Graduate Econometrics
Lecture 8: Course Revision

Yonas Alem

Department of Economics
University of Gothenburg

January 11, 2015

- One is interested in the best linear combination of $x_2, ..., x_K$ and a constant gives a good approximation of the dependent variable y
- Write an arbitrary linear combination:

$$\tilde{\beta}_1 + \tilde{\beta}_2 x_2 + ... + \tilde{\beta}_K x_K, \tag{1}$$

- Where $\tilde{\beta}_1, ..., \tilde{\beta}_K$ are constants to be chosen
- Index the observations by $i, i = 1, ..., N$
- The difference between the observed value of y and its linear approximation is given by

$$y_i - [\tilde{\beta}_1 + \tilde{\beta}_2 x_{i2} + ... + \tilde{\beta}_K x_{iK}]. \tag{2}$$

- Re-write [2] as:

$$y_i - x_i'\tilde{\beta}. \tag{3}$$

- The objective is to choose values for $\tilde{\beta}_1, ..., \tilde{\beta}_K$ such that these differences are small

- OLS: choose $\tilde{\beta}$ such that the sum of squared differences is as small as possible. i.e

$$S(\tilde{\beta}) \equiv \sum_{i=1}^{N}(y_i - x_i'\tilde{\beta})^2 \tag{4}$$

- FOC: Differentiate $S(\tilde{\beta})$ w.r.t the vector $\tilde{\beta}$, which gives the following K conditions (**Normal Equations**)

$$-2\sum_{i=1}^{N} x_i(y_i - x_i'\tilde{\beta}) = 0 \tag{5}$$

or

$$(\sum_{i=1}^{N} x_i x_i')\tilde{\beta} = \sum_{i=1}^{N} x_i y_i. \tag{6}$$

- The solution to the minimization problem, for b is given by

$$b = (\sum_{i=1}^{N} x_i x_i')^{-1} \sum_{i=1}^{N} x_i y_i \tag{7}$$

- One can confirm from the SOCs that b indeed corresponds to a minimum Thus,

$$\hat{y}_i = x_i' b \tag{8}$$

- $[A1]: E\{\varepsilon_i\} = 0, \quad i = 1, ..., N$
- $[A2]: E\{x_i \varepsilon_i\} = 0$
- $[A3]: V\{\varepsilon_i\} = \sigma^2, \quad i = 1, ..., N$
- $[A4]: Cov\{\varepsilon_i, \varepsilon_j\} = 0, \quad i \neq j$
- For exact statistical inference from a given sample, of N observations one needs the additional distributional assumption (normality), i.e.,

$$\varepsilon_i \sim NID(0, \sigma^2) \qquad [A5] \tag{9}$$

- $[A1] - [A5] \Longrightarrow b \sim N(\beta, \sigma^2 (X'X)^{-1})$ and $b_k \sim N(\beta_k, \sigma^2 c_{kk})$
- Where c_{kk} is the (k, k) element in $(X'X)^{-1}$
- Under the Gauss-Markov conditions, OLS is **BLUE**!

The OLS Estimator
Summary Statistics

- Three commonly used summary statistics in multiple regression:
  1. The standard error of the regression
  2. The regression R^2, and;
  3. The adjusted R^2 denoted $\bar{R}^2$
- All three statistics measure how well the OLS estimate of the multiple regression line describes, or "fits", the data

The *t-test*

- Let $H_0 : \beta_k = \beta_k^0$ where β_k^0 is a specific value chosen by the researcher
- If this is true, the **statistic:**

$$t_k = \frac{b_k - \beta_k^0}{se(b_k)} \tag{10}$$

- Has a t distribution with $N - K$ degrees of freedom. If H_0 is not true, the alternative, $H_1 : \beta_k \neq \beta_k^0$ holds
- One rejects H_0 if the probability of observing a value of $|t_k|$ or larger is smaller than a given **significance level** α, often 5% $\Longrightarrow$ the critical values $t_{N-K;\alpha/2}$:

$$P\{|t_k| > t_{N-K;\alpha/2}\} = \alpha \tag{11}$$

The *t-test* cont.

