

MACS30100 Homework1

Takuyo Ozaki

1/19/2020

Building models

Deviant aggressive behavior

We assume rational individuals under incomplete information. They maximize their expected utility (= expected benefit minus expected cost) for an individual action Y :

$$E[U_i(Y)] = \sum E[B_i(Y)] - \sum E[C_i(Y)]$$

Under this assumption, individuals will choose the deviant action when “expected utility from deviant behavior” exceeds “expected utility from normal behavior”:

$$E[U_i(Y_{deviant})] - E[U_i(Y_{normal})] > 0$$

In other words, individuals will choose the deviant action when “expected benefit minus cost from deviant behavior” exceeds “expected benefit minus cost from normal behavior”:

$$(\sum E[B_i(Y_{deviant})] - \sum E[C_i(Y_{deviant})]) - (\sum E[B_i(Y_{normal})] - \sum E[C_i(Y_{normal})]) > 0$$

To reduce deviant aggressive behavior in this model, we need to increase $E[B_i(Y_{normal})]$ or $E[C_i(Y_{deviant})]$, or decrease $E[B_i(Y_{deviant})]$ or $E[C_i(Y_{normal})]$, which lead to decrease $E[U_i(Y_{deviant})] - E[U_i(Y_{normal})]$.

We can apply this model to theory 1 to 4 and propose the appropriate social policies to reduce deviant aggressive behavior.

- Theory1: Education policy is an appropriate example. The government should provide the scholarship with “good” students who contribute to the society (e.g., volunteer) while “bad” students who did deviant aggressive behavior (e.g. drug, violence) should be expelled from the school. By doing this, we can increase $E[B_i(Y_{normal})]$ and $E[C_i(Y_{deviant})]$, which could suppress the deviant actions.
- Theory2: The welfare program is an appropriate example. The government should supply income support or earned income tax credit with low-income people, some of who are doing deviant behavior by being frustrated with their life. By reducing the inequality between deviant people and personal authorities, deviant people do not need to show the hostility toward personal authorities. So, we can decrease $E[B_i(Y_{deviant})]$, which could decrease the deviant actions.

- Theory3: Labor policy is an appropriate example. The government should ban job discrimination based on gender, race, age, and criminal record. Moreover, the government should support deviant people (e.g., criminals) to get jobs. By eliminating the discrimination and increasing the job opportunities, deviant people are likely to profit from conforming to the social rules. Thus, we can increase $E[B_i(Y_{normal})]$, which could reduce the deviant actions.
- Theory4: Housing policy is an appropriate example. The government should give opportunities for deviant people to live in the public housing of many different areas. By isolating from other deviant subcultures, they increase the cost of contacting with other deviant people. Thus, we can increase $E[C_i(Y_{deviant})]$, which could suppress the deviant actions.

Waiting until the last minute

- a. The reason why people often do things at the last minute can come from the following human natures.
 - First, people tend to overestimate present things ("present bias") and underestimate future things. As a result, people are likely to begin to do things that they face at present, while they are likely to procrastinate tasks where the deadline is not coming up soon.
 - Second, people tend to complicate things when there is plenty of time to do. On the other hand, if people have less available time, people need to focus on only the necessary things to accomplish a task. As a result, people are likely to begin to do things when the deadline is approaching.
- b. Generally speaking, we tend to have "present bias", i.e. get more utility at present than in the future. Suppose there are only two periods, present t , and future $t + 1$. In the simplest model, the utility is described by:

$$U_t = \mu(c_t) + \delta\mu(c_{t+1})$$

where U_t is the sum of utility at the time of t , $\mu(c_t)$ is the utility from the action c at the time of t , δ is the discount rate ($\delta < 1$).

This model explains the observation (waiting until the last minute) because we get more utility from future tasks when time goes. For example, at the time of t , the utility from doing the task to accomplish a task at the time of $t+1$ is small ($U_t = \delta\mu(c_{t+1})$), but at the time of $t+1$, the utility from it can become larger ($U_{t+1} = \mu(c_{t+1})$).

- c. Generally speaking, tasks are likely to complicate to fill the available time for its completion (the Parkinson's law). In other words, if you have more available time to accomplish a task, you would think the task more complicated and would need to spend more time. If you have less available time, you would need to focus only on the necessary things to accomplish a task. In the simplest model to explain Parkinson's law, we can induce the following equations:

$$E[C(Y)] = T(Y) * D(Y) * Q(Y)$$

where $E[C(Y)]$ is the expected cost to achieve a task Y , $T(Y)$ is the available time to achieve Y , $D(Y)$ is the difficulty of Y , and $Q(Y)$ is the quantity of Y . When the benefit of achieving the task ($E[B(Y)]$) exceeds the cost of achieving the task ($E[C(Y)]$), we begin to do Y . This model explains the observation (waiting until the last minute) because when the deadline is approaching (i.e., $T(Y)$ get smaller), we can start the task since the cost of achieving a task ($E[C(Y)]$) can get smaller than the benefit ($E[B(Y)]$).

