

Problem Set #1

MACS 30100, Dr. Evans

Shuting Chen

Problem 1. Classify a model from a journal.

Parts (a) - (c).

Detailed citation of the article:

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg. 2017. "Peer Effects in the Workplace." *American Economic Review*, 107(2): 425-56.

Description of the theoretical model from the selected article:

The authors develop a principal-agent model of unobserved worker effect in which peer-induced productivity effects arise through social pressure and knowledge spillover and translate into peer-related wage effects.

Specifically, they consider a firm of N workers and the production function for worker i is defined as

$$f_i = y_i + \varepsilon_i = a_i + e_i (1 + \lambda^K \bar{a}_{\sim i}) + \varepsilon_i, \quad (1)$$

where y_i is worker i 's productive capacity, which depends on individual ability a_i , individual effort e_i , and average peer ability (excluding worker i) $\bar{a}_{\sim i}$. Besides, the knowledge spillover is captured by λ^K .

Apart from modelling individual output f_i , the authors construct the following disutility function by including cost of effort and peer pressure

$$c_i = C(e_i) + P(e_i, \bar{f}_{\sim i}) = k e_i^2 + \lambda^P (m - e_i) \bar{f}_{\sim i} \quad (2)$$

where λ^P and m can be treated as both the "strength" and the "pain" from peer pressure and $\bar{f}_{\sim i}$ is the average peer output (excluding worker i).

Part (d).

Exogenous variables: a_i , e_i , $\bar{a}_{\sim i}$ and $\bar{f}_{\sim i}$

Endogenous variables: f_i and c_i

Part (e).

The model is static since it does not have any time index for each variable. Moreover, due to the nonlinear term $e_i \bar{a}_{\sim i}$ in function (1) and quadratic cost of effort in function (2), the model is nonlinear. Besides, this model is stochastic since it has ε_i , which is random productivity shock that is beyond the workers' control.

Part (f).

Presumably, one could include a variable representing average peer effort (excluding worker i) in the production function for each worker (i.e. $\bar{e}_{\sim i}$). This should be reasonable since workers may increase their production when they perceive that others work harder than themselves.

Problem 2. Make your own model.

Model.

To model whether someone decides to get married, which is a binary response variable, I am going to use the following logistic regression model

$$P(y_i = 1|\mathbf{x}) = G(\beta_0 + \beta_1 r_i + \beta_2 w_i + \beta_3 a_i + \beta_4 pm_i + \beta_5 h_i + \beta_6 eth_i) \quad (3)$$

where $G(\cdot)$ is the cumulative distribution function (CDF) for a standard logistic random variable, r_i is a dummy variable indicating whether individual i is in a stable relationship, w_i represents i 's total wealth including income, saving, real estate, etc., a_i represents i 's age in numerical numbers, pm_i illustrates the parents marital status (divorced or not) for individual i , h_i records whether i is in a stable health status without suffering from severe disease within two years, and eth_i is ethnicity.

Data Generating Process.

We could generate data of y_i as follows

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0 \text{ (get married)} \\ 0, & \text{otherwise (not get married)} \end{cases} \quad (4)$$

where y_i^* is a latent variable specified as

$$y_i^* = \beta_0 + \beta_1 r_i + \beta_2 w_i + \beta_3 a_i + \beta_4 pm_i + \beta_5 h_i + \beta_6 eth_i + \varepsilon_i, \quad (5)$$

and ε_i follows the standard logistic distribution.

Key Factors.

Among these six chosen variables, I think the first three variables, which are whether individuals are in a stable relationship, individual's wealth condition, and age, have most important impact on someone's decision of marriage. Firstly, being in a relationship is a prerequisite for considering marriage. However, most responsible people would not consider marriage until they have established a healthy and stable relationship with their partner. Thus, whether or not being in a stable relationship would be a key factor when someone considers whether to get married. Besides, people would commonly take their wealth level into account before marriage. Marriage actually means that two people form a new family, which involves setting up a new home and presumably raising the next generation in the near future. All of these require a solid source of wealth. In addition, someone would like to get married because they are at marriageable age. Within a certain age range, people are under more pressure to get married when they become older.

Preliminary Test.

A preliminary test could be conducted by using data collected from a relatively small scale survey designed for the desired variables. After collecting data, we could estimate the logit model (3) by maximum likelihood estimation and implement statistical tests such as the likelihood ratio test to see the joint significance of chosen factors. Moreover, we could measure the goodness-of-fit of the model by pseudo R^2 ,

$$\tilde{R}^2 = 1 - \frac{L_{ur}}{L_0},$$

where L_{ur} is the log-likelihood for the unrestricted model and L_0 is the log-likelihood for the model estimated with only an intercept term.