Topics in Labor and Demographic Economics: Midterm Problem Set

INSTRUCTIONS (READ CAREFULLY)

The problem set will be split into two sections to make sure that you are making adequate progress on it before it is due, since the first part involves being able to properly clean up the data and prepare it for data analysis.

First component (submission) [due May 14th, 2025 at 11:59PM]:
Submit via e-mail to me a report in the form of a pdf (can also be an R markdown) showing the progress you have made for Question 1 [dealing with the data cleaning and analysis of the problem set]. This involves submitting an initial version of the figures and tables to be submitted for Question 1.

Second component (final submission:)
Submit all of the files related to your solution to this problem set via email to me (andrea.flores@fgv.br) by **WEDNESDAY, JUNE 1ST at 23:59h**. Your solution to this problem set should consist of **(1)** a pdf with your responses to each item as if it was a report (for items just requiring coding with no discussion, include a reference to the relevant part of your code); and **(2)** code files.

Go as far as you can. Even if you don't manage to complete all items in a question, partial credit will be granted generously if you clearly describe the issues you faced in implementing the estimation method and intuitively explain how you think these issues could be addressed.

Good luck!

Preliminaries: Data Download and Setup

We will be combining two different data sources for this problem set. The objective is for you to get acquainted with the process of setting up, cleaning, and analyzing data that can be used to estimate labor supply models considering potential differences in household behavior by household structure (headed by a married couple, by a cohabiting couple or by a single person). The data to be downloaded include the following:

1. ENOE (National Survey of Occupation and Employment)
2. ENIGH (National Survey of Household Income and Expenditure)

Question 1: Data Preparation and Analysis

(a) Use the ENOE data set that you have downloaded and plot the yearly mean wages, mean hours worked (unconditional and conditional) and mean employment rates of women aged 25-65 in the downloaded sample. Interpret the yearly trends presented in the plots.

(b) Generate **three** tables, one for married women aged 25-65, another one for cohabiting women aged 25-65, and another one for single women aged 25-65. In each table – formatted as below – report the average labor force participation rates and weekly earnings by age and education. Interpret the relevant empirical patterns.

	Labor Force Participation				**Earnings**			
	All	*Less than HS*	*High School*	*College or Above*	*All*	*Less than HS*	*High School*	*College or Above*
All								
Less than 35 with Child Younger than 16								
Less than 35 with No Small Children								
35-54								
55 and Older								

(c) Focusing on partnered women aged 25-65, compute the ratio of women's housework hours to their spouses' housework hours (that is, women's weekly hours spent in housework divided by the weekly hours spent by their spouses in housework). Generate a graph plotting how this ratio has evolved over time for married and cohabiting women (separately)

(d) Using the ENIGH data, aggregate the different expenditures variables in the `concentradohogar` table to generate the following expenditure categories

- Food
- Transportation
- Health Services
- Housekeeping Services
- Health Insurance
- Home Insurance
- Utilities
- Child Care
- Education
- Rent

In two different tables – one for households headed by a married couple and another one for households headed by a cohabiting couples – report the mean and median values of these households' expenditures on the different categories listed above

Question 2: Semi-Parametric, Structural Estimation

Suppose that a partnered woman's utility over leisure and consumption bundles follows the functional form:

$$U(C, L; \mathbf{x}, \epsilon) = \beta \log(L) + (1 - \beta) \log(C)$$

where $\beta \in (0, 1)$ and $\beta = \mathbf{x}'\boldsymbol{\theta} + \epsilon$. The wage equation is

$$w(z, \xi) = z'\gamma + \xi$$

As covered in class, the observed wage equation can be written as

$$w(z, \xi) = z'\gamma + M(\Pr(P = 1|y, z, x)) + u$$

Using (1) the sub-sample of married women between the ages of 25 and 65 and (2) the sub-sample of cohabiting women between the ages of 25-65, implement the following estimation steps (separately):

(a) Non-parametrically estimate $\Pr(P = 1|y, z, x)$ using a kernel regression where z includes completed education and age and x includes a constant, age and current number of children

(b) Considering the observed wage equation described above, use Robinson's partial regression model to estimate γ. See pg. 157 in the MicroeconometricsMatlab.pdf file located in the Readings subfolder of the shared Dropbox folder. With your results, predict the wages for the sub-sample of women who don't work.

(c) We want to also account for the intensive margin in our estimation of β. With the predicted wages computed in part (b) estimate β via maximum likelihood considering that we observe $h_i > 0$ for women working (upon setting up the likelihood function described in slide 42 of the StaticIntensive-Margin_Lecture slides).

(d) Do you fail to reject the theoretical implications of the model for any of the sub-samples? Discuss.

Question 3: Marshallian Labor Supply

Take the model we considered in the Estimation Lecture of the Static Intensive Labor Supply:
Let the direct utility follow a Stone-Geary form

$$U = B_0 \ln(L - \gamma_L) + B_1 \ln(C - \gamma_C)$$

where

$$B_0 + B_1 = 1$$
$$C_i - \gamma_C > 0; \quad L - \gamma_L > 0$$
$$B_0 = x'\tilde{B}_0 + \epsilon$$

(a) Suppose that the price of consumption is 1. What are the Marshallian demand functions for consumption and leisure?

(b) Under what conditions does the following condition hold?

$$\mathbb{E}[x'(\gamma_L w + x'\tilde{B}_0(wT + y - \gamma_C - \gamma_L w) - wL)] = 0$$

Hint: We already know this would require an interior solution, but what else should we be willing to assume regarding tastes for this moment condition to hold?

(c) Derive a similar moment using the Marshallian demand for the consumption of C. Combine this moment with the moment in part (b) to estimate the parameters of the model using GMM with data from the ENIGH data set used in Question 1 (aggregating all expenditure categories listed in item (d) to get a measure of total household consumption/expenditures) but restrict your sample to two main sub-samples:

- Sub-sample 1: Keep only the observations of married men who are between 25 and 65 years old at the time of the survey
- Sub-sample : Keep only the observations of cohabiting men who are between 25 and 65 years old at the time of the survey

(d) Compute S_{00} for the two sub-samples separately. Do you fail to reject the theoretical implications of the model? Discuss.

Question 4: Intertemporal Model of Labor Supply

Consider a simple 3-period model of consumption and labor supply with $T = 3$ denoting the retirement stage. That is, $L_3 = L_0$ where L_0 denotes total time endowment. Preferences are temporally separable so that at age 1, the agent solves the following life-cycle maximization problem:

$$\max \sum_{t=1}^{3} \beta^{t-1}[\alpha \ln C_t + (1-\alpha) \ln L_t]$$

subject to

$$\begin{aligned} \sum_{t=1}^{3} \frac{1}{(1+r)^{t-1}} C_t &\leq A_1 + \sum_{t=1}^{2} \frac{1}{(1+r)^{t-1}} w_t (L_0 - L_t) \\ L - L_1 &\geq 0 \\ L - L_2 &\geq 0 \end{aligned}$$

(a) Derive the Marshallian and Frischian labor supply functions. Consider cases with both interior and corner solutions for different periods.

(b) Within each case, characterize λ and derive expressions for changes in λ in response to changes in (w_1, w_2) and A_1 (initial wealth). *Note that in this case, λ is time-invariant as it would be the Lagrange multiplier associated with the lifetime budget constraint.*

(c) **Bonus:** Describe how you would solve the model (describe the optimal policy functions) if you observe (w_1, w_2, A_1) and knew that

$$\begin{bmatrix} \alpha \\ \beta \\ r \\ L_0 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.9 \\ 0.08 \\ 8700 \end{bmatrix}$$