- d. The followings are the two predictions derived from the model (b) and (c).

- (b1) The model(b) predicts that most people would not save money until the last minute of the retirement, but people who do not overestimate the present things would save money. The followings are the assumptions.
 - There are only two periods; present t , and future $t + 1$.
 - People earn money m at present, and retire(= earn no money) in the future.
 - The utility function is $U = \sqrt{m}$, where m is consumption amount of money.
 - The sum of utility is $U_t = \mu(c_t) + \delta\mu(c_{t+1})$, where $\delta(< 1)$ is the discount rate.

In this case, the sum of utility is maximized when people consume m and save no money at present because the utility from money at present is more weighted than the one in the future (i.e., $\delta < 1$). However, when people do not overestimate the present thing (i.e., $\delta = 1$), it is maximized when people consume $1/2m$ at the time of t , and $1/2m$ at the time of $t + 1$ because the present utility from money at present is the same weight as the one in the future. Thus, this model explains that most people do not save until their retirement.

- (b2) The model(b) predicts that non-health-conscious people usually do not go on a diet while health-conscious people do it. The followings are the assumptions.
 - There are only two periods; present t , and future $t + 1$.
 - People can choose: (1) go on a diet or (2) do not go on a diet.
 - When people go on a diet at t , people can be healthy at $t + 1$.
 - When people do not go on a diet at t , people can cause disease at $t + 1$.
 - The utility from a diet is $\mu(diet) = -d$ if go on a diet, $\mu(nondiet) = d$ if not.

- The utility from health condition is $\mu(health) = h$ if health, and $\mu(disease) = -h$ if not
- The sum of utility is $U_t = \mu(c_t) + \delta\mu(c_{t+1})$, where $\delta(< 1)$ is the discount rate.

In this case, the sum of utility is $U_t = -d + \delta * h$ if go on a diet, and $U_t = d - \delta * h$ if not. Thus, non-health-conscious people ($d > h$) do not go on a diet while health-conscious people ($d < h$) go on a diet. Thus, this model explains that non-health-conscious people do not go on a diet.

- (c1) The model(c) predicts that workers would begin to do a difficult task when the deadline is approaching while they would begin to do an easy task soon. The assumptions and notations are the same as my answer (c). $E[C(Y)] = T(Y) * D(Y) * Q(Y)$. And people begin to do the task when $E[B(Y)] - E[C(Y)] > 0$. In other words, people begin to do the task when the following inequations are satisfied:

$$E[B(Y)] > E[C(Y)] = T(Y) * D(Y) * Q(Y)$$

Thus, the model predicts when a task is easy (i.e., $D(Y)$ is small), people tend to begin to do the task soon because the benefit $E[B(Y)]$ can exceed the cost $E[C(Y)]$, even if there is enough available time (i.e., $T(Y)$ is big).

- (c2) The model(c) predicts that when the assignment has a lot of questions, students are less likely to begin to solve them as soon as the professor publishes the assignment. The assumptions are the same as my answer (c). And Y is the assignment, not task. In this case, the model(c) predicts that if the quantity of the assignment (i.e. $Q(Y)$) is so large, students begin to do the assignment when the deadline is coming up soon (i.e., $T(Y)$ get smaller) and the benefit $E[B(Y)]$ can exceed the cost $E[C(Y)]$.

Selecting and fitting a model

Question 1

- When the sample size n is extremely large, and the number of predictors p is small, the performance of a flexible learning method is expected to be better than an inflexible method because sample size is so large that the model fit more parameters and the smaller number of predictors reduce the variance of the models.
- When the number of predictors p is extremely large and the number of observations n is small, the performance of a flexible learning method is expected to be worse than an inflexible method because there is more possibility of overfitting.
- When the relationship between the predictors and response is highly non-linear, the performance of a flexible learning method is expected to be better than an inflexible method because it might be less restrictive on the shape of fit.

- d. When the variance of the error terms $\sigma^2 = Var(\epsilon)$ is extremely high, the performance of a flexible learning method is expected to be worse than an inflexible method because there is more possibility of overfitting.

Question 2

- Bias can decrease with more model flexibility because it is easier to fit the model since it has fewer assumptions made about the shape of the fit
- Variance can increase with more model flexibility because while new data can affect more on the parameter estimates of the flexible model while it does not affect so much on one of the inflexible models like a linear regression model
- Training error can decrease with higher flexibility because the flexible model can fit more, which can reduce MSE on the training data, even though it is overfitting
- Test error has a U-shaped curve because the expected test MSE for a given value x_0 is the sum of three fundamental quantities: the variance of $f(x_0)$, the squared bias of $f(x_0)$ and the variance of the error variance terms
- Irreducible error is always the same value, which is not related to the model fit