University of Minnesota
ECON 8186
Homework 1
Angelo Mendes and Johanna Torres Chain

Prof. Fatih Guvenen
January 31, 2023

BENCHMARK CONSUMPTION UTILITY FUNCTIONS

We start from our *income-fluctuations problem*:

$$V(a_t; y_t) = \max_{c_t} \left(u(c_t) + \delta \mathbb{E}[V(a_{t+1}; y_{t+1}) \mid y_t] \right)$$

$$\text{s.t.} \quad \begin{aligned} & c_t + a_{t+1} = y_t + (1+r)a_t \\ & \log y_t = \rho \log y_{t-1} + \eta_t \\ & a_t \geq \underline{a} \end{aligned}$$

(a) We start by the variational approach perturbing consumption level c_t by $\epsilon > 0$:

$$\underbrace{-u'(c_t)\epsilon}_{\text{utility lost by consuming } c_t - \epsilon} \quad + \underbrace{\delta \mathbb{E}[(1+r)u'(c_{t+1})\epsilon + \epsilon\theta u'(c_{t+1})]}_{\text{utility gain at } t+1} = 0$$

Note that the utility gain at $t+1$ by reducing consumption at t can be splited in two terms. First, the usual marginal utility by consuming $(1+r)\epsilon$ at $t+1$. Second, the term in blue expresses the utility gain tomorrow by reducing the *habit* term given by a share of the consumption at t. We thus have:

$$u'(c_t) = \delta \mathbb{E}[(1+r+\theta)u'(c_{t+1})]$$

$$1 = \delta(1+r+\theta)\mathbb{E}\left[\left(\frac{c_{t+1} - \theta c_t}{c_t - \theta c_{t-1}}\right)^{-\alpha}\right]$$

We now derive the same equation by the FOC and the Envelope condition.

$$V(a_t; y_t) = \max_{a_{t+1}} \{u(y_t + (1+r)a_t - a_{t+1}) + \delta \mathbb{E}[V(a_{t+1}; y_{t+1}) \mid y_t]\}$$

Using the functional form for the utility function and the budget constraint:

$$V(a_t; y_t) = \max_{a_{t+1}} \left\{ \left(\frac{(y_t + (1+r)a_t - a_{t+1}) - \theta(y_{t-1} + (1+r)a_{t-1} - a_t)}{1-\alpha} \right)^{1-\alpha} + \delta \mathbb{E}[V(a_{t+1}; y_{t+1})] \right\}$$

Computing the marginal utility we have the following expression:

$$u'(c_t) = \left(\underbrace{(y_t + (1+r)a_t - a_{t+1})}_{c_t} - \theta \underbrace{(y_{t-1} + (1+r)a_{t-1} - a_t)}_{c_{t-1}} \right)^{-\alpha}$$

FOC:

$$-u'(c_t) + \delta\mathbb{E}[V_a(a_{t+1};y_{t+1})] = 0$$

Envelope:

$$V_a(a_t;y_t) = (1+r+\theta)\times\Big((y_t + (1+r)a_t - a_{t+1}) - \theta(y_{t-1} + (1+r)a_{t-1} - a_t)\Big)^{-\alpha}$$

$$V_a(a_{t+1};y_{t+1}) = (1+r+\theta)\times\Big(\underbrace{(y_{t+1} + (1+r)a_{t+1} - a_{t+2})}_{c_{t+1}} - \theta\underbrace{(y_t + (1+r)a_t - a_{t+1})}_{c_t}\Big)^{-\alpha}$$

We then have that:

$$1 = \delta(1+r+\theta)\mathbb{E}\left[\left(\frac{c_{t+1} - \theta c_t}{c_t - \theta c_{t-1}}\right)^{-\alpha}\right]$$

If we don't have to deal with endogenous habit formation the Euler Equation takes a more usual formulation:
FOC:

$$u'(c_t) = \delta\mathbb{E}[V(a_{t+1};y_{t+1})]$$

Envelope:

$$V_a(a_t;y_t) = (1+r)u'(c_t) \Longrightarrow V_a(a_{t+1};y_{t+1}) = (1+r)u'(c_{t+1})$$

