Fixed Effects and First Differences

AME Week 2

September 2022

1 Introduction

The purpose of this and next week's exercises is to estimate two basic linear panel data models with unobserved effects. The two models make different assumptions about the correlation between observed and unobserved components and it it is important to understand which set of assumptions are the most appropriate in empirical applications. Next week's exercise goes through an econometric test procedure (the Hausman test) that tests the assumptions of the two models against each other. This week's exercise starts out by estimating the unobserved effects model allowing for arbitrary contemporaneous correlation between the unobserved individual effect and the explanatory variables. We shall use two estimators: The Fixed-Effects (FE) estimator and the First-Difference (FD) estimator.

2 Linear Panel Data Models

Consider the following linear model,

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{1}$$

where $i = 0, ..., N$ indexes the cross sectional unit that is observed (e.g., households), and $t = 0, ..., T$ denotes time (e.g. weeks, years). $\boldsymbol{x}_{it}$ is a $1 \times K$ vector of regressors, $\boldsymbol{\beta}$ contains the $K \times 1$ parameters to be estimated, c_i is an unobserved individual specifc component which is constant across time periods, and u_{it} is an unobserved random error term.

If c_i turns out to be an additional error term uncorrelated with the regressors in the sense of $E[c_i\boldsymbol{x}_{it}] = 0$ for all t, then $\boldsymbol{\beta}$ can be consistently estimated by pooled OLS (POLS) (as $N \to \infty$ for fixed T), albeit in an inefficient manner. To see this, consider the joint error termt $v_{it} = c_i + u_{it}$, and note that,

$$E[v_{it}\boldsymbol{x}_{it}] = E[c_i\boldsymbol{x}_{it}] + E[u_{it}\boldsymbol{x}_{it}] = \mathbf{0}, \tag{2}$$

so that the usual OLS assumptions are satisfied. Conversely, if c_i is systematically related to one or more of the observed variables in the sense of $E[c_i\boldsymbol{x}_{it}] \neq \mathbf{0}$, then the POLS estimator is **not** consistent for $\boldsymbol{\beta}$.

2.1 Example

When might $E[c_i \boldsymbol{x}_{it}] \neq \mathbf{0}$? Consider a model designed to investigate if union membership affects wages. The model explains wages as a function of experience and their union membership.

$$\log(\text{wage}_{it}) = \beta_1 \text{exper}_{it} + \beta_2 \text{exper}^2_{it} + \beta_3 \text{union}_{it} + c_i + u_{it}, \tag{3}$$

where c_i is an individual-specific constant that summarizes innate and unobserved characteristics. If people select into union or non-union jobs based on which sector rewards their innate characteristics best, then $E[union_{it} c_i] \neq 0$. For this reason, it doesn't seem reasonable to use OLS on the pooled data.

In this example, the inconsistency of OLS is caused by the presence of c_i. The conventional approach to deal with this problem in linear panel data models is to transform equation (1) such that c_i vanishes, and the transformed model allows $\boldsymbol{\beta}$ to be estimated by OLS. Because the model is linear, we may rid ourselves of c_i using relatively simple, linear, transformations. In the following, we shall consider two such transformations: i) the **within-groups** transformation, and ii) the **first-difference** transformation.

3 Fixed Effects (FE) and Within-Groups Transformation

The within-groups transformation subtracts from each variable its mean within each cross sectional unit. Consequently, every time-invariant variables disappear when using this transformation. To make the within-groups transformation more explicit, take the average of equation (1) across T for each i to obtain

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta} + c_i + \bar{u}_i, \tag{4}$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$, and $T^{-1} \sum_{t=1}^{T} c_i = c_i$. Subtract equation (4) from equation (1) to get

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \boldsymbol{\beta} + (c_i - c_i) + (u_{it} - \bar{u}_i) \tag{5}$$

$$\Leftrightarrow \ddot{y}_{it} = \ddot{\mathbf{x}}_{it} \boldsymbol{\beta} + \ddot{u}_{it}. \tag{6}$$

