Project 1: Linear Panel Data Models

Due: Sunday October 4th, 2020 at 22:00

1 Fixed-Effects Theory

This question is based on the model considered in Mundlak (1961). Let

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1.1}$$

where $\mathbf{x}_{it}$ and $\boldsymbol{\beta}$ are $K \times 1$ vectors, and u_{it} an unobserved scalar random component, which we impose assumptions on below. Each α_i is here an "individual-specific fixed effect" in the sense of being a nonrandom intercept specific to i.

The system (1.1) may be written in the stacked form

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u},$$

where we have defined

$$\underbrace{\mathbf{y}}_{NT\times 1} := \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix}, \quad \underbrace{\mathbf{X}}_{NT\times K} := \begin{bmatrix} \mathbf{x}_{11}' \\ \vdots \\ \mathbf{x}_{T1}' \\ \vdots \\ \mathbf{x}_{N1}' \\ \vdots \\ \mathbf{x}_{NT}' \end{bmatrix}, \quad \underbrace{\mathbf{u}}_{NT\times 1} := \begin{bmatrix} u_{11} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix},$$

$$\underbrace{\mathbf{D}}_{NT\times N} := \begin{bmatrix} \mathbf{1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{1} \end{bmatrix}, \quad \underbrace{\boldsymbol{\alpha}}_{N\times 1} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \quad \underbrace{\boldsymbol{\gamma}}_{(N+K)\times 1} := \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix}, \quad \underbrace{\mathbf{Z}}_{NT\times(N+K)} := \begin{bmatrix} \mathbf{D} & \mathbf{X} \end{bmatrix},$$

and the displayed $\mathbf{0}$ and $\mathbf{1}$ are short for $T \times 1$ vectors of zeros and ones, respectively. Assume that $\mathrm{E}[\mathbf{u}|\,\mathbf{Z}] = \mathbf{0}_{NT\times 1}$ and $\mathrm{var}(\mathbf{u}|\,\mathbf{Z}) = \sigma^2\mathbf{I}_{NT}$ for some constant $\sigma^2 > 0$. Under these assumptions, the best (in the sense of variance minimization) linear unbiased estimator of $\boldsymbol{\gamma}$

is the (pooled) OLS estimator

$$\widehat{\boldsymbol{\gamma}} := \begin{bmatrix} \widehat{\boldsymbol{\alpha}} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix} := \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}. \tag{1.2}$$

This regression is sometimes referred to as *the long regression* because it includes the individual indicators $\mathbf{D}$ in addition to $\mathbf{X}$ and estimates the α_i's and $\boldsymbol{\beta}$ simultaneously.

In the 1960s computational power was a true bottleneck. In particular, inversion of an $NT \times NT$ matrix such as $\mathbf{Z}'\mathbf{Z}$ was essentially impossible for even moderately large NT. The purpose of this exercise is to use least-squares and linear-algebra manipulations to bypass some of these computational issues.

The starting point of our analysis is the matrix

$$\mathbf{M} := \mathbf{I}_{NT} - \mathbf{D}\left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'.$$

When multiplied with an $NT \times 1$ vector $\mathbf{v}$, $\mathbf{Mv}$ delivers the residuals from regressing the elements of $\mathbf{v}$ on the rows of $\mathbf{D}$. The matrix $\mathbf{M}$ is therefore sometimes referred to as a *residual maker*.

(a) Show that $\mathbf{MD} = \mathbf{0}_{NT\times N}$ and that $\mathbf{M}$ is both *symmetric* ($\mathbf{M}' = \mathbf{M}$) and *idempotent* ($\mathbf{MM} = \mathbf{M}$).[1]

(b) Let $\widehat{\mathbf{u}} := \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\gamma}}$ be the residuals from the long regression. Show that $\mathbf{M}\widehat{\mathbf{u}} = \widehat{\mathbf{u}}$. [*Hint:* Inspect the first-order condition (FOC) for a minimum.]

(c) Show that $\widehat{\boldsymbol{\beta}}$ may be written as

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{M}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}\mathbf{y}. \tag{1.3}$$

This is sometimes referred to as *the short regression*. Are there any potential caveats from implementing this expression for $\widehat{\boldsymbol{\beta}}$ computationally? Justify your answer.

(d) Let "dots" denote a within-transformed variable (deviation from i-specific means), such

[1] A few reminders about matrix algebra: for conformable matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, we have $(\mathbf{A}')' = \mathbf{A}$, $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, $(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$, $\det(\mathbf{A}') = \det(\mathbf{A})$, and $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$. For a constant $c \in \mathbb{R}$, $(c\mathbf{A})' = c\mathbf{A}'$.

as

$$\ddot{y}_{it} := y_{it} - \overline{y}_i, \quad \overline{y}_i := \frac{1}{T}\sum_{t=1}^{T} y_{it},$$
$$\text{and} \quad \ddot{\mathbf{x}}_{it} := \mathbf{x}_{it} - \overline{\mathbf{x}}_i, \quad \overline{\mathbf{x}}_i := \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it},$$

Show that $\mathbf{My} = \ddot{\mathbf{y}}$ and $\mathbf{MX} = \ddot{\mathbf{X}}$, where $\ddot{\mathbf{y}}$ and $\ddot{\mathbf{X}}$ arise from stacking the $\ddot{y}_{it}$s and $\ddot{\mathbf{x}}_{it}$'s, respectively, over first t and then i. [*Hint:* Calculate $\mathbf{D}'\mathbf{D}$ and $\mathbf{D}'\mathbf{y}$.]

