Empirical IO: Problem Set - 1

Not all questions need to be answered; only five are required. The choices are given below:

1) Do either Question 1 or Question 2
2) Do either Question 3 or Question 4
3) Do either Question 5 or Question 6
4) Do either Question 7 or Question 8
5) Do either Question 9 or Question 10

The total marks for the problem set are 150. This will then be normalized to 20 using:

$$\text{Final Grade} = \frac{\text{your group's marks}}{150} \times 20.$$

What to submit: The problem set should be submitted in LaTeX and the R script. Your R code should be capable of (re-)producing the numbers that you present in the LaTeX tables.
Submission Deadline: The problem set must be submitted by **11-October-2025 23:59:59**. A one-hour delay leads to a deduction of 5 points (i.e., the maximum grade will be 15), a two-hour delay results in a deduction of 10 points, . . . , and a delay of at least four hours results in a deduction of 20 points. The solutions will be released on 12-October-2025 as soon as the deadline passes. When you submit the problem set, please email both me and Tushar.

Q1. Cournot Cement Example — Equilibrium Computation and OLS Drawbacks

A: Theoretical Model [15 marks]

Consider a single market and period. Set all shocks to zero, and ignore the floor operator.

$$\text{Demand: } P = \beta_0 - \beta_1 Q, \tag{1}$$

$$\text{Marginal cost: } MC(q) = \gamma_0^{MC}. \tag{2}$$

1) Write down a firm's variable profit function (for now ignore fixed costs) under the given demand and marginal cost as a function of q. Note that $Q = q + \tilde{Q}$, where $\tilde{Q}$ is all other firms' aggregated output. $\tilde{Q}$ will appear in the profit function but is beyond the firm's control. [1 mark]

2) Derive the first-order condition for a firm under Cournot competition. Note that $Q = q + \tilde{Q}$, where $\tilde{Q}$ is all other firms' total output. [1 mark]

3) Using the first-order condition, argue that the equilibrium level of output must be identical for all firms, therefore $q = \frac{Q}{N}$. [1 mark]

4) Solve for q and P using the FOC you derived and the demand function. You should obtain

$$q^* = \frac{\beta_0 - \gamma_0^{MC}}{\beta_1(1+N)}.$$

[3 marks]

5) Show that $\frac{dq^*}{dN} < 0$. [1 mark]

6) Show that the variable profit of the firm in equilibrium can be written as:

$$\pi^{variable} = \beta_1 \cdot (q^*)^2 . \tag{3}$$

[Hint: Use the FOC $P + \frac{\partial P}{\partial q} \cdot q - \gamma_0^{MC} = 0$ to substitute for P in the variable profit function. Do not forget to evaluate $\frac{\partial P}{\partial q}$.] [3 marks]

7) Show that the variable profit function is strictly decreasing in N. [2 marks]

8) Assume that

$$FC = \gamma_{FC,0}. \tag{4}$$

How many firms enter in equilibrium? Explain the equation you use to find the equilibrium number of firms that enter. [Hint: Find where total profit = variable profit − fixed cost = 0. Assume continuous N.] [3 marks]

B: Empirical Model [15 marks]

Create one regression table for all parts. Clearly label them either in the column or manually add table notes. Assume linear demand, quadratic variable costs, and free entry. Consider the following regression versions of the equations we derived earlier:

$$\text{Demand: } Q_{mt} = \frac{\beta_0}{\beta_1} - \frac{1}{\beta_1} \cdot P_{mt} + \frac{\beta_X}{\beta_1} \cdot X_{mt}^D + \frac{\epsilon_{mt}}{\beta_1}, \tag{5}$$

$$\text{Cournot Condition: } P_{mt} = \gamma_0^{MC} + \beta_1 \frac{Q_{mt}}{N_{mt}} + \gamma_X^{MC} \cdot X_{mt}^{MC} + \mu_{mt}, \tag{6}$$

$$\text{Entry: } \left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \frac{\gamma_0^{FC}}{\beta_1} + \frac{\gamma_X^{FC}}{\beta_1} \cdot X_{mt}^{FC} + \frac{e_{mt}}{\beta_1}. \tag{7}$$

Here, ϵ_{mt}, μ_{mt}, and e_{mt} are independent from each other and also independent of X_{mt}^D, X_{mt}^{MC}, and X_{mt}^{FC}. However, note that P_{mt}, Q_{mt}, and N_{mt} are endogenous variables (as solved in the previous question).

1) Write down the demand equation and entry condition as linear regressions such that coefficients are linear. Note that $\frac{1}{\beta}$ enters non-linearly. [1 mark]

2) What are valid instruments for Q_{mt}, $\frac{Q_{mt}}{N_{mt}}$, and P_{mt} and why? Write the condition under which these are valid. [1 mark]

3) Load the dataset called `cement_example_pset.csv`. Provide the code in the R script and create the summary statistics table. [2 marks]

4) Estimate the entry equation. [2 marks]

5) Run OLS for the Cournot Condition. Then instrument for $\frac{Q_{mt}}{N_{mt}}$ using X_{mt}^{FC} in the Cournot Condition regression. Run the regressions. [3 marks]

6) Run OLS for the demand regression. Then instrument for P_{mt} using X_{mt}^{MC} in the demand regression. Run the regression. [3 marks]

7) Recover the estimates of the original parameters γ_0^{FC}, γ_X^{FC}, β_0, β_X, and β_1. [3 marks]

Q2. Drawbacks of Running OLS under the Cost-Minimization Model [30 marks]

A: Theoretical Model [15 marks]

Let the production function be $Y = A\,K^{\alpha_K} L^{\alpha_L} M^{\alpha_M}$ with input prices (W_K, W_L, W_M).

