1 Mundlak approach

Let's consider a simple example

$$y_{it} = \mathbf{x}_{it}\beta + \alpha_i + \epsilon_{it},$$

$$\alpha_i = \bar{\mathbf{x}}_i\gamma + v_i,$$

$$\mathbb{E}(\alpha_i|\mathbf{x}_i) = \bar{\mathbf{x}}_i\gamma,$$

where $i = 1, \dots, N$, $t = 1, \dots, T$, $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}$ and v_i is a time-invariant unobservable that is uncorrelated to the regressors. The key to the Mundlak approach is to determine if α_i and $\mathbf{x}_{it}$ are correlated. We can substitute $\alpha_i = \bar{\mathbf{x}}_i\gamma + v_i$ into $y_{it} = \mathbf{x}_{it}\beta + \alpha_i + \epsilon_{it}$ which gives us

$$y_{it} = \mathbf{x}_{it}\beta + \bar{\mathbf{x}}_i\gamma + v_i + \epsilon_{it},$$

Thus, our Wald test is

$$H_0 : \gamma = 0,$$

$$H_A : \gamma \neq 0,$$

and Wald statistic is given by $W = \frac{\hat{\gamma}}{var(\hat{\gamma})}$ which is also same as the Huasman statistics.

1. Reject the null hypothesis: This suggests there is statistical evidence that there is correlation between the time-invariant unobservables and your regressors, namely, the fixed-effects assumptions are satisfied.

2. Do not reject the null hyopthesis: This suggest that the generated regressors are zero, there is evidence of no correlation between the time-invariant unobservable and your regressors; that is, the random effects assumptions are satisfied.

Note remember for a random effect model the $cov(a_i\mathbf{x}_{it}) = 0$ has to hold to ensure consistency for the random effector estimator $\hat{\beta}_{RE}$ and this lead to $var(\hat{\beta}_{RE}) < var(\hat{\beta}_{FE})$. Thus, random effector estimator is a more consistent estimator than the fixed effect estimator. However, when $cov(a_i\mathbf{x}_{it}) \neq 0$, then $\hat{\beta}_{FE}$ is the sole consistent estimator out of the two estimators.

2 Concentrated likelihood

A key difficulty in typical maximization problems is the high dimensionality of θ. One can reduce the number of dimensions if a subset of the equations can be solved and if the solution can be substituted back into the likelihood function. Let the parametric vector be partitioned into $\theta = (\theta_1', \theta_2')'$. Suppose that given any θ_2, one can find the optimal value of θ_1 as a function of θ_2 by solving the first-order conditions

$$\frac{\partial}{\partial \theta_1} L(\theta|\mathbf{y}) = 0 \Longrightarrow \hat{\theta}_1 = \hat{\theta}_1(\theta_2),$$

Substituting this function into the original log-likelihood yields the **concentrated loglikelihood function**

$$L_c(\theta_2|\mathbf{y}) = L_c(\hat{\theta}_1(\theta_2), \theta_2|\mathbf{y}),$$

Since the concentrated log-likelihood $L_c(\theta_2|\mathbf{y})$ is a function of θ_2 only, the dimension of the maximization problem is reduced. Once the MLE for θ_2 is obtained, say, by numerical methods, we can then compute the MLE for $\hat{\theta}_1$ analytically via $\hat{\theta}_1(\theta_2)$. We can use a concentrate likelihood function via the OLS, for instance, recall from last week's tutorial MLE for OLS is

$$\hat{\theta}_{MLE} = \underset{\theta}{\text{argmax}} ln(L(\beta, \sigma^2|\mathbf{y})),$$

where $\theta = (\beta, \sigma^2)$ and $ln(L(\beta, \sigma^2|\mathbf{y})) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta)$. We can first take the FOC with respect to β

$$\frac{\partial ln(L(\beta, \sigma^2|\mathbf{y}))}{\partial \beta} = -\frac{1}{2\sigma^2}(2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'\mathbf{y}) = 0 \Rightarrow \hat{\beta}_{MLE} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

Then the concentrated likelihood function is derived by subsituting $\hat{\beta}_{MLE}$ back into the log-likelihood function

$$ln(L_c(\hat{\beta}_{MLE}, \sigma^2|\mathbf{y})) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\hat{\beta}_{MLE}'\mathbf{X}'\mathbf{y} + \hat{\beta}_{MLE}'\mathbf{X}'\mathbf{X}\hat{\beta}_{MLE}),$$

then we take the FOC of the above concentrated likelihood function with respect to σ^2

$$\frac{\partial ln(L_c(\hat{\beta}_{MLE}, \sigma^2|\mathbf{y}))}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{y} - \mathbf{X}\hat{\beta}_{MLE})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{MLE}) = 0 \Longrightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta}_{MLE})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{MLE}),$$