(e) Show that $\widehat{\boldsymbol{\beta}}$ obtained from the short regression is actually *the within estimator*. That is, this estimator may be obtained from the pooled OLS regression of $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$.

(f) Show that each $\widehat{\alpha}_i$ in (1.2) may be expressed as $\widehat{\alpha}_i = \overline{y}_i - \overline{\mathbf{x}}_i'\widehat{\boldsymbol{\beta}}$. [*Hint:* FOC.]

(g) Show that the residuals $\widehat{\mathbf{u}}$ from the pooled OLS regression of y_{it} on $\mathbf{z}_{it}$ are the same as the residuals $\widehat{\ddot{\mathbf{u}}}$ from the pooled OLS regression of $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$. [*Hint:* Use (f).]

(h) Show that the conditional variance $\text{var}(\widehat{\boldsymbol{\gamma}}\,|\mathbf{Z}) = \sigma^2\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$. [*Hint:* $\text{var}\left(\mathbf{Av}\right) = \mathbf{A}\text{var}\left(\mathbf{v}\right)\mathbf{A}'$ for a linear transformation $\mathbf{Av}$ of a random vector $\mathbf{v}$.]

(i) The lower-right $K \times K$ block of the matrix $\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$ is given by

$$\left(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{D}\left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{X}\right)^{-1}.$$

(Trust us here.) Show that $\text{var}(\widehat{\boldsymbol{\beta}}\,|\mathbf{Z}) = \sigma^2(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}$.

(j) Essentially any commercial software package asked to conduct a pooled OLS regression of $\ddot{y}_{it}$ on $\ddot{\mathbf{x}}_{it}$ will automatically calculate and provide default standard errors and confidence intervals for each $\widehat{\beta}_k$ in the usual way. How must these standard errors be adjusted in order to account for the fact that we have transformed the y_{it}'s and $\mathbf{x}_{it}$'s? Does the adjustment matter? Justify your answer. [*Hint:* Use (g) and (i).]

2 Production Functions

In many macroeconomic models, firms are assumed to have constant returns to scale in their production function, F, which we measure by total sales in the data. If the inputs in production are capital, K, and labor, L, then constant returns to scale means that

$F(\lambda K, tL) = \lambda F(K, L)$ for all $\lambda \geq 0$. In this assignment, we assume that the production technology of French manufacturing firms is Cobb-Douglas,

$$F(K, L) = AK^{\alpha}L^{\beta},$$

where A is typically thought of as "total factor productivity" (TFP), and α and β are parameters of interest.

In order to empirically investigate the production technology of firms, you are provided with a panel dataset on French manufacturing firms. Specifically, the dataset covers $N = 441$ firms over the years 1967 to 1979. The three accompanying files, `ldsa.dat`, `lemp.dat` and `lcap.dat`, contain respectively the variables `LDSA` (log of deflated sales), `LEMP` (log of employment), and `LCAP` (log of adjusted capital stock), respectively. Each of the variables have had their cross-sectional means subtracted, but we will ignore the econometric issues that this imposes throughout. To read in the data, use Matlab's `load` function. Each of the three datasets are in wide form ($N \times T$).

(1) Formulate a linear panel data model for firm sales, and estimate the parameters α and β using different relevant estimators from the course. When comparing estimators, be explicit about the assumptions each requires for consistency or efficiency. Discuss these assumptions in the empirical context and give an example of firm behavior that satisfies or invalidates these assumptions.

(2) Use the French data to test the assumption that production exhibits constant returns to scale. Discuss the assumptions required for your test to be valid.

(3) Criticize the validity of your conclusion in (2), focusing on the issue you find most critical about the whole empirical exercise. Organize the discussion as follows:

   (a) describe the real-world behavior underlying or causing the issue, including why you find it so critical,

   (b) formally explain the econometric problem this causes, and

   (c) give a suggestion for how the analysis could be examined or even improved (e.g. with more variables, a different model, exploiting a policy reform, etc.).

2.1 Hints

(a) Remember to define all mathematical variables and distinguish between, e.g., the true value and an estimate hereof. Regarding matrices: specify the dimensions if they are not entirely clear.

(b) When using an estimator, carefully discuss which assumptions are required to derive the estimator and discuss whether these assumptions are satisfied empirically. If not, what are the consequences for the estimator in question?

(c) If you have used several estimators, discuss which estimator appears more suitable for the current data and application.

(d) Be precise about the statistical tests you use for testing various hypotheses. Explain which null hypothesis you are testing, what decision rule you are employing, and what your conclusions are. If a variance matrix has been estimated, discuss the assumptions it relies on. If you have several, compare their implications and pick the best one for the empirical context.

Formal Requirements

- You must hand in a report containing your answers to all parts of the assignment.
- The report must be written in English and uploaded to Peergrade via Absalon as one single PDF file.
- You must obey the following page constraints:
  - Problem 1: No constraints,
  - Problem 2: Maximum four normal pages of text plus two pages of output.[2]
- You are allowed to work in groups of up to three people (not necessarily in the same exercise class as yours). List all group members on the front page of your report.
- The assessment criteria are posted on the course page on Absalon.

References

MUNDLAK, Y. (1961): "Empirical production function free of management bias," *Journal of Farm Economics*, 43, 44–56.

[2]One normal page of text is defined as: Font size = 12p, 1.5 line spacing, and margins of 2.5 cm.