1) Write the cost-minimization problem that defines the cost function $C(Y)$. [1 mark]

2) Write the Lagrangian. [1 mark]

3) Derive the FOCs with respect to K, L, and M. [3 marks]

4) Solve for K, L, M from the FOCs and substitute into the expenditure $C = W_L L + W_K K + W_M M$. [3 marks]

5) Use the constraint $Y = AK^{\alpha_K} L^{\alpha_L} M^{\alpha_M}$ to solve for the multiplier λ. [1 mark]

6) Substitute λ back into the FOCs to obtain closed-form demands $K(Y, \mathbf{W})$, $L(Y, \mathbf{W})$, $M(Y, \mathbf{W})$. [3 marks]

7) Obtain the cost function $C(Y, \mathbf{W})$. Then compute $MC(Y) = \frac{dC}{dY}$ and discuss the sign of $C''(Y)$ in terms of $\alpha \equiv \alpha_K + \alpha_L + \alpha_M$. [3 marks]

B: Empirical Model [15 marks]

Create one regression table for all parts. Clearly label them either in the column or manually add table notes.

Let the log-linear production function be

$$\log(Y_{it}) = \alpha_K \cdot \log(K_{it}) + \alpha_L \cdot \log(L_{it}) + \alpha_M \cdot \log(M_{it}) + \omega_{it} + \epsilon_{it}. \tag{8}$$

1) Upload the dataset: `cost_minimization_data.csv`. The code should be in the `R-script` and create a summary-statistics table for the dataset. [1 mark]

2) Assume that you observe ω_{it}. Run regression 8 using `feols` and report the estimates. [2 marks]

3) Suppose you observe $R_{it} = P_{it} \times Y_{it}$. Then run the following regression:

$$\log(R_{it}) = \beta_K \cdot \log(K_{it}) + \beta_L \cdot \log(L_{it}) + \beta_M \cdot \log(M_{it}) + \omega_{it} + \epsilon^*_{it}. \tag{9}$$

Are β_K, β_L, and β_M different from α_K, α_L, and α_M? What could be the reason? [3 marks]

4) Suppose now $\log(P_{it}) = \log(P_{industry,t}) + \mu_{it}$. However, $P_{industry,t}$ is unknown. Write down a modification of regression 9 that can absorb the variation explained by $P_{industry,t}$ in P_{it} and yield better estimates for elasticities (coefficients of regression 9). Run the modified regression using `feols`. Suppose you observe $P_{industry,t}$ (provided in the dataset). Then write down and run another regression that can recover the elasticities. [4 marks]

5) Now assume ω_{it} is not observed and re-run all three regressions again. Are the estimates different from the results obtained by running 8? Why? Also provide reasons behind the direction of the bias between 8 and the modified regression of 8 where ω_{it} is not observed. [5 marks]

Q3. Drawbacks of Observing Physical Output vs. Revenue [30 marks]

A: Revenue Regression with Isoelastic Demand Assumption [12 marks]

Create one regression table for all parts. Clearly label them either in the column or manually add table notes.

Start from the physical-output equation (logs):

$$y_i = \alpha_L \ell_i + \alpha_K k_i + \omega_i + \varepsilon_i.$$

Assume you observe revenue $R_i = P_i Y_i$ and inverse demand $\log(P_i) = a_i - \beta \log(Y_i)$ with $\beta > 1$.

1) Derive an expression for $r_i \equiv \ln R_i$ and show that a revenue regression can be written as

$$r_i = \alpha_L^\star \ell_i + \alpha_K^\star k_i + \omega_i^\star + \varepsilon_i^*,$$

with $\alpha_j^\star = \gamma\, \alpha_j$. Derive γ from the model. [2 marks]

2) The dataset `revenue_production.csv` contains Y, R, L, K, and TFP `omega`. Run the following regressions:

   a) the physical-output regression: y_i on ℓ_i, k_i; [4 marks]

   b) the revenue regression: r_i on ℓ_i, k_i, ω_i (assuming TFP is observed). [4 marks]

3) Discuss how $\alpha_L^\star$ vs. α_L affects conclusions about returns to scale, and how the interpretation of $\omega_i^\star$ differs from ω_i. [2 marks]

B: Simultaneity and OLS Bias [6 marks]

Assume

$$\text{(log-PF)} \quad y_i = \alpha_0 + \alpha_L \ell_i + \omega_i, \qquad \text{(log-LD)} \quad \ell_i = \gamma_0 + \gamma_1\, y_i.$$

Please note that since there is is a constant term in (log-PF) therefore $E[\omega_i] = 0$.

1) Solve the system for y_i and ℓ_i in terms of $\omega_i, \alpha_0, \alpha_l \gamma_0, \gamma_1$.

2) Derive $\text{Cov}(\ell_i, \omega_i)$ and $\text{Cov}(\ell_i, y_i)$. *(State any mean/independence assumptions you use.)*

3) Interpret the sign of $\text{Cov}(\ell_i, \omega_i)$ and the implication for OLS bias when regressing y_i on ℓ_i.