3 Fixed effect model

Consider a panel data model

$$y_{it} = \mathbf{x}_{it}\beta + \alpha_i + u_{it}, \tag{1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The first fixed effect (FE) assumption is strict exogeneity of the explanatory variables conditional on α_i which is $\mathbb{E}(u_{it}|\mathbf{x}_i, \alpha_i) = 0$. We also should mention the sequential moment restrictions:

$$\mathbb{E}(u_{it}|\mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}, \alpha_i) = 0,$$

$$\mathbb{E}(y_{it}|\mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}, \alpha_i) = \mathbb{E}(y_{it}|\mathbf{x}_{it}, a_i) = \mathbf{x}_{it}\beta + \alpha_i,$$

What does this restriction mean? After $\mathbf{x}_{it}$ and α_i have been controlled for, no past value of $\mathbf{x}_{it}$ affect the expected value of y_{it}. This condition is more natural than the strict exogeneity assumption, which requires conditioning on future values of $\mathbf{x}_{it}$ as well.

To derive the FE estimator, we first need to define (1) as averages:

$$\bar{y}_i = \bar{\mathbf{x}}_i\beta + \alpha_i + \bar{u}_i, \tag{2}$$

where $\bar{y}_{it} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$ and $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$. Next, we substract (2) from (1) we get

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta + (\alpha_i - \alpha_i) + u_{it} - \bar{u}_i,$$

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\beta + \tilde{u}_{it}, \tag{3}$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$ and $\tilde{u}_{it} = u_{it} - \bar{u}_i$. Similar to OLS, the FE estimator is

$$\hat{\beta}_{FE} = (\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}\tilde{y}_{it}), \tag{4}$$

To get it in sample error form, we substitute (3) from (4)

$$\hat{\beta}_{FE} = (\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}(\tilde{\mathbf{x}}_{it}\beta + \tilde{u}_{it})),$$

$$\hat{\beta}_{FE} = (\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it}\beta + (\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}\tilde{u}_{it}),$$

Since we know $\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it} = \mathbf{I}$, then

$$\hat{\beta}_{FE} = \beta + (\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}\tilde{u}_{it}),$$

Note from the strict exogeneity assumption of $\mathbb{E}(u_{it}|\mathbf{x}_i, \alpha_i) = 0$, it follows that u_{it} and $\bar{u}_i$ are uncorrelated with $\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_i$ for $t = 1, \dots, T$. Thus, this imples $\tilde{\mathbf{x}}_{it}^{'}\tilde{u}_{it} = \tilde{\mathbf{x}}_{it}^{'}u_{it}$.

$$\hat{\beta}_{FE} - \beta = (\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}u_{it}),$$

Then we want to find

$$\underset{N\longrightarrow\infty}{plim}\,(\hat{\beta}_{FE} - \beta) = \underset{N\longrightarrow\infty}{plim}\,((\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}u_{it})),$$

We want to get the above term is 'sample moment form'

$$\underset{N\longrightarrow\infty}{plim}\,(\hat{\beta}_{FE} - \beta) = \underset{N\longrightarrow\infty}{plim}\,((\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{\mathbf{x}}_{it}^{'}u_{it})),$$

Then applying the weak law of large numbers and continous mapping theorem, we get

$$(\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1} \underset{p}{\longrightarrow} \mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it})^{-1},$$

$$(\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{x}}_{it}^{'}u_{it}) \underset{p}{\longrightarrow} \mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}u_{it}),$$

Then applying the Slutsky theorem,

$$\underset{N\longrightarrow\infty}{plim}\,(\hat{\beta}_{FE} - \beta) = \underset{N\longrightarrow\infty}{plim}\,(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}\tilde{\mathbf{x}}_{it}))^{-1} \times \underset{N\longrightarrow\infty}{plim}\,(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}u_{it})),$$

If the strict exogeneity assumption of $\mathbb{E}(u_{it}|\mathbf{x}_i, \alpha_i) = 0$ holds, then $\mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}u_{it}) = 0$ and the above term becomes zero

$$\underset{N\longrightarrow\infty}{plim}\,(\hat{\beta}_{FE} - \beta) = 0.$$

If the strict exogeneity assumption does not hold, then under the sequential moment restriction, $\mathbb{E}(\tilde{\mathbf{x}}_{it}^{'}u_{it}) =$

$\mathbb{E}((\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' u_{it}) = -\mathbb{E}(\bar{\mathbf{x}}_i' u_{it})$ because $\mathbb{E}(\mathbf{x}_{it} u_{it}) = 0$ and so $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\tilde{\mathbf{x}}_{it}' u_{it}) = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\bar{\mathbf{x}}_i' u_{it}) = -\mathbb{E}(\bar{\mathbf{x}}_i' \bar{u}_i)$. If $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\tilde{\mathbf{x}}_{it}'\tilde{\mathbf{x}}_{it}))^{-1}$ is bounded, $var(\bar{\mathbf{x}}_i)$ and $var(\bar{u}_i)$ are order of $\frac{1}{T}$, and the time-series is weakly dependent, then the inconsistency from using fixed effects when the strict exogeneity assumption fails is of order $\frac{1}{T}$.