Therefore:

$$u'(c_t) = (1+r)\delta\mathbb{E}[u'(c_{t+1})|y_t]$$

$$(b) \quad 1 = \delta(1+r)\mathbb{E}\left[\left(\frac{c_{t+1} - \theta\overline{C}_t}{c_t - \theta\overline{C}_{t-1}}\right)^{-\alpha}\middle| y_t\right]$$

$$(c) \quad 1 = \delta(1+r)\mathbb{E}\left[\left(\frac{c_{t+1} - \theta\overline{C}_{t+1}}{c_t - \theta\overline{C}_t}\right)^{-\alpha}\middle| y_t\right]$$

By the variational argument:

$$-\epsilon u'(c_t) + \delta(1+r)\mathbb{E}[u'(c_{t+1})\epsilon|y_t] = 0$$

Substituting the marginal utilities we get the same results as using FOC and Envelope.

DATA WORK - US Current Population Survey data

In this section, we present some data from IPUMS and run a Mincer regression. We downloaded the following variables:

1. *age*
2. *sex*
3. *educ99*: educational attainment. From this variable we construct 4 dummies: *(i)* high school degree; *(ii)* some college; *(iii)* college degree; and *(iv)* graduate education, or at least some post-baccalaureate education.
4. *uhrsworkt*: hours usually worked per week at all jobs
5. *uhrsworkly*: hours usually worked per week last year
6. *wkswork1*: weeks worked last year
7. *incwage*: wage and salary income

After that, we created a sample following the instructions described in the problem set:

- Keep if age is between 25 and 60, inclusive.
- Drop individuals with missing data on incwage, usual weekly hours, weeks worked, age, education, and gender. Keep if state is missing.
- Drop if INCWAGE is less than $2,000 (2012 dollars).
- Drop if hourly wage > $500, annual hours < 50 hrs.

(a) Tabulate summary statistics for the raw data you downloaded and for the final sample you created with the selection criteria just described: report the sample size; and the mean, median, standard deviation, min, max, skewness, kurtosis, and interquartile range for all the variables you downloaded. (Prepare two tables, one for each sample, with the same format.)

Table 1: Summary - Raw data

Variable	Obs	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	IQ Range
gender1	201,398	0.49	0.50	0	1	0.06	1.00	1
age	201,398	35.36	22.27	0	85	0.24	2.07	17
educ1	201,398	0.21	0.41	0	1	1.40	2.95	0
educ2	201,398	0.20	0.40	0	1	1.46	3.15	0
educ3	201,398	0.13	0.34	0	1	2.16	5.68	0
educ4	201,398	0.07	0.26	0	1	3.33	12.07	0
wkswork1	201,398	22.82	24.71	0	52	0.24	1.13	52
uhrsworkly	201,398	524.89	480.33	1	999	-0.03	1.00	0
incwage	201,398	2.32	4.22	0.00	10.00	1.27	2.61	10.30
hwage	99,402	22.34	40.23	0	3000	20.28	1077.00	9.94
anhrswork	201,398	900.60	1061.69	0	5148	3.63	14.90	2080

We now drop the variables suggested in the enunciate of the problem.

Table 2: Summary - Treated data

Variable	Obs	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	IQ Range
gender1	71,548	0.52	0.50	0	1	-0.07	1.01	1
age	71,548	42.00	9.87	25	60	0.03	1.89	16
educ1	71,548	0.27	0.44	0	1	1.05	2.10	1
educ2	71,548	0.28	0.45	0	1	0.97	1.94	1
educ3	71,548	0.24	0.43	0	1	1.24	2.53	0
educ4	71,548	0.13	0.34	0	1	2.18	5.77	0
wkswork1	71,548	48.62	8.99	1	52	-2.93	10.98	0
uhrsworkly	71,548	40.70	10.20	1	99	0.41	7.43	0
incwage	71,548	49884.48	57179.62	2000	1168999	7.80	107.71	38000
hwage	71,548	24.53	25.09	0.39	500	7.50	97.52	16.61
anhrswork	71,548	1994.43	635.29	50	5148	-0.25	5.14	260

(b) Mincer wage regression

Define $y_i := \log(Y_i)$ as the log income (incwage). Then, we run the following version of the Mincer wage regression:

$$y_i = d_{\text{age}} + d_{\text{educ}} + d_{\text{gender}} + d_{\text{state}}$$

where d_j indicates a dummy for variable j, and the regression is run in a single cross-section (2012). Dummy for age 25 is omitted for normalization. Education is the last degree completed and is a categorical variable with 4 possible values: (i) high school degree; (ii) some college; (iii) college degree; and (iv) graduate education, or at least some post-baccalaureate education. (The fifth dummy, for less-than-highschool education, is omitted for normalization.)