This simple manipulation of the empirical model has eliminated c_i by subtracting the mean within each i-group. This is called the *within transformation*, and a within-transformed variable is denoted $\ddot{y}_{it} = y_{it} - \bar{y}_i$. The parameters of interest, $\boldsymbol{\beta}$, can be estimated by OLS on the transformed data, i.e.

$$\hat{\boldsymbol{\beta}}_{FE} = (\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1} \ddot{\mathbf{X}}' \ddot{\mathbf{y}}, \tag{7}$$

where $\ddot{\mathbf{X}}$ is the $NT \times K$ matrix and $\ddot{\mathbf{y}}$ the $NT \times 1$ vector arising from stacking the observables of (6), i.e., $\ddot{\mathbf{x}}'_{it}$ and $\ddot{y}_{it}$, over first t and then i.

3.1 FE Assumptions

Let $\ddot{\mathbf{X}}_i$ denote the $T \times K$ matrix arising from stacking $\ddot{\mathbf{x}}_{it}'$ over t. (We here keep the i subscript to avoid a clash of notation.) We invoke the following assumptions

$$\begin{aligned} FE.1 &: E[u_{it}|\mathbf{x}_{i1},..,\mathbf{x}_{iT},c_i] = 0, \quad \text{for } t = 1,\dots,T, \\ FE.2 &: \text{Rank } E[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i] = K, \quad \text{for } i = 1,\dots,N \\ FE.3 &: E[\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i,c_i] = \sigma_u^2\mathbf{I}_T, \quad \text{for } i = 1,\dots,N. \end{aligned}$$

Under the strict exogeneity assumption ($FE.1$) and the rank condition ($FE.2$), the FE estimator, $\hat{\boldsymbol{\beta}}_{FE}$, consistently estimate $\boldsymbol{\beta}$ as $N \to \infty$ for fixed T. Under FE.3, $\hat{\boldsymbol{\beta}}_{FE}$ is also asymptotically efficient. (But the latter assumption is not needed for consistency.)

In order to perform inference on the obtained parameter estimates, we need standard errors of the parameter estimates. If the unobservables $\{u_{it}\}_{t=1}^T$ of (1) satisfy $FE.3$, then the variance-covariance matrix of the FE estimator may be estimated by

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{FE}) = \hat{\sigma}_u^2(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}, \tag{8}$$

where $\hat{\sigma}_u^2 := \widehat{\mathbf{u}}'\widehat{\mathbf{u}}/[N\,(T-1)-K]$ and $\widehat{\mathbf{u}} := \ddot{\mathbf{y}} - \ddot{\mathbf{x}}\boldsymbol{\beta}$ so that $\widehat{\mathbf{u}}'\widehat{\mathbf{u}} = \sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2$.

4 Transforming data using the ‘perm‘ function

The main challenge in implementing (3) in Python lies in de-meaning the variables, i.e., constructing $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and $\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$. On the *individual level*, this can be done by pre-multiplying equation (1) by a transformation matrix

$$\mathbf{Q}_T := \mathbf{I}_T - \begin{pmatrix} 1/T & \dots & 1/T \\ \vdots & \ddots & \vdots \\ 1/T & \dots & 1/T \end{pmatrix}_{T\times T}. \tag{9}$$

However, even though $\mathbf{Q}_T\mathbf{y}_i = \ddot{\mathbf{y}}_i$, we can’t simply multiply the full data vector, $\mathbf{y} = (\mathbf{y}_1,\dots,\mathbf{y}_N)'$, with $\mathbf{Q}_T$ since it needs to be done for each individual. Towards this end, the Python function `perm`$(P, \mathbf{x})$ picks out the elements of the input-vector (here $\mathbf{x}$) and premultiplies with the input-matrix $\mathbf{P}$ for one individual at the time (using a *for* loop). For example,

$$\texttt{perm}\left(\mathbf{Q}_T, \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}\right) = \begin{pmatrix} \mathbf{Q}_T\mathbf{y}_1 \\ \vdots \\ \mathbf{Q}_T\mathbf{y}_N \end{pmatrix} = \begin{pmatrix} \ddot{\mathbf{y}}_1 \\ \vdots \\ \ddot{\mathbf{y}}_N \end{pmatrix} = \ddot{\mathbf{y}}. \tag{10}$$

The same goes for $\mathbf{x}$-input. (You may want to take a look under the hood of this function.)