4 First difference model

We can take the first difference of (1) across time T,

$$y_{i2} - y_{i1} = (\mathbf{x}_{i2} - \mathbf{x}_{i1})\beta + u_{i2} - u_{i1},$$
$$\vdots$$
$$y_{iT} - y_{iT-1} = (\mathbf{x}_{iT} - \mathbf{x}_{iT-1})\beta + u_{iT} - u_{iT-1},$$

and we can define it as

$$\triangle y_{it} = \triangle \mathbf{x}_{it}\beta + \triangle u_{it}, \tag{5}$$

where $\triangle y_{it} = y_{it} - y_{it-1}$, $\triangle \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$ and $\triangle u_{it} = u_{it} - u_{it-1}$. Similar to the FE estimator, the FD estimator will be

$$\hat{\beta}_{FD} = (\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}(\triangle \mathbf{x}_{it}^{'}\triangle y_{it}), \tag{6}$$

Then to get it in sampling error form, we substitute (5) from (6)

$$\hat{\beta}_{FD} = (\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}(\triangle \mathbf{x}_{it}^{'}(\triangle \mathbf{x}_{it}\beta + \triangle u_{it})),$$

$$\hat{\beta}_{FD} = \sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it}\beta + \sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle u_{it},$$

Since $\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it} = \mathbf{I}$, then

$$\hat{\beta}_{FD} = \beta + \sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle u_{it},$$

Then we want to find

$$\underset{N \longrightarrow \infty}{plim}\,(\hat{\beta}_{FD} - \beta) = \underset{N \longrightarrow \infty}{plim}\,(\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle \mathbf{x}_{it})^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\triangle \mathbf{x}_{it}^{'}\triangle u_{it}),$$

and we first want to get it in sample moment form

$$\underset{N \longrightarrow \infty}{plim}\ (\hat{\beta}_{FD} - \beta) = \underset{N \longrightarrow \infty}{plim}\ (\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \triangle \mathbf{x}_{it}^{'} \triangle \mathbf{x}_{it})^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \triangle \mathbf{x}_{it}^{'} \triangle u_{it}),$$

Then applying the weak law of large numbers and continous mapping theorem, we get

$$(\frac{1}{N} \sum_{i=1}^{N} \triangle \mathbf{x}_{it}^{'} \triangle \mathbf{x}_{it})^{-1} \underset{p}{\longrightarrow} \mathbb{E}(\triangle \mathbf{x}_{it}^{'} \triangle \mathbf{x}_{it})^{-1},$$

$$(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{it}^{'} \triangle u_{it}) \underset{p}{\longrightarrow} \mathbb{E}(\mathbf{x}_{it}^{'} \triangle u_{it}),$$

Then applying the Slutsky theorem,

$$\underset{N \longrightarrow \infty}{plim}\ (\hat{\beta}_{FD} - \beta) = \underset{N \longrightarrow \infty}{plim}\ (\frac{1}{T} \sum_{t=2}^{T} \mathbb{E}(\triangle \mathbf{x}_{it}^{'} \triangle \mathbf{x}_{it}))^{-1} \times \underset{N \longrightarrow \infty}{plim}\ (\frac{1}{T} \sum_{t=2}^{T} \mathbb{E}(\triangle \mathbf{x}_{it}^{'} \triangle u_{it})),$$

The FD estimator is consistent if $\mathbb{E}(\triangle \mathbf{x}_{it}^{'} \triangle u_{it}) = 0$ then

$$\underset{N \longrightarrow \infty}{plim}\ (\hat{\beta}_{FD} - \beta) = 0,$$

Recall the sequential moment restriction, which is

$$\mathbb{E}(u_{it} | \mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}, \alpha_i) = 0,$$

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}, \alpha_i) = \mathbb{E}(y_{it} | \mathbf{x}_{it}, a_i) = \mathbf{x}_{it}\beta + \alpha_i,$$

What does this restriction mean? After $\mathbf{x}_{it}$ and α_i have been controlled for, no past value of $\mathbf{x}_{it}$ affect the expected value of y_{it}. This condition is more natural than the strict exogeneity assumption, which requires conditioning on future values of $\mathbf{x}_{it}$ as well. If the FD estimator is inconsistent, then the sequential moment restriction implies

$$\mathbb{E}(\triangle \mathbf{x}_{it}^{'} \triangle u_{it}) = \mathbb{E}(\mathbf{x}_{it}^{'} u_{it}) - \mathbb{E}(\mathbf{x}_{it-1}^{'} u_{it-1}) - \mathbb{E}(\mathbf{x}_{it-1}^{'} u_{it}) - \mathbb{E}(\mathbf{x}_{it}^{'} u_{it-1}) = -\mathbb{E}(\mathbf{x}_{it}^{'} u_{it-1}) \neq 0,$$

Under stationarity condition, $\mathbb{E}(\mathbf{x}_{it}^{'} u_{it-1})$ does not depend on t and therefore the FD estimator inconsistency does not depend on T. Therefore, if our choice were between FE and FD, we would tend to prefer FE because, when $T > 2$, FE can have less bias as $N \longrightarrow \infty$.