- The **two-tailed** critical value for $\alpha = 0.05$ is $1.96 \implies$ at the 5% level, H_0 will be rejected if

$$|t_k| > 1.96 \quad (12)$$

A simple t-Test - Confidence Interval

- A **confidence interval**: the interval of all values for β_k^0 for which $H_0 : \beta_k = \beta_k^0$ is not rejected by the t-tests.

$$-t_{N-K;\alpha/2} < \frac{b_k - \beta_k}{se(b_k)} < t_{N-K;\alpha/2}, \tag{13}$$

or

$$b_k - t_{N-K;\alpha/2} se(b_k) < \beta_k < b_k + t_{N-K;\alpha/2} se(b_k), \tag{14}$$

- Using the standard normal approximation, a 95% confidence interval (setting $\alpha = 0.05$ for β_k is given by the interval

$$[b_k - 1.96 se(b_k), b_k + 1.96 se(b_k)] \tag{15}$$

- In repeated sampling, 95% of these intervals will contain the true value β_k

Joint Test of Significance: $F-$Test

- Tests whether the increase in R^2 moving from a restricted model to the more general model is significant:

$$F = \frac{(R_1^2 - R_0^2 / J}{(1 - R_1^2)/(N - K)} \tag{16}$$

- Where R_1^2 and R_0^2 refer to those of the unrestricted and the restricted models respectively
- A special case of this $F - test$: standard test automatically supplied by regression a package
- Tests the null hypothesis that the partial slope coefficients are equal to zero, i.e, tests $H_0 : \beta_2 = \beta_3 = ... = \beta_K = 0$. The test statistic is given by

$$F = \frac{(S_0 - S_1)/(K - 1)}{S_1/(N - K)} \tag{17}$$

Joint Test of Significance: $F-$Test cont.

- By construction, R^2 of $S_0 = 0 \Longrightarrow$

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)} \tag{18}$$

- Has an F distribution with $(K-1)\&(N-K)$ degrees of freedom, denoted as F^{K-1}_{N-K}

- **Multicollinearity:** when an approximate linear relationship among the explanatory variables leads to unreliable regression estimates
- **Outliers**: an outlier is an observation that deviates markedly from the rest of the sample
- **Missing Observations:** when some information is missing for a part or all of the sample

- Testing misspecification
- Selecting regressors
  - Should be done based on economic theory
- Akaike's Information Criterion (AIC) & Schwarz Bayesian Information Criterion (BIC)

$$AIC = log\frac{1}{N}\sum_{i=1}^{N} e_i^2 + \frac{2K}{N} \tag{19}$$

$$BIC = log\frac{1}{N}\sum_{i=1}^{N} e_i^2 + \frac{K}{N} logN \tag{20}$$

- Used when:
  - Alternative models are not nested
  - Economic theory provides no guidance on selecting the right model

- We saw that OLS is BLUE when assumptions [A1]-[A4] are fulfilled
- Heteroskedasticity and Autocorrelation $\Longrightarrow$ [A3] & [A4] no longer hold $\Longrightarrow$ OLS is not efficient and the standard errors will be misleading in hypothesis testing
- Heteroskedasticity $\Longrightarrow$ different error terms do not have identical variances $\Longrightarrow$ the diagonal elements of the covariance matrix are not the same
  - Different groups in the sample (e.g. males and females) have different variances
  - The variation of unexplained household savings increases with income (just as savings itself)
- Autocorrelation arises in cases where the data has time dimension
  - The covariance matrix is nondiagonal such that different error terms are correlated.