C: When Measurement Noise is Correlated with Inputs [12 marks]

Let

$$\ln P_i = \ln P_0 + \mu_i, \qquad y_i = \alpha_0 + \alpha_L \ell_i + \omega_i + \varepsilon_i, \qquad r_i = \ln P_i + y_i.$$

Assume all residuals have mean zero, unit variance, and are mutually independent.

1) Write the estimating equation for regressing r_i on ℓ_i in the form

$$r_i = \beta_0 + \beta_1 \ell_i + e_i.$$

Find β_0, β_1, and e_i in model primitives. What is $\text{Var}(e_i)$? [2 marks]

2) Under the conditions $\mathbb{E}[\mu_i(\ell_i - \bar{\ell})] = \mathbb{E}[\omega_i(\ell_i - \bar{\ell})] = \mathbb{E}[\varepsilon_i(\ell_i - \bar{\ell})] = 0$, show that $\hat{\beta}_1^{\text{OLS}} = \alpha_L$. [2 marks]

3) Now let $\ell_i = \gamma_0 + \gamma_\mu \mu_i + u_i$ with $\gamma_\mu < 0$. Express

$$\mathbb{E}\big[e_i \cdot (\ell_i - \bar{\ell})\big]$$

in terms of the model parameters. Here u_i has unit variance and is independent of everything. [2 marks]

4) Show that $\hat{\beta}_1 < \alpha_L$ and explain the intuition. [2 marks]

5) Further assume $\ell_i = \gamma_0 + \gamma_\mu \mu_i + \gamma_\omega \omega_i + e_i$ with e_i independent of everything. Show that [2 marks]

$$\mathbb{E}\big[e_i(\ell_i - \bar{\ell})\big] = \gamma_\mu + \gamma_\omega.$$

6) State the condition (an inequality in γ_ω, given $\gamma_\mu < 0$) under which $\hat{\alpha}_L^{\text{OLS}} < \alpha_L$. [2 marks]

Question 4

A: Three-Stage Least Squares vs. IV Method [20 marks]

Now suppose the data-generating process is given by:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{10}$$
$$x_i = \gamma_0 + \gamma_1 z_i + e_i, \tag{11}$$
$$z_i = \delta_0 + \delta_1 w_i + u_i, \tag{12}$$

with

$$\varepsilon_i = \theta_0 e_i + \mu_{1i}, \tag{13}$$
$$e_i = \rho_0 u_i + \mu_{2i}, \tag{14}$$
$$w_i \perp\!\!\!\perp \mu_{1i}, \mu_{2i}, \text{ and } u_i. \tag{15}$$

[Hint: if $A_i \perp\!\!\!\perp B_i \Rightarrow \text{Cov}(A_i, B_i) = 0$]

Please note that since there is is a constant term in (log-PF) therefore $E[\epsilon_i] = E[e_i] = E[u_i] = 0$. This also ensures that $E[\mu_{1i}] = E[\mu_{2i}] = 0$.

1) Is x_i exogenous? Why? [2 marks]

2) Is z_i a valid instrument for x_i? Why? [2 marks]

3) Is w_i a valid instrument for z_i? Why? [2 marks]

4) Can we say w_i is a valid instrument for x_i? Why? [2 marks]

[**HINT: To show whether something is exogenous you need to show that the error term is not correlated with the main dependent variable. To show something is an instrument (or not an instrument) you need to show that the exogeneity condition and relevance conditions are satisfied (or one of those is not satisfied).**]

5) Consider the following IV estimator:

$$\hat{\beta}^{IV} = \frac{S_{zy}}{S_{zx}}.$$

Is this a consistent estimator of β? Prove it. Find the bias (from its zero). [3 marks]

6) Now consider the following IV estimator:

$$\hat{\beta}_w^{IV} = \frac{S_{wy}}{S_{wx}}.$$

Is this a consistent estimator of β? Find the bias (from its zero). [3 marks]

7) Next, consider the following 3–stage least squares procedure.

Stage 1:

$$z_i = \delta_0 + \delta_1 w_i + u_i,$$

obtain $\hat{z}_i = \delta_0^{OLS} + \delta_1^{OLS} w_i$.

Stage 2:

$$x_i = \gamma_0 + \gamma_1 \hat{z}_i + v_i,$$

obtain $\hat{x}_i = \gamma_0^{2SLS} + \gamma_1^{2SLS} z_i$, where γ_1^{2SLS} is the OLS estimator above.

Stage 3:

$$y_i = \alpha + \beta \hat{x}_i + \varepsilon_i,$$

to obtain $\hat{\beta}^{3SLS}$ as the OLS estimator of the above. Is $\hat{\beta}^{3SLS}$ consistent? Prove whether it is or not. [6 marks]

B: Control Function within IV [2 marks]

Prove that in the following setup, the obtained residuals $e_i = X_i - \delta_0 - \delta_1 Z_i$ are a valid control function.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{16}$$

$$X_i = \delta_0 + \delta_1 Z_i + e_i, \tag{17}$$

$$\text{such that} \quad \text{Cov}(Z_i, e_i) = \text{Cov}(Z_i, \epsilon_i) = 0. \tag{18}$$

C: Tricky Regressions [8 marks]