5 First-difference estimation (FD)

The within transformation is one particular transformation that enables us to get rid of c_i. An alternative is the first-difference transformation. To see how it works, lag equation (1) one period and subtract it from (1) such that

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad t = 2, \ldots, T, \tag{11}$$

where $\Delta y_{it} := y_{it} - y_{it-1}$, $\Delta \mathbf{x}_{it} := \mathbf{x}_{it} - \mathbf{x}_{it-1}$ and $\Delta u_{it} := u_{it} - u_{it-1}$. As was the case for the within transformation, first differencing eliminates the time invariant component c_i. Note, however, that one time period is lost when differencing.

5.1 FD Assumptions

$$\begin{aligned}
FD.1 &: \quad E[u_{it}|\mathbf{x}_{i1}, .., \mathbf{x}_{iT}, c_i] = 0 \quad t = 1, \ldots, T, \ \text{(as in } FE.1\text{)} \\
FD.2 &: \quad \text{Rank } E[\Delta \mathbf{X}_i' \Delta \mathbf{X}_i] = K, \quad \text{for } i = 1, \ldots, N, \\
FD.3 &: \quad E[\mathbf{e}_i \mathbf{e}_i' | \mathbf{x}_i, c_i] = \sigma_e^2 \mathbf{I}_{T-1} \ \text{ with } \mathbf{e}_i := \Delta \mathbf{u}_i, \quad \text{for } i = 1, \ldots, N.
\end{aligned}$$

Under the strict exogeneity assumption ($FD.1$) and the rank condition ($FD.2$), the FD estimator

$$\hat{\boldsymbol{\beta}}_{FD} = \left(\Delta \mathbf{X}' \Delta \mathbf{X}\right)^{-1} \Delta \mathbf{X}' \Delta \mathbf{y} \tag{12}$$

consistently estimates $\boldsymbol{\beta}$ (as $N \to \infty$ for fixed T). If also FD.3 holds, then $\hat{\boldsymbol{\beta}}_{FD}$ is asymptotically efficient. (Again, the latter assumption is not needed for consistency).

Under $FD.3$, $u_{it} = u_{it-1} + e_{it}$ follows a random walk. This is the opposite extreme relative to assumption $FE.3$, where the u_{it} are assumed to be serially uncorrelated. In many cases, the truth is likely to lie somewhere in between. The variance-covariance matrix of the FE estimator may be estimated by

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{FD}) = \hat{\sigma}_e^2 \left(\Delta \mathbf{X}' \Delta \mathbf{X}\right)^{-1} \tag{13}$$

where $\hat{\sigma}_e^2 := \hat{\mathbf{e}}' \hat{\mathbf{e}} / [N(T-1) - K]$ and $\hat{e}_{it} := \Delta y_{it} - \Delta \mathbf{x}_{it} \widehat{\boldsymbol{\beta}}$.

Notice how we, both in the case of FE and FD, manipulate the model in a way that allows the standard OLS assumptions to hold on the *transformed* data, and then simply treat the transformed model as if it was our model of interest. Under exogeneity ($FE.1/FD.1$) the choice between first difference and the within estimator pertains to efficiency considerations, and the choice hinges on the assumptions made about the serial correlation of the errors ($FE.3/FD.3$).

To estimate the coefficients in (5) in Python, we must first difference all the variables, i.e construct $\Delta y_{it} = y_{it} - y_{it-1}$ and $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$. This can be done by premultiplying the variables in levels (y_i and $\mathbf{x}_i$) by the transformation matrix $\mathbf{D}$ given by

$$\mathbf{D} := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}_{T-1\times T} . \tag{14}$$

(Can you see why $\mathbf{D}$ gets the job done?)