

## Problem Set 1

MACS 30010, Dr. Evans

Fiona Fan

**Problem 1** Classify a model from a journal (5 points).

**Part (a, b & c).**

Chetty (2015) lays out a framework of how behavioral economics can positively influence public policy modeling. To conceptualize, Chetty introduced a representative-agent model, as shown in (1). Here, the agent tries to maximize her experienced utility  $u(c)$  (the actual well-being) by choosing a set of tax rates  $t$  and "nudges"  $n$  as described in Thaler and Sunstein (2008). Policy-makers architect nudges to direct the public towards a pre-assumed better outcome, like placing the healthy food in more salient places in a cafeteria while hiding away the junk food.  $d$  represents the set of ancillary conditions that resemble nudges, which influence people's behaviors but cannot be manipulated by policy-makers, unlike  $n$ .  $c$  represents a set of different consumption goods or consumption at different times.  $\bar{R}$  is a revenue requirement.  $v(c)$  is her decision utility (the objective to maximize).  $p$  is the pretax prize vector for consumption good  $c$ , and  $Z$  is the individual's wealth.

$$\max_{t,n} u(c) \text{ s.t.} \quad (1)$$

$$t \cdot c = \bar{R} \quad (2)$$

$$c = \arg \max_c \{v(c|n, d) \text{ s.t. } (p + t) \cdot c = Z\}. \quad (3)$$

**Part (d).**

Here in the model,  $u(c)$ ,  $v(c|n, d)$ ,  $c$ ,  $t$ ,  $n$  are endogenous, determined internally by the model in a two-step process. First  $c$  is determined to maximize the agent's decision utility  $v(c)$ , and then  $t, n$  are determined to maximize the agent's experienced utility  $u(c)$ .  $d, p, \bar{R}, Z$  are exogenously given for the maximization.

**Part (e).**

This model is static as no time-dependent changes are introduced. It is highly non-linear as the endogenous variables are interdependent. It is deterministic as no randomness is introduced.

**Part (f).**

Endogenously chosen leisure can be added to the model. I will give one example of how below. Suppose tax  $t$  is consumption tax only, then the model becomes:

$$\max_{t,n} u(c) \text{ s.t.} \quad (4)$$

$$t \cdot c = \bar{R} \quad (5)$$

$$c, l = \arg \max_{c,l} \{v(c, l|n, d) \text{ s.t. } (p + t) \cdot c + (1 - l) \cdot w = Z\}. \quad (6)$$

,where  $w$  is the exogenous wage and  $l$  is the endogeneous leisure. The total amount of time is assumed to be 1. Here, experienced utility that the policy maker

uses to make decision is still determined by  $c$ . However, when choosing  $c$  to maximize decision utility, the agents take  $l$ , leisure, into consideration as well.

### Bibliography:

Chetty, Raj. "Behavioral economics and public policy: A pragmatic perspective." The American Economic Review 105, no. 5 (2015): 1-33.

**Problem 2** Make your own model. (5 points).

#### Part (a, b & c).

Since here the outcome variable is binary, I will use logistic regression to model the decision to get married.

$$Pr(Y_i = 1 | X_{1i}, X_{2i}, \dots, X_{Ki}) = F(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \dots + \beta_k X_{ki} + \epsilon_i) \quad (7)$$

where

$$F(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Here,  $Y$  is the outcome variable whose values 1 (get married) and 0 (not married).  $X_1$  is age, a continuous numerical variable;  $X_2$  is ethnicity, a categorical variable;  $X_3$  is zipcode, treated as a categorical variable;  $X_4$  is the number of years of higher education, a continuous variable;  $X_5$  is the number of years of parental higher education, also a continuous variable;  $X_6$  is whether the parents divorced, a binary variable;  $X_7$  is the subject's current dating status, a categorical variable;  $X_8$  is the subject's religion, a categorical variable;  $X_9$  is the income level of the subject, a continuous variable.  $\epsilon$ , drawn from is the error term, representing haphazard events that can prompt a person to make the decision, like the legalization of same-sex marriage.  $\epsilon$  is drawn from a log-normal distribution  $LN(0, \sigma^2)$ .

#### Part (d).

The variables that could bear most significance, in my opinion, are age, income level of the subject, current dating status, and religion. The rest are supplementary control variables.

#### Part (e).

I chose the variables described in Part (a-c) based on my intuition and some literature review. The factors described in Part (d), in my opinion, have more salient causal influence over people's decision to get married, as compared to the rest, for the following reasons:

1. age
  - Older people are more likely to be urged into marriage.
2. income level of the subject
  - More financially well-off people might be less likely to seek the stability of a marriage.
3. current data status

- People in a stable relationship are more likely to decide to marry.

#### 4. religion

- Certain religions prompt believers to get married within a certain age range.

#### **Part (f).**

A test on the significance of these factors could be a regression run on the available census data, or, (since the dating status of the participant might be unavailable) a small scale survey. A preliminary logistic regression (9) will not only tell us about the significance of the factors, but will also help us calibrate our model.

$$Pr(Y'_i = 1 | X'_{1i}, X'_{2i}, \dots, X'_{Ki}) = F(\beta'_0 + \beta'_1 X'_{1i} + \beta'_2 X'_{2i} + \beta'_3 X'_{3i} \dots + \beta'_k X'_{ki} + \epsilon'_i) \quad (9)$$

Additionally, if we want to see if our model holds up against real-world data, we can simulate the data, using randomly generated  $X_{1i}, X_{2i}, \dots, X_{Ki}$  and  $\epsilon$  with similar mean and variance to the survey data, to predict  $Y_i$ . Then we compare the simulated  $Y_i$  with  $Y'_i$  and see if the two groups are significantly different. If not, then we can say model is somewhat robust.