The seniority gap (between age 25 and 55)
By our definition of the dummy variables (since we excluded the dummy for 25-year-old observations) the seniority gap is given by the coefficient of $d_{\text{age}=55} = 0.60273083$. This implies that holding everything else constant, income is predicted to be 82.7% greater for an individual who is 55 than for an individual who is 25.

The gender gap
In our regression, the dummy for men is excluded for normalization, then the gender gap is just the negative of the coefficient of $-d_{\text{female}} = 0.41574597$. This implies, holding everything else constant that income is predicted to be 41.6% greater for a men than for women.

The college premium
The college premium is the difference between the coefficients for the dummies $d_{\text{college}} - d_{\text{highschool}}$. Than in the Stata set-up correspond to $educ_4$ and $educ_2$. The college premium is estimated then to be 0.58384675, implying that income is predicted to 79.3% greater

for a college graduate than for a high school graduate.

States with lowest and highest average log wage and the wage level
Based on the state dummies the state with the lowest income is Idaho and the state with the highest is Washington D.C. To calculate the average income level in each state we run a separate regression, $y_i = \gamma_0 + \gamma_1 d_{\text{state}}$.
The average log income level then is $d_{\text{Washington D.C.}} + \alpha_0 = 10.819499$ and $d_{\text{Idaho}} + \alpha_0 = 10.253148$. This implies income levels of $\$49,985$ and $\$28,370$, respectively.

Table 3 summarizes all the results.

Table 3: Results for the Mincer wage regression

Variable	Value
Seniority gap	0.60273083
Gender gap	0.41574597
College premium	0.58384675
Lowest wage (Idaho)	\$28,370
Highest wage (District of Columbia)	\$49,985

(c) Expanded mincer wage regression

Again, define $y_i := \log(Y_i)$. Then, we run the following version of the Mincer wage regression.

$$y_i = \alpha_0 d_{\text{age}} + \alpha_1 d_{\text{educ}} + \alpha_2 d_{\text{gender}} + \alpha_3 d_{\text{age}} \times d_{\text{educ}} + \alpha_4 d_{\text{gender}} \times d_{\text{educ}} + \alpha_5 d_{\text{gender}} \times d_{\text{age}} + \alpha_6 d_{\text{state}} + \epsilon$$

This version of the Mincer includes the interaction of age, gender, and education dummies. By doing this we allow for heterogeneity in the effects of schooling between gender and across ages.

The seniority gap (between age 25 and 55) for men and women by education level
Five different education levels are defined so we estimate 10 seniority gaps (5 per woman and 5 per man).

For men:

- Less-than-high school education
$$\text{Seniority Gap} = d_{\text{age=55}} = 0.54360885$$
- High school degree
$$\text{Seniority Gap} = d_{\text{age=55}} + d_{\text{age=55}} \times d_{\text{highschool}} = 0.61655359$$
- Some college
$$\text{Seniority Gap} = d_{\text{age=55}} + d_{\text{age=55}} \times d_{\text{somecollege}} = 0.6364491$$

- College degree
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{college}} = 0.72868877$$
- Graduate Education
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{gradschool}} = 1.0060805$$

For women:

- Less-than-high school education
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{female}} = 0.38733227$$
- High school degree
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{female}} + d_{\text{age}=55} \times d_{\text{highschool}} = 0.46027702$$
- Some college
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{female}} + d_{\text{age}=55} \times d_{\text{somecollege}} = 0.48017252$$
- College degree
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{female}} + d_{\text{age}=55} \times d_{\text{college}} = 0.57241219$$
- Graduate Education
$$\text{Seniority Gap} = d_{\text{age}=55} + d_{\text{age}=55} \times d_{\text{female}} + d_{\text{age}=55} \times d_{\text{gradschool}} = 0.84980397$$

Table 4 summarizes the results for the seniority gap by gender and education.