- In order to claim that the standard errors generated from OLS are misleading due to heteroskedasticity, we need to perform a standard test
- A number of tests are available
- If these tests do not reject $H_0 :=$ heteroskedasticity, there is no problem in using OLS
- If we however reject $H_0 :=$, we should find a solution
  - Use EGLS
  - Use Heteroskedasticity-consistent standard errors for the OLS estimator
  - Revise the specification of our model

Multiplicative Heteroskedasticity

- The alternative hypothesis for this test is:

$$\sigma_i^2 = \sigma^2 exp\{z_i'\alpha\} \tag{21}$$

where z_i is a $J-$dimensional vector. The null hypothesis of homoskedasticity corresponds to $\alpha = 0$, so the problem under test is

$$H_0 = \alpha = 0 \quad versus \quad H_1 : \alpha \neq 0 \tag{22}$$

- H_0 can be tested using the residuals from the least square regression using the standard $F-$test for the hypothesis that all the parameters except the intercept are equal to zero

The Breusch-Pagan (BP) Test

- The simplest version of the BP test can be computed as the number of observations multiplied by the R^2 of an auxiliary regression (a regression of e_i^2, i.e., the OLS residuals) on z_i and a constant
  - The resulting test statistic given by $\xi = NR^2$ is asymptotically χ^2 distributed with J d.f

The White Test

- Obtain NR^2 in the regression of e_i^2 on a constant and all (unique) first moments, second moments, and cross-products of the original regressors
- This test statistic is asymptotically distributed as χ^2 with P d.f, where P is the number of regressors in the auxiliary regression, excluding the intercept.

- So far, in OLS, we assumed independence. $E\{x_i\varepsilon_i\} = 0$.
- In other words, all the explanatory variables are exogenous.
- There are a number of cases in economics where this assumption is unrealistic.
- When a variable is endogenous, the error term will be correlated with the explanatory variable. Thus, OLS is no more unbiased and inconsistent.
- The linear model no longer corresponds to a conditional expectation or a best linear approximation
- Many reasons for contemporaneous correlation between the error term and one or more of the X variables.

1. Introduction of a lagged dependent variable
2. Measurement error in the explanatory variables
3. Omitted variable bias
4. Simultaneity and reverse causality

- Consider the linear wage model

$$y_i = x_{1i}'\beta_1 + x_{2i}\beta_2 + \varepsilon_i \tag{23}$$

- To make the conditional expectation (the best linear approximation) of y_i given x_{1i} and x_{2i}, we needed to impose

$$E\{\varepsilon_i x_{1i}\} = 0 \tag{24}$$

and

$$E\{\varepsilon_i x_{2i}\} = 0 \tag{25}$$

- If not, the model no longer corresponds to $E\{y_i | x_{1i}, x_{2i}\} \Longrightarrow$ OLS will be biased and inconsistent
- In the above wage equation, "ability" or "intelligence" (which is unobserved and hence included in ε_i) would be correlated with "education"

Instrumental Variables (IV) Estimation
Single Endogenous Regressor and Single IV cont.

- In such a case, $E\{\varepsilon_i x_{2i}\} \neq 0$ and we say that x_{2i} is endogenous
- Under additional model identifying assumptions, we would be able to derive another estimator
- The conditions in (24) and (25) are called **moment conditions**
- These conditions would be sufficient to identify the unknown parameters in the model

- The K parameters in β_1 and β_2 should be such that the following K equalities hold:

$$E\{(y_i - x_{1i}'\beta_1 - x_{2i}\beta_2)x_{1i}\} = 0 \tag{26}$$

$$E\{(y_i - x_{1i}'\beta_1 - x_{2i}\beta_2)x_{2i}\} = 0 \tag{27}$$

- These conditions are imposed on the estimator when estimating OLS through the corresponding sample moments
- That is, the OLS estimator $b = (b_1', b_2)'$ for $\beta = (\beta_1', \beta_2)'$ is solved from

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - x_{1i}'b_1 - x_{2i}b_2)x_{1i} = 0 \tag{28}$$

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - x_{1i}'b_1 - x_{2i}b_2)x_{2i} = 0 \tag{29}$$

- These are the first-order conditions for the minimization of the least square criterion and the number of conditions = K (number of unknown parameters)
    - b_1 and b_2 can be solved uniquely from (28) and (29)
- When (25) is violated, (29) drops out and we can no longer solve for b_1 and $b_2 \Longrightarrow \beta_1$ and β_2 are no longer identified
- Identification requires at least one additional moment condition which is possible when we have what is called an **Instrumental Variable (IV)**
- An instrumental variable z_{2i} in this case is a variable such that: $E\{\varepsilon_i z_{2i}\} = 0$ **(the IV is exogenous)** and $E\{z_{2i}x_{2i}\} \neq 0$ **(the IV is relevant)**

Instrumental Variables (IV) Estimation
Single Endogenous Regressor and Single IV cont.