Consider the following setup. Initially, the econometrician had data on Y_i and X_i and then they were able to obtain data on W_i.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{19}$$

$$\text{Cov}(X_i, \epsilon_i) \neq 0, \tag{20}$$

$$\epsilon_i = \gamma W_i + u_i, \tag{21}$$

$$\text{such that} \quad \text{Cov}(X_i, u_i) = 0, \quad \text{Cov}(W_i, X_i) \neq 0, \quad \text{and } \text{Cov}(W_i, u_i) \neq 0. \tag{22}$$

1) Is W_i an instrument for X_i? If yes, why? If no, then why? [2 marks]

2) Is W_i a control function for X_i? If yes, why? If no, then why? [2 marks]

3) Combine both relations, i.e., regressions 20 and 22, to produce a single regression. List the endogenous variable(s). [2 marks]

4) Describe a strategy to consistently estimate the regression you obtained in the previous part. [2 marks]

Question 5 [30 points]

The data file `spanish dairy farms.dta` [use package `haven`, or click on `Import Dataset` and select the option to load from `Stata`] contains annual information from 247 dairy farmers in the region of Asturias (Spain) between 1993 and 1998. The dataset includes information on the production of milk in physical units, the number of cows, the amount of feed, labor, and land. Consider a Cobb–Douglas function for the production of milk in terms of the following inputs: cows, feed, labor, and land:

$$\text{MILK} = A \cdot \text{COWS}^{\alpha_C}\text{FEED}^{\alpha_F}\text{LABOR}^{\alpha_L}\text{LAND}^{\alpha_D}.$$

Use this dataset to implement the following estimators and hypothesis tests. Provide the code in `R` and the table of estimation results in LaTeX.

Question 5.1 (4 points)

a) Write the production function in logarithms. Interpret it as a linear regression model. [1]

b) Estimate by OLS with time dummies. Comment on the results. [2]

c) Test the null hypothesis $\alpha_C + \alpha_F + \alpha_L + \alpha_D = 1$. Comment on the results. [1]

Question 5.2 (6 points)

a) Estimate by Fixed Effects with time dummies. Comment on the results.

b) Test the null hypothesis $\alpha_C + \alpha_F + \alpha_L + \alpha_D = 1$. Comment on the results.

c) Test the null hypothesis of no time-invariant unobserved heterogeneity: $\eta_i = 0$ for every firm i. Comment on the results.

Question 5.3 (6 points)

a) Estimate by Fixed Effects – Cochrane–Orcutt with time dummies. Comment on the results.

b) Test the null hypothesis $\alpha_C + \alpha_F + \alpha_L + \alpha_D = 1$. Comment on the results.

c) Test the four over-identifying restrictions of the model. Comment on the results.

Question 5.4 (6 points)

a) Estimate by Arellano–Bond without time dummies and non-serially correlated transitory shock. Comment on the results.

b) Test the null hypothesis $\alpha_C + \alpha_F + \alpha_L + \alpha_D = 1$. Comment on the results.

Question 5.5 (6 points)

a) Estimate by Arellano–Bond without time dummies but with an AR(1) transitory shock. Comment on the results.

b) Test the null hypothesis $\alpha_C + \alpha_F + \alpha_L + \alpha_D = 1$. Comment on the results.

c) Test the four over-identifying restrictions of the model. Comment on the results.

QUESTION 5.6 (2 points)

Based on the previous results, select your preferred estimates of the production function. Explain your choice.

Question 6 [30 points]

The data file `INDIAKLEMS09012024.csv` contains annual information from multiple industries in India. The dataset includes information on gross output (GO_r), capital (K_r), employment (EMP), and intermediate inputs of energy (II_E_r), materials (II_M_r), and services (II_S_r). Consider the following Cobb–Douglas production function:

$$GO_{it} = A \cdot K_{it}^{\alpha_K} \cdot EMP_{it}^{\alpha_L} \cdot II\_E_{it}^{\alpha_E} \cdot II\_M_{it}^{\alpha_M} \cdot II\_S_{it}^{\alpha_S}.$$

Taking logs, the regression specification is:

$$\log(GO_{it}) = \alpha_K \log(K_{it}) + \alpha_L \log(EMP_{it}) + \alpha_E \log(II\_E_{it}) + \alpha_M \log(II\_M_{it}) + \alpha_S \log(II\_S_{it}) + \omega_{it} + \varepsilon_{it}.$$

Use this dataset to implement the following estimators and hypothesis tests. Provide the code in `R` and report the estimation results.

Question 6.1 (4 points)

a) Write the production function in logarithms. Interpret it as a linear regression model. [1]

b) Estimate by OLS with time dummies. Comment on the results. [2]

c) Test the null hypothesis $\alpha_K + \alpha_L + \alpha_E + \alpha_M + \alpha_S = 1$. Comment on the results. [1]

Question 6.2 (6 points)

a) Estimate by Fixed Effects with time dummies. Comment on the results.

b) Test the null hypothesis $\alpha_K + \alpha_L + \alpha_E + \alpha_M + \alpha_S = 1$. Comment on the results.

c) Test the null hypothesis of no time-invariant unobserved heterogeneity: $\eta_i = 0$ for every industry i. Comment on the results.

