Homework #1

Byoungchan Lee
ECON5345 Spring 2026
Due: Feb 8, 2026, 10:00 PM

Reminder: Empirical and simulation exercises must be submitted with "stand-alone" code and data (if any) in an appropriate format. The homework is due Sunday (data, code, and word-processed report by email to **the instructor before 10 PM Sunday**). Also, submit your word-processed report via **Gradescope**.

1. (Working, 1960). Problem 8.3 in Romer's textbook (4th edition). Actual data give not consumption at a point in time, but average consumption over an extended period, such as a quarter. This problem asks you to examine the effects of this fact.

Suppose that monthly consumption follows a random walk: $C_t = C_{t-1} + e_t$, where e_t is white noise. Suppose, however, that the data provide average consumption over three-period intervals (quarterly); that is, one observes $\frac{C_1+C_2+C_3}{3}$, $\frac{C_4+C_5+C_6}{3}$, and so on.

a) Find an expression for the change in measured consumption from one three-period interval to the next in terms of the e's.
b) Is the change in measured consumption uncorrelated with the previous value of the change in measured consumption? In light of this, is measured consumption a random walk?
c) Given your result in part a), is the change in consumption from one three-period interval to the next necessarily uncorrelated with anything known as of the first of these three-period intervals? Is it necessarily uncorrelated with anything known as the three-period interval immediately preceding the first of the three-period intervals?
d) Suppose that $C_0 = 0$ and $e_t \sim iidN(0,1)$. Simulate the measured consumption series, $\frac{C_1+C_2+C_3}{3}$, $\frac{C_4+C_5+C_6}{3}$, ... for 10,000 periods (HINT: normrnd, cumsum, reshape, and mean functions in Matlab). Construct a series of the change in measured consumption (HINT: diff function in Matlab). Compute and plot ACF/PACF of the change in measured consumption (HINT: autocorr and parcorr in Matlab).
e) Download (seasonally adjusted) monthly and quarterly real personal consumption expenditure of non-durable goods in the US (FRED database). Plot ACF/PACF of the changes of each series. Discuss your results.

- ACF is the autocorrelation function defined as $\rho(k) = \gamma(k)/\gamma(0)$, where $\gamma(k) = cov(x_t, x_{t-k})$.
- PACF is the partial autocorrelation function. $\phi(1) = \rho(1)$, and $\phi(k) = corr\big(x_{t+k} - P(x_{t+k}|x_{t+k-1}, \dots, x_{t+1}), x_t - P(x_t|x_{t+k-1}, \dots, x_{t+1})\big)$ for $k > 1$, where $P(x|y)$ means a projection of x on y.
- For a white noise process, it is known that $\rho(k) = \phi(k) = 0$ for any $k > 0$.

2. This problem is designed to familiarize you with a numerical optimization routine in your favorite software. Specifically, I ask you to replicate the results in Figure 4 of Matejka (REStud, 2016) by solving a constrained numerical optimization problem. To be clear, it is okay if you cannot understand every theoretical detail of rational inattention below. My objective is to provide you with numerical exercises based on macroeconomic questions.

a) Consider a seller who faces the demand function $p^{-\frac{d+1}{d}}$, where p denotes the price. The unit input cost is 1. Therefore, the profit function is given by $\Pi(p,d) = p^{-\frac{d+1}{d}}(p-1)$. Show that the profit-maximizing pricing strategy $p^*(d) = 1 + d$ under perfect information (i.e., d is observable without any constraints).

b) Suppose that the seller has a prior on the distribution of d: $d \sim Uniform\left(\frac{1}{9}, \frac{1}{2}\right)$. Let's denote the corresponding pdf as $g(d)$. We assume that this seller is rationally inattentive. That is, fully observing the realization of d and determining p accordingly is beyond the seller's information processing capacity. Instead, the seller decides on a joint distribution $f(p,d)$ that maximizes the expected profit while satisfying the information constraint. Formally,

$$\max_{f(p,d)} \int\int \Pi(p,d) f(p,d)\, dd\, dp$$

subject to

$$\int f(p,d)dp = g(d) \quad for\ all \quad d,$$
$$f(p,d) \geq 0 \quad for\ all \quad p\ and\ d,$$
$$I(p,d) \equiv \int\int \log_2 \frac{f(p,d)}{f_p(p) f_d(d)} f(p,d)\, dd\, dp \ \leq \kappa.$$

To find a solution to this maximization problem, we consider a discretized optimization problem. Suppose that p and d can take value from $\{p_1, \dots, p_{N_p}\}$ and $\{d_1, \dots, d_{N_d}\}$, respectively. Under the assumptions above, we set $d_1 = \frac{1}{9}$, $d_{N_d} = \frac{1}{2}$, $p_1 = p^*(d_1)$, and $p_{N_p}(d_{N_d}) = p^*(d_{N_d})$. Next, we investigate a probability mass function (pmf) $\{f_{i,j} | 1 \leq i \leq N_p, 1 \leq j \leq N_d\}$ such that $f_{i,j} = prob(p = p_i, d = d_j)$ for all i and j. We similarly discretize a pdf $g(d)$ as $\{g_j | 1 \leq j \leq N_d\}$ such that $g_j = prob(d = d_j)$. Our objective is to find the optimal pmf:

$$\max_{\{f_{i,j}\}} \sum_{i,j} \Pi(p_i, d_j) f_{i,j}$$

subject to

$$\sum_i f_{i,j} = g_j \quad for\ all \quad j,$$
$$f_{i,j} \geq 0 \quad for\ all \quad i\ and\ j,$$
$$\sum_{i,j} \log_2 \frac{f_{i,j}}{f_{p,i} f_{d,j}} f_{i,j} \ \leq \kappa,$$

where $f_{p,i} = prob(p = p_i) = \sum_j f_{i,j}$ and $f_{d,j} = prob(d = d_j) = \sum_i f_{i,j}$. We constructed a constrained optimization problem with $N_p * N_d$ variables.

