Problem Set 2

Erich Battistin
University of Maryland, CEPR, FBK-IRVAPP and IZA

Applied Econometrics I - AREC623

These problems are intended to encourage critical thinking about the various concepts we have covered, provide you with opportunities to practice at manipulating random variables and their distributions, and highlight how different models can be applied in practical scenarios. All the questions can be addressed using the material we have covered in previous lectures, but there may be instances where certain parts of a question present challenges. In such cases, do not hesitate to bring up your questions during your TA session. Collaborate with your peers to collectively tackle the challenging aspects of these questions.

Question 1

One important pattern that has emerged in the literature over the last two decades is that households allocate a substantial portion of their expenditures to a limited number of items. In simple terms, individual households have increasingly focused their spending on a select few items as time has passed. On the other hand, however, households buy very different items, meaning that the aggregate spending concentration has declined over time.

To investigate these trends, you have been provided with panel data on the shopping habits of a large sample of households over three years: 2012, 2017 and 2022. Let i denote a household, c a spending category (e.g., beverages, cosmetics), and j a specific product within the category. Define E_{icjt} as the total expenditure of household i in year t on product j in category c. The share of household i spending on product j is:

$$X_{icjt} \equiv \frac{E_{icjt}}{\sum_j E_{icjt}}, \tag{1}$$

which represents the proportion of total spending allocated to product j.

Denote by C_{ict} the *variance* of (1) calculated across products of category c. This variance is a measure of household's i product concentration in year t for purchases made in category c, with higher variance signalling increased concentration. Using this notation, the empirical evidence on shopping patterns documented in the literature can be summarized as follows:

while each household is directing more of its spending toward its preferred spending categories, there is also a growing diversity in the categories that different households choose to consume.

Assume that the list of categories c remains the same between 2012 and 2022 and that it is possible to track all households over this period (i.e., no attrition in the panel). I would like you to work on the following questions.

(a) Consider the following regression function:

$$C_{ict} = \alpha_0 + \alpha_1 \mathbb{1}(t = 2017) + \alpha_2 \mathbb{1}(t = 2022) + \varepsilon_{ict},$$

where, for example, $\mathbb{1}(t = 2022)$ denotes a variable which takes value one for observations in 2022 and zero otherwise. Demonstrate that the OLS estimation of this regression function yields a value of the coefficient α_1 that approximates the quantity:

$$E\left(C_{ict} | \mathbb{1}(t = 2017) = 1\right) - E\left(C_{ict} | \mathbb{1}(t = 2012) = 1\right).$$

Make all steps of your proof explicit and **use at most one page to answer**.

(b) Consider an indicator Z_i, measured in 2012 (the first wave of data), taking value one if household i had any children and zero otherwise. Change the regression function in point (a) by adding Z_i as follows:

$$C_{ict} = \alpha_0 + \alpha_1 \mathbb{1}(t = 2017) + \alpha_2 \mathbb{1}(t = 2022) + \alpha_3 Z_i + \varepsilon_{ict}.$$

Derive the expression for the quantity approximated by the OLS estimate of the coefficient α_1 in this regression function. Make all steps of your proof explicit and **use at most one page to answer**.

(c) Consider an indicator Z_{it}, measured in each year t, taking value one if household i in year t had any children and zero otherwise. Change the regression function in point (a) by adding Z_{it}, as follows:

$$C_{ict} = \alpha_0 + \alpha_1 \mathbb{1}(t = 2017) + \alpha_2 \mathbb{1}(t = 2022) + \alpha_3 Z_{it} + \varepsilon_{ict},$$

where the assumption $\alpha_3 \neq 0$ is maintained throughout. Using your solutions to points (a) and (b), state under which conditions the OLS estimate of the coefficient α_1 in this regression function approximates the quantity:

$$E\left(C_{ict} | \mathbb{1}(t = 2017) = 1\right) - E\left(C_{ict} | \mathbb{1}(t = 2012) = 1\right).$$

Use at most 10 lines to answer this question.

(d) How do you interpret the coefficient α_1 of the equation in point (a)? What is the sign you would expect to see for this coefficient given the context on spending patterns explained in the introduction of this problem set? Why? **Use at most** 10 **lines to answer this question**.

(e) Consider the following regression function:

$$C_{ict} = \beta_1 + \sum_{h=2}^{n} \beta_h \mathbb{1}(h = i) + \eta_{ict},$$

where $\mathbb{1}(h = i)$ denotes a variable which takes value one for household i and zero otherwise. Derive the expression for the quantity approximated by the OLS estimate of the coefficients β_h in this regression function. Make all steps of your proof explicit and **use at most one page to answer this part**. Moreover, **use at most** 10 **lines** to answer the following additional question: what is the sign you would expect to see for these coefficients given the context on spending patterns explained in the introduction of this problem set? Why?