Question 6.3 (6 points)

a) Estimate by Fixed Effects – Cochrane–Orcutt with time dummies. Comment on the results.

b) Test the null hypothesis $\alpha_K + \alpha_L + \alpha_E + \alpha_M + \alpha_S = 1$. Comment on the results.

c) Test the four over-identifying restrictions of the model. Comment on the results.

Question 6.4 (6 points)

a) Estimate by Arellano–Bond without time dummies and non-serially correlated transitory shock. Comment on the results.

b) Test the null hypothesis $\alpha_K + \alpha_L + \alpha_E + \alpha_M + \alpha_S = 1$. Comment on the results.

Question 6.5 (6 points)

a) Estimate by Arellano–Bond without time dummies but with an AR(1) transitory shock. Comment on the results.

b) Test the null hypothesis $\alpha_K + \alpha_L + \alpha_E + \alpha_M + \alpha_S = 1$. Comment on the results.

c) Test the four over-identifying restrictions of the model. Comment on the results.

Question 6.6 (2 points)

Based on the previous results, select your preferred estimates of the production function. Explain your choice.

Question 7 [30 Marks]

A: Arellano–Bond with Serial Correlation [10 Marks]

Consider the following set of assumptions:

$$
\begin{aligned}
\text{PF} \quad & y_{it} = \alpha_L \cdot \ell_{it} + \alpha_K \cdot k_{it} + \eta_i + \delta_t + u_{it} && (23)\\
\text{LD} \quad & \ell_{it} = \beta_1^{LD} \ell_{i,t-1} + \beta_2^{LD} k_{i,t-1} + \beta_3^{LD} \omega_{it} + \beta_4^{LD} r_{it} && (24)\\
\text{KD} \quad & k_{it} = \beta_1^{KD} \ell_{i,t-1} + \beta_2^{KD} k_{i,t-1} + \beta_3^{KD} \omega_{it} + \beta_4^{KD} r_{it} && (25)\\
\text{FE-1} \quad & \omega_{it} = \eta_i + \delta_t + u_{it} && (26)\\
\text{AR(1)} \quad & u_{it} = \rho \cdot u_{it-1} + a_{it} && (27)
\end{aligned}
$$

1) First, ignore LD and KD and focus on PF, FE-1, and AR(1). Follow the same steps as in the Cochrane–Orcutt estimator to derive the quasi-difference equation. Convert the quasi-difference equation into a linear regression that can be estimated (again, as in Cochrane–Orcutt; relabel the parameters). Call this regression `qdPF`. Do not forget the fixed effects in the derivation. [2 marks]

2) Now use `qdPF` to derive a differenced version of `qdPF`, as in the Arellano–Bond estimator. Call this `DqdPF`. [2 marks]

3) Write down all the endogenous variables in the new regression. Use LD and KD to find instruments for the endogenous variables. No need to prove—just argue. *(Hint: Number of instruments $\geq$ number of endogenous variables > 2.)* [2 marks]

4) Consider `INDIAKLEMS09012024.csv`. Define the production function by

$$VA\_r_{it} = A \cdot K_{it}^{\alpha_K} \cdot L_{it}^{\alpha_L}.$$

Use `pgmm` in the `plm` package to estimate the above function using the Arellano–Bond estimator with serial correlation. Ignore time dummies for the estimation.

Hint: Steps

Step 1 First create all the variables you need in `qdPF`.

Step 2 Create the instruments.

Step 3 Declare/create the `pdata.frame` object.

Step 4 Run `pgmm( qdPF | instruments, [everything else the same as before] )`.

[4 marks]

B: OP Extension — Levinsohn–Petrin [20 Marks]

Levinsohn and Petrin (2023) noted that investment often tends to be zero; therefore, invertibility of f_K is not valid, as many values of TFP ω_{it} are mapped to 0. One must rely on a more flexible input than investment to create the control function. To this purpose, they propose using the material demand equation to formulate the control function. The following system does precisely that:

$$\text{PF} \quad y_{it} = \alpha_L \cdot \ell_{it} + \alpha_K \cdot k_{it} + \alpha_M \cdot m_{it} + \omega_{it} + e_{it} \tag{28}$$
$$\text{LD} \quad \ell_{it} = f_L(k_{i,t}, \omega_{it}, r_{it}) \tag{29}$$
$$\text{KD} \quad i_{it} = f_K(k_{i,t}, \omega_{it}, r_{it}) \tag{30}$$
$$\text{MD} \quad m_{it} = f_M(k_{i,t}, \omega_{it}, r_{it}) \tag{31}$$
$$\text{LP-1} \quad \omega_{it} = \phi_t(k_{i,t}, m_{it}, r_{it}) \tag{32}$$
$$\text{OP-2} \quad r_{it} = r_t \quad \text{(can use fixed effects)} \tag{33}$$
$$\text{OP-3} \quad \omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} \tag{34}$$
$$\text{OP-4} \quad k_{it} = (1-\delta)\, k_{it-1} + i_{it-1} \tag{35}$$

1) Suppose $\phi_t = c_K k_{it} + c_M m_{it} + c_{KM}\, k_{it} \cdot m_{it} + \delta_t$. Derive the first-stage (linear) regression for this setup. Highlight which elasticities are identified from the first stage. *[Hint: φ_t from the notes $\neq \phi_t$. Here, ϕ_t is $f_M^{-1}(\dots)$.]* [3 marks]