- In this case we have

$$E\{(y_i - x'_{1i}\beta_1 - x_{2i}\beta_2)z_{2i}\} = 0 \tag{30}$$

- Such an IV would be referred to as "exogenous" and would be sufficient to the model's K parameters
- Condition (30) is known as the **exclusion restriction**
- The **IV estimator** $\hat{\beta}_{IV}$ can then be solved from

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - x'_{1i}\hat{\beta}_{1,IV} - x_{2i}\hat{\beta}_{2,IV})x_{1i} = 0 \tag{31}$$

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - x'_{1i}\hat{\beta}_{1,IV} - x_{2i}\hat{\beta}_{2,IV})z_{2i} = 0 \tag{32}$$

- Solving these equations analytically gives the IV estimator as follows.

$$\hat{\beta}_{IV} = (\sum_{i=1}^{N} z_i x_i')^{-1} (\sum_{i=1}^{N} z_i y_i) \quad (33)$$

- where $x_i' = (x_{1i}', x_{2i})$ and $z_i' = (x_{1i}', z_{2i})$
- Do you see what happens when $z_{2i} = x_{2i}$?
- Identification of the model and consistency of the IV estimator requires that the moment conditions uniquely identify the parameters of interest

- This is equivalent to saying that π_2 in the following equation is significantly different from zero

$$x_{2i} = x'_{1i}\pi_1 + z_{2i}\pi_2 + v_i \tag{34}$$

- z_{2i} should also not be a linear combination of the elements in $x_1 i$
- If these conditions are satisfied, we say that the instrument is **relevant** (testable by $H_0 : \pi_2 = 0$)
- The IV estimator therefore would be implemented using a two-stage framework
  - Stage 1: Estimate (34) (the reduced form equation), and get the predicted values of x_{2i} (the endogenous variable)
  - Stage 2: Run OLS regression of the model using predicted values of from stage 1 instead of the endogenous varible (i.e., use $\hat{x}_{2i}$ in place of x_{2i})

- Practical challenges with the IV estimator
  1. Finding an exogenous and relevant instrument
  2. High standard errors compared to OLS
- Note however that the moment conditions we stated earlier are identifying, **they cannot be tested statistically**
- They can however be tested if there are more conditions than actually needed for identification
- One can however test endogeneity of x_{2i} using a variant of the Hausman test called (Durbin-Wu-Hausman test) by comparing the OLS and IV estimators for β provided that the instrument z_{2i} is valid

- Durbin-Wu-Hausman test: steps
- Step 1: estimate a reduced-form equation explaining x_{2i} from x_{1i} and z_{2i} and save the residuals, say $\hat{\upsilon}_i$
- Step 2: add the residuals to the mode of interest and estimate an OLS model of

$$y_i = x'_{1i}\beta_1 + x_{2i}\beta_2 + \hat{\upsilon}_i\gamma + e_i \tag{35}$$

- One can test the endogeneity of x_{2i} by performing a standard t-test on $\gamma = 0$

- Starting assumptions of ML:
  - The conditional distribution of an observed phenomenon is known, except for a finite number of unknown parameters.
  - These parameters will be estimated by taking those values for them that give the observed values the highest probability, the highest likelihood.
  - It provides an approach of estimating a set of parameters characterizing a distribution, if we know, or assume we know, the form of this distribution

Cross-sectional Binary Choice Models

- Used to model phenomenal that are of discrete nature
  - Do married women participate in the labor force?
  - Which sections of society are poor?
  - What are the determinants of an agricultural technology adoption?
- For such kinds of models, OLS is generally inappropriate - we rather use binary choice models (estimated using Maximum Likelihood Method)
- Mostly (although not exclusively) the problems analyzed are micro-economic nature

Cross-sectional Binary Choice Models Cont.