Table 4: Seniority gap

Education	Male	Female
Less-than-high school education	0.54360885	0.38733227
High school degree	0.61655359	0.46027702
Some college	0.6364491	0.48017252
College degree	0.72868877	0.57241219
Graduate Education	1.0060805	0.84980397

The average gender gap and the gender gap for college graduates at ages 30, 40, and 50.
The average Gender Gap is defined as:

$$\mathbb{E}(ln(w)|g = \text{male}) - \mathbb{E}(ln(w)|g = \text{female})$$

In order to calculate these expectations, we run the following regression:

$$y_i = \beta_0 + d_{\text{gender}} = d_{\text{state}}$$

As in the expanded mincer, we assumed that the gender gap is independent of the state. Then the average gender gap is just the coefficient: $-d_{\text{female}} = 0.35949086$ (negative since d_{male} is normalized).
Finally, the gender gap for college graduates at ages 30, 40, and 50.

- College graduates at the age of 30
  $$\text{Gender Gap} = -(d_{\text{female}} + d_{\text{age=30}} \times d_{\text{female}} + d_{\text{college}} \times d_{\text{female}}) = 0.31485055$$
- College graduates at the age of 40
  $$\text{Gender Gap} = -(d_{\text{female}} + d_{\text{age=40}} \times d_{\text{female}} + d_{\text{college}} \times d_{\text{female}}) = 0.403228$$
- College graduates at the age of 50
  $$\text{Gender Gap} = -(d_{\text{female}} + d_{\text{age=50}} \times d_{\text{female}} + d_{\text{college}} \times d_{\text{female}}) = 0.49470976$$

Table 5 summarizes the results of the gender gap for college graduates of different ages.

Table 5: Gender gap for college graduates (*educ3=1*)

Age	Value
30	0.31485055
40	0.403228
50	0.49470976
Avg	0.35949086

The college premium, separately for each gender, at age 45.

- Men at the age of 45
  $$\text{College Premium} = d_{\text{college}} - d_{\text{highschool}} + d_{\text{age=45}} \times d_{\text{college}} - d_{\text{age=45}} \times d_{\text{highschool}} = 0.57401622$$
- Women at the age of 45
  $$\text{College Premium} = d_{\text{college}} - d_{\text{highschool}} + d_{\text{age=45}} \times d_{\text{college}} - d_{\text{age=45}} \times d_{\text{highschool}} + d_{\text{female}} \times d_{\text{college}} - d_{\text{female}} \times d_{\text{highschool}} = 0.57982136$$

Table 6 summarizes the results for the college premium by gender at the age of 45.

Table 6: College premium

Gender	Value
Men	0.57982136
Women	0.57401622

The states with lowest and highest average log wage and the wage level
Again, as in part B, Idaho has the lowest average wage and Washington D.C. has the highest. Using the same methodology as in part, the average log income level in D.C. is estimated to be $49,985, and the average log income level in Idaho is estimated to be $28,370.

(d) Discuss how your results in parts b and c compare

First, let's compare the senior gap among the different mincer regressions. In the exercise in B, we assumed the gap to be constant across genders and education levels, therefore we find a gap of 0,6027 (all the comparatives are in units of logged wage), this implies that much of the lifetime return comes after age 25. In part C, we allowed the gap to vary across gender and education. We found then that the gap is increasing in the education level for both, men and women, and that is larger for men for all education levels. In general, lifetime earnings tend to be increasing in education and tend to be larger for men.

Then, in B we also calculated the gender gap unconditionally on education level or age. In C, we allowed this gap to vary across ages and education. Unconditionally, we got a gender gap of 0.4157, but conditioning in college graduation, we found that is smaller at the age of 30 (0.3148) but increasing in age. Therefore the gender gap at the age of 50 for college graduates is larger (0.4947) than the unconditional one in B.

Next, we calculated the college premium. As in the previous exercises in B, the unconditional college premium is 0.5838. When we allowed this premium to vary across gender and age, we found that only conditioning in sex, this premium is similar at the age of 45 (0.57 for both, but slightly higher for men). This implies that at age of 45, the return to college is similar for men and women, and is close to the unconditional college premium.

Finally, in both exercises (B and C) we calculate the states with the lowest and highest average log wage and wage level. We found that regardless that the state premium changes slightly in B to C (coefficients), Idaho and Washington D.C. have the lowest and highest coefficient respectively.