2) Suppose $\omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} = h(\omega_{it-1}) + \xi_{it}$, where $h(\omega) = \pi_0 + \pi_1\omega + \pi_2\omega^2$. Derive the second-stage regression and argue that the remaining elasticities are identified. You will need m_{it-2} as an instrument for m_{it}. [3 marks]

3) Load the dataset `hospc_nber.csv` (Equipment is material for hospitals). Ignore the parametric assumptions on ϕ_t and h. Use 2, 3, and 4 degree polynomials for ϕ_t/φ_t and 3, 5, and 7 degree polynomials for h. Create one table for all three combinations:

   (a) $\phi_t/\varphi_t \approx 2^{\text{nd}}$ poly, $h \approx 3^{\text{rd}}$ poly [4 marks]

   (b) $\phi_t/\varphi_t \approx 3^{\text{rd}}$ poly, $h \approx 5^{\text{th}}$ poly [4 marks]

   (c) $\phi_t/\varphi_t \approx 4^{\text{th}}$ poly, $h \approx 7^{\text{th}}$ poly [4 marks]

   Note: in all the regressions, the ϕ_t polynomial is interacted with year (treat year as a discrete factor when interacting). Estimate the above specifications.

4) Which case do you prefer and why? Consider the warnings you get while running the regressions as well for this part. This is open-ended. [2 marks]

Question 8 [30 Marks]

A: OP Extension — ACF [20 Marks]

Ackerberg, Caves, and Frazer (2015) note that if labor hiring depends on current TFP (ω_{it}), then labor demand is essentially collinear with the control function. Therefore, one should recover that elasticity in the second stage. To recover this, they assume hiring also has higher frictions than materials and that labor contracts tend to be signed before material demands are made. Consider the following set of assumptions:

$$\begin{aligned}
\text{PF} \quad & y_{it} = \alpha_L\,\ell_{it} + \alpha_K\,k_{it} + \alpha_M\,m_{it} + \omega_{it} + e_{it} && (36)\\
\text{LD} \quad & \ell_{it} = f_L(l_{it-1}, k_{it}, \omega_{it}, r_{it}) && (37)\\
\text{KD} \quad & i_{it} = f_K(k_{it}, \omega_{it}, r_{it}) && (38)\\
\text{ACF-MD} \quad & m_{it} = f_M(k_{it}, \ell_{it}, \omega_{it}, r_{it}) && (39)\\
\text{ACF-1} \quad & \omega_{it} = \phi_t(k_{it}, \ell_{it}, m_{it}, r_{it}) && (40)\\
\text{OP-2} \quad & r_{it} = r_t \quad \text{(can use fixed effects)} && (41)\\
\text{OP-3} \quad & \omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} && (42)\\
\text{OP-4} \quad & k_{it} = (1-\delta)\,k_{it-1} + i_{it-1} && (43)
\end{aligned}$$

1) Suppose $\phi_t = c_K k_{it} + c_M m_{it} + c_L \ell_{it} + \delta_t$. Derive the first-stage (linear) regression for this setup. Highlight that none of the elasticities are identified from the first stage. *[Hint: φ_t from the notes $\neq \phi_t$. Here, ϕ_t is $f_M^{-1}(\dots)$.]* [3 marks]

2) Suppose $\omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} = h(\omega_{it-1}) + \xi_{it}$, where $h(\omega) = \pi_0 + \pi_1\omega + \pi_2\omega^2$. Derive the second-stage regression and argue that all elasticities are identified if one has an instrument for ℓ_{it}. Find the instrument as well. *[Hint: To obtain the instrument use LD and lagged LD.]* [3 marks]

3) Load the dataset `hospc_nber.csv` (Equipment is material for hospitals). Ignore the parametric assumptions on ϕ_t and h. Use 2, 3, and 4 degree polynomials for ϕ_t (or φ_t) and 3, 5, and 7 degree polynomials for h. Create one table for all three combinations:

   (a) $\phi_t/\varphi_t \approx 2^{\text{nd}}$ poly, $h \approx 3^{\text{rd}}$ poly [4 marks]

   (b) $\phi_t/\varphi_t \approx 3^{\text{rd}}$ poly, $h \approx 5^{\text{th}}$ poly [4 marks]

   (c) $\phi_t/\varphi_t \approx 4^{\text{th}}$ poly, $h \approx 7^{\text{th}}$ poly [4 marks]

   Note: in all the regressions, the ϕ_t polynomial is interacted with year (treat year as a discrete factor when interacting). Estimate the above specifications and remember that in the second stage ℓ_{it} needs an instrument.

4) Which case do you prefer and why? Consider the warnings you obtain while running the regressions for this part as well. This is open-ended. [2 marks]

B: Olley–Pakes with Serial Correlation [10 Marks]

Consider the following set of assumptions:

$$\begin{aligned}
\text{PF} \quad & y_{it} = \alpha_L\,\ell_{it} + \alpha_K\,k_{it} + \omega_{it} + e_{it} && (44)\\
\text{LD} \quad & \ell_{it} = f_L(k_{it}, \omega_{it}, r_{it}) && (45)\\
\text{KD} \quad & i_{it} = f_K(k_{it}, \omega_{it}, r_{it}) && (46)\\
\text{OP-1} \quad & \omega_{it} = \phi_t(k_{it}, i_{it}, r_{it}) && (47)\\
\text{OP-2} \quad & r_{it} = r_t \quad \text{(can use fixed effects)} && (48)\\
\text{OP-3} \quad & \omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} && (49)\\
\text{OP-4} \quad & k_{it} = (1-\delta)\,k_{it-1} + i_{it-1} && (50)\\
\text{AR(1)} \quad & \xi_{it} = \rho\,\xi_{it-1} + a_{it} \quad \text{such that } a_{it} \perp\!\!\!\perp k_{it} && (51)
\end{aligned}$$

