate income will also be affected by state policy, even national policy. Suppose that the national government are now encouraging the development of high-tech companies, and local governments have their own definitions of “high-tech” companies. Based on their understanding, state governments publish local policy to help related corporates, which might influence how corporates treat their customers and employees. However, this high-tech policy factor is not clearly included in the covariates variables, or fixed effects variables.

Therefore, taking policy factor into account might help improving the current model.

2. Make my own model

a) b) c)

The model of whether someone decides to get married:

- $y_{married} = \beta_0 status + \beta_1 age + \beta_2 gender + \beta_3 education + \beta_4 income$
  $+ \beta_5 personality + \epsilon$

  $y_{married}$: 1 = get married, 0 = not get married

  $status$: the relationship status of the person, such as dating, in relationship, engaged, divorced once, *etc.*

d)

I think the most important factors are status and age.

- $status$:

  Personally, I believe that it is the most determining factors in the model. Those who have engaged are much more likely to decide to get married, compared to people who are still looking for their dates. After all, getting married involves two individuals, and making a promise, that is, engagement here, towards each other can help people make their decisions of getting married.

- age:

  Age is another crucial factor in this model. There is a certain period of time, when most people feel that it might be appropriate and necessary to start a family. Typically, people aged 25 to 35 have higher probabilities to decide to get married. Before this period, people are too young to take responsibilities for another person’s happiness. After this period, people might easily lose passion and courage to switch to a new life. Therefore, the timing really matters when someone decides to get married or not.

e)

The inspiration comes from a lecture about intimate relationship in developmental psychology during my undergraduate study. I combined the existed theories in psychology and my personal understanding to build this model. I choose these variables because they contribute more to the decision of getting married or not than others, and I will absolutely consider these factors when I make my own decision.

The first two factors have been explained in part d). Gender should be included, as women and men have different concerns when they think about marriage. Usually, women might desire more to get married than men. Education and income are also important in this model, as people with higher education background and income tend to be more cautious when making the decision of getting married. As for personality, it also plays a part in this decision. Certain personality traits, including openness and conscientiousness, positively contribute to the likelihood of getting married, while other traits don't.

In sum, I decide to choose $status, age, gender, education, income, personality$ into the model, given their universal influence when people decide whether to get married.

f)

I plan to collect some data in two ways.

- Open-source datasets:

  I intend to collect the demographical information from open-source datasets, and extract the marital status, age, gender, highest education, income and personality test results to run the model.

- Simple survey: Different from the first method, I can also collect some data from people around me. Participants will provide their demographical information and finish a personality test. Then I can use the data to test my model.

I will be able to calculate the values of β_0 to β_5 using the data I collect, and perform the significant test of regression coefficient β. If any β is significantly different from 0, then the corresponding independent factor should be viewed as a significant one.

Interviewing might be a useful supplement here. I plan to interview two groups of people. The first one is married group, and the second one is not married group. I will collect their data for variables in the model and listen to their opinions about their decisions of getting married or not. Then I can compare the group differences on these independent variables. If these two groups are really different on $status, age, gender, education, income, personality$, it might be the signal that this variable is significant.