Use uniform grids for both p and d. Also, assume that $N_p = N_d = 20$ (the paper uses 70). $\kappa = 0.5$. Solve the above maximization problem and draw a figure similar to the right panel in Figure 4, Matejka (REStud, 2016).

To help you solve this question, I will provide my code for a similar question in a working paper version of Matejka's paper. In the working paper, he considers the following question. $\Pi = p^{-\theta}(p - \mu)$, where $\theta = 3$. Now, the marginal cost μ is stochastic: $\mu \sim Uniform(0.8, 1.2)$. The optimal pricing strategy under full information is $p_{opt}(\mu) = \frac{\theta}{\theta-1}\mu$. When the seller is rationally inattentive, the optimization problem becomes

$$\max_{f(p,\mu)} \int\int \Pi(p,\mu) f(p,\mu)\, d\mu\, dp$$

subject to

$$\int f(p,\mu) dp = g(\mu) \quad for\ all \quad \mu,$$
$$f(p,\mu) \geq 0 \quad for\ all \quad p\ and\ \mu,$$
$$I(p,\mu) \equiv \int\int \log_2 \frac{f(p,\mu)}{f_p(p) f_\mu(\mu)} f(p,\mu)\, d\mu\, dp \ \leq \kappa.$$

He assumes $\kappa = 1$. My code for the discretized version of this problem is a hint for the problem you need to solve.

In my sample code, I increased `'MaxFunctionEvaluations'` and decreased `'OptimalityTolerance'` and `'StepTolerance'` from their default values. Try a version with the default option and the adjusted option separately. Comment on the differences in results.

3. (Types of data). The level of individual labor productivity (in logarithms) is often modeled as an AR(1) process as follows:

$$y_{it} = \mu_i + \rho y_{i,t-1} + e_{it}, \qquad (*)$$

where μ_i reflects individual-specific factors, such as education, experience, age, etc., and $e_{it} \sim iidN(0, \sigma^2)$. Below, we write $y_{i,t-1}$ as x_{it}.

a) (Cross-section). Suppose that an econometrician observes $\{y_{i1}, y_{i0}\} = \{y_{i1}, x_{i1}\}$ for $i = 1, \ldots, N$. In this case, the data are cross-sectional. Can the econometrician identify ρ in the model (*) above? What if they assume $\mu_i = \mu$ for all i? Does the OLS work in this case when $N \to \infty$?

b) (Timeseries). Suppose that an econometrician observes $\{y_{1t}, y_{1,t-1}\} = \{y_{1t}, x_{1t}\}$ for $t = 1, \ldots, T$. In this case, the data are timeseries. Does the OLS work concerning the estimation of ρ in this case, when $T \to \infty$?

Suppose that an econometrician observes $\{y_{it}, x_{it}\}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In this case, we have panel data. When N is large but T is small, we often call it a short panel. When T is large, the panel estimation usually works well for a similar reason to the timeseries case.

c) (Nickell bias). Consider a short panel. When a regression model includes a lagged dependent variable (in our case, $y_{i,t-1}$) and fixed effects (μ_i), it is known that the OLS estimator is often biased. This bias is often referred to as Nickell bias (Nickell, 1981, Econometrica). This question asks you to simulate the model (*), estimate ρ based on the simulated short panel data, and assess the size of bias for different values of N and T. For simulation exercises, use $\rho = 0.9$, $\sigma = 1$, and $\mu_i = 0$ for all i for simplicity. For a given value of N and T, simulate panel data as follows:

(i) Draw y_{i0} from $N\left(0, \frac{\sigma^2}{1-\rho^2}\right)$, which is the stationary distribution of y_{it} when $\mu_i = 0$. We will learn more about the stationarity and an AR(1) process later in this course (HINT: normrnd function in Matlab).

(ii) Generate y_{it} for $t = 1$ by drawing e_{i1} from $N(0, \sigma^2)$ and using the model (*).

(iii) Repeat (ii) for $t = 2, \dots, T$ and $i = 1, \dots, N$.

The OLS with fixed effects are often estimated using demeaned data. That is, the model (*) implies that:

$$y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = \rho\left(y_{i,t-1} - \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}\right) + \left(e_{it} - \frac{1}{T}\sum_{t=1}^{T} e_{it}\right).$$

Thus, we can estimate ρ by running a pooled OLS of $y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ on $y_{i,t-1} - \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$.

Simulate the data for $T = 2$ and $N = 100$. What is your estimated ρ? How about when $N = 500$ and $N = 1000$? Does the OLS estimate of ρ converge to the true value when N ncreases?

d) Repeat c) with $T = 5$ and $T = 20$. Use $N = 1000$ for this question. Comment on your results.