1) Suppose $\phi_t = c_1 k_{it} + c_2 k_{it}^2 + d \cdot i_{it} + \delta_t$. Derive the first-stage (linear) regression for this setup. Highlight which elasticities are identified from the first stage. *[Hint: φ_t from the notes $\neq \phi_t$. Here, ϕ_t is $f_K^{-1}(\dots)$.]* [2 marks]

2) Suppose $\omega_{it} = \mathbb{E}[\omega_{it} \mid \omega_{it-1}] + \xi_{it} = h(\omega_{it-1}) + \xi_{it}$. Derive the quasi-differenced version of this regression by following the same steps as in the Cochrane–Orcutt estimator. [2 marks]

3) Suppose $h(\omega) = \pi_0 + \pi_1\omega + \pi_2\omega^2$. Derive the second-stage regression. [2 marks]

4) Load the dataset `RFSD.csv` (Ignore variables that are not needed in the question). Ignore the parametric assumptions on ϕ_t and h. Use 2, 3, and 4 degree polynomials for ϕ_t/φ_t and 3, 5, and 7 degree polynomials for h. Create one table for all three combinations:

   (a) $\phi_t/\varphi_t \approx 2^{\text{nd}}$ poly, $h \approx 3^{\text{rd}}$ poly

   (b) $\phi_t/\varphi_t \approx 3^{\text{rd}}$ poly, $h \approx 5^{\text{th}}$ poly

   (c) $\phi_t/\varphi_t \approx 4^{\text{th}}$ poly, $h \approx 7^{\text{th}}$ poly

   Note: in all the regressions, the ϕ_t/φ_t polynomial is interacted with year (treat year as a discrete factor when interacting). [4 marks]

Question 9 [30 Marks]

Part A: 20 Marks

Table 1: LP (2003): Elasticity Estimates

Input	ISIC 311	ISIC 381
Unskilled labor	0.139 (0.010)	0.172 (0.033)
Skilled labor	0.051 (0.009)	0.188 (0.025)
Electricity	0.085 (0.007)	0.081 (0.015)
Fuels	0.023 (0.004)	0.020 (0.011)
Materials	0.500 (0.078)	0.420 (0.091)
Capital	0.240 (0.053)	0.290 (0.094)
No. obs.	6115	1394

1) For each ISIC, which inputs have the largest elasticities? Compare the roles of *skilled labor* and *capital* across the four industries, emphasizing which industry seems capital-intensive vs. which is skilled-labor-intensive. [2 marks]

2) Treat the production function as Cobb–Douglas in the six inputs listed. Comment on returns to scale for each industry, assuming the standard errors of returns to scale are 0.059 and 0.075 across the two industries, respectively. [2 marks]

3) Suppose $A = \exp(\text{TFP}) = 1$ and

$$L_{\text{unskill}} = 100, \quad L_{\text{skill}} = 50, \quad M = 300, \quad K = 100, \quad E = 20, \quad F = 10. \tag{52}$$

   In which industry are (a) unskilled and skilled labor, (b) unskilled labor and capital, and (c) skilled labor and capital most complementary? *[Hint: The degree of complementarity was discussed in class; there is a specific measure for it.]* [3 marks]

4) Assume the same values of inputs and TFP as before. In addition, assume the price of output is $P = 3$, and input prices are $W_{ul} = 5$, $W_{sl} = 25$, $W_E = 2$, $W_F = 8$, $W_M = 1$, and $W_K = 10$. For each industry, which input is least optimally employed? Which input is most optimally employed? Which is the most profitable industry out of all four? *[Hint: The worse FOC holds for an input the worse is its optimal allocation.]* [4 marks]

5) For the same set of input values as before, which industry has the lowest marginal rate of techinical substitution between capital and skilled labor? [3 marks]

6) Given the input prices and the levels of inputs in the previous parts, which inputs are over-utilized in production and which ones are under-utilized across the two industries? *[Hint: What would a profit-maximizing firm do?]* [4 marks]

7) Given the input prices and input values, what are the markups $\frac{P-MC}{MC}$ across the two industries? [2 marks]

Part B

This is a hospital-level production function estimation. The underlying data are a panel in firm and year. The production function being estimated is:

$$Y_{it} = A(q_{it}) \cdot K_{it}^{\alpha_K} \cdot L_{it}^{\alpha_L} \quad \text{s.t.} \quad A(q_{it}) = \exp(a_0 + a_q\, q_{it} + \omega_{it})\,. \tag{53}$$

Note that the inputs are L_{it} and K_{it}. The variable q_{it} is quality, which is not an input but influences total factor productivity.

Table 2: Production Function Estimates

	OLS	FE	OP
Quality effort, α_q	−0.0028	−0.0018	−0.0124
	(0.0007)	(0.0004)	(0.0042)
Capital, β_k	0.4607	0.1788	0.5134
	(0.0209)	(0.0514)	(0.0468)
Labor, β_ℓ	0.6723	0.1855	0.2453
	(0.0149)	(0.0119)	(0.0319)

Standard errors in parentheses.

1) Comment on the covariance of inputs and TFP that may explain the changes in elasticity estimates from OLS to FE. [1 mark]

2) Comment on the covariance of inputs and TFP that may explain the changes in elasticity estimates from FE to OP. [1 mark]

3) Is there a trade-off between inputs? Do hospitals with the above production function estimates face a moral dilemma? Comment. [1 mark]

4) For a hospital with $\omega_{it} = 0$ and $a_0 = 0$, what is the profit-maximizing set of inputs L_{it} and K_{it}, along with the quality choice $q_{it} \geq 0$, when $P = W_k = W_\ell = W_q = 1$? Assume that $C(K, L, q) = W_k \cdot K + W_\ell \cdot L + W_q \cdot exp(q)$ [3 marks]

5) Now suppose everything remains the same as in the previous part, but $P = 1 + \exp(q_{it})$. What is the optimal choice of inputs and quality? [4 marks]

Table 3: LP (2003): Elasticity Estimates

Input	ISIC 321	ISIC 331
Unskilled labor	0.130 (0.024)	0.193 (0.034)
Skilled labor	0.155 (0.026)	0.133 (0.030)
Electricity	0.005 (0.019)	0.047 (0.021)
Fuels	0.038 (0.010)	0.021 (0.014)
Materials	0.500 (0.118)	0.550 (0.086)
Capital	0.180 (0.095)	0.190 (0.090)
No. obs.	1129	1032

Question 10 [30 Marks]

Part A: 20 Marks

1) For each ISIC, which inputs have the largest elasticities? Compare the roles of *skilled labor* and *capital* across the four industries, emphasizing which industry is capital-intensive vs. which is skilled-labor-intensive. [2 marks]

2) Treat the production function as Cobb–Douglas in the six inputs listed. Comment on returns to scale for each industry, assuming the standard errors of returns to scale are 0.113 and 0.157, respectively. [2 marks]

3) Suppose $A = \exp(\text{TFP}) = 1$ and

$$L_{\text{unskill}} = 100, \quad L_{\text{skill}} = 50, \quad M = 300, \quad K = 100, \quad E = 20, \quad F = 10. \tag{54}$$

In which industry are (a) unskilled and skilled labor, (b) unskilled labor and capital, and (c) skilled labor and capital most complementary? *[Hint: The degree of complementarity was discussed in class; there is a specific measure for it.]* [3 marks]

4) Assume the same values of inputs and TFP as before. In addition, assume the price of output is $P = 3$ and input prices are $W_{ul} = 5$, $W_{sl} = 25$, $W_E = 2$, $W_F = 8$, $W_M = 1$, and $W_K = 10$. For each industry, which input is least optimally employed? Which input is most optimally employed? Which is the most profitable industry out of all four? *[Hint: The worse FOC holds for an input the worse is its optimal allocation.]* [4 marks]

5) For the same set of input values as before, which industry has the lowest marginal rate of substitution between capital, skilled labor, and unskilled labor? [3 marks]

6) Given the input prices and the levels of inputs, which inputs are over-utilized in production and which ones are under-utilized across the two industries? *[Hint: What would a profit-maximizing firm do?]* [4 marks]

7) Given the input prices and input values, what are the markups $\frac{P-MC}{MC}$ across the two industries? [2 marks]

Part B

This is a hospital-level production function estimation. The underlying data are a panel by firm and year. The production function being estimated is:

$$Y_{it} = A(q_{it}) \cdot K_{it}^{\alpha_K} \cdot L_{it}^{\alpha_L} \quad \text{s.t.} \quad A(q_{it}) = \exp(a_0 + a_q \cdot q_{it} + \omega_{it}) . \tag{55}$$

Note that the inputs are L_{it} and K_{it}. The variable q_{it} denotes quality, which is not an input but influences total factor productivity.

Table 4: Production Function Estimates

	OLS	FE	OP
Quality effort, α_q	-0.0018	-0.0028	-0.0124
	(0.0004)	(0.0007)	(0.0042)
Capital, β_k	0.1788	0.4607	0.5134
	(0.0514)	(0.0209)	(0.0468)
Labor, β_ℓ	0.1855	0.6723	0.2453
	(0.0119)	(0.0149)	(0.0319)

Standard errors in parentheses.

1) Comment on the covariance between inputs and TFP that may explain the changes in elasticity estimates from OLS to FE. [1 mark]

2) Comment on the covariance between inputs and TFP that may explain the changes in elasticity estimates from FE to OP. [1 mark]

3) Is there a trade-off between inputs? Do hospitals with the above production function estimates face a moral dilemma? Comment. [1 mark]

4) For a hospital with $\omega_{it} = 0$ and $a_0 = 0$, what is the profit-maximizing set of inputs L_{it} and K_{it}, along with the quality choice $q_{it} \geq 0$, when $P = W_k = W_\ell = W_q = 1$? Assume that $C(K, L, q) = W_k \cdot K + W_\ell \cdot L + W_q \cdot exp(q)$ [3 marks]

5) Now suppose everything remains the same as in the previous part, but $P = 1 + \exp(q_{it})$. What is the optimal choice of inputs and quality? [4 marks]