- Suppose we want to study the impact of income (assumed as the only variable here) on the probability of owning a car:

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i = x_i'\beta + \varepsilon_i \tag{36}$$

- Where, $y_i = 1$ if family i owns a car, 0 if family i does not own a car
- $x_i = (x_{i1}, x_{i2})'$
- The standard assumptions:

$$E\{\varepsilon_i|x_i\} = 0 \quad such \;\; that \quad E\{y_i|x_i\} = x_i'\beta \Longrightarrow \tag{37}$$

$$E\{y_i|x_i\} = 1.P\{y_i = 1|x_i\} + 0.P\{y_i = 0|x_i\} \tag{38}$$

$$= 1.P\{y_i = 1|x_i\} = x_i'\beta \tag{39}$$

Cross-sectional Binary Choice Models Cont.

Why OLS does not work well

- Thus, the linear model implies that $x_i'\beta$ is a probability and should therefore lie between 0& 1.
- This is only possible if the x_i values are bounded and if certain restrictions on β are satisfied.
  - Hard to achieve this in practice
- Another fundamental problem:
  - ε_i in equation 36 has a highly non-normal distribution and suffers from heteroscedasticity
  - Because y_i has only two possible outcomes, so does the error term for a given value of x_i

Cross-sectional Binary Choice Models Cont.

- We therefore use binary choice models (or univariate dichotomous models)
- Describe the probability $y_i = 1$ directly (but derived from an underlying latent variable model (see next pages)
- The general formulation is:

$$P\{y_i = 1|x_i\} = G(x_i, \beta) \tag{40}$$

  for some function $G(.)$

- Usually, we assume:

$$G(x_i, \beta) = F(x_i'\beta) \tag{41}$$

Cross-sectional Binary Choice Models Cont.

- Common choices of F are the standard normal distribution

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \Phi(z)\,dz, \qquad (42)$$

giving rise to the so-called **Probit Model**, and the standard logistic function given by:

$$L(x'\beta) = \frac{e^{x'\beta}}{(1 + e^{x'\beta})} \qquad (43)$$

leading to the **Logit Model**

Cross-sectional Binary Choice Models Cont.

- A third option is a uniform distribution over the interval [0,1] with distribution function:

$$F(x'\beta) = 0, x'\beta < 0; \tag{44}$$

$$F(x'\beta) = x'\beta, 0 \leq x'\beta \leq 1; \tag{45}$$

$$F(x'\beta) = 1, x'\beta > 1. \tag{46}$$

- Leading to what is called the **Linear Probability Model** (the OLS model), except that the probabilities are set to 0 or 1 if $x_i'\beta$ exceeds the lower upper limit respectively.

Cross-sectional Binary Choice Models Cont.

- Probit and logit are more common on applied work.
- Both the standard normal and the standard logistic random variable have an expectation of zero, while the latter has a variance of $\pi^2/3$ instead of 1.
- Correcting for the scaling difference would give similar results
- Apart from their signs, the coefficients in these binary choice models are not easy to interpret directly
- One needs to compute the marginal effects of changes in the explanatory variables

Cross-sectional Binary Choice Models Cont.

- For a continuous explanatory variable, x_{ik}, say the marginal effect is defined as the partial derivative of the probability that y_i equals one.
- For the three models above, we obtain

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k; \tag{47}$$

$$\frac{\partial L(x_i'\beta)}{\partial x_{ik}} = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2}\beta_k \tag{48}$$

$$\frac{\partial (x_i'\beta)}{\partial x_{ik}} = \beta_k; (or \;\; 0) \tag{49}$$

- Very often binary choice modes are derived from underlying behavioral model - following the latent model approach.

$$y^* = x_i'\beta + \epsilon_i \tag{50}$$

- y^* is referred to as the latent variable because it is unobserved
- Assume a probability model of working where a person chooses to work if the utility difference exceeds a certain threshold level
- Thus, one observes $y_i = 1$ (working) if and only if $y_i^* > 0$, and $y_i = 0$ (not working) otherwise.
- Hence,

$$P\{y_i = 1\} = P\{y_i^* > 0\} = P\{x_i'\beta + \epsilon_i > 0\} = P\{-\epsilon_i \leq x_i'\beta\} = F(x_i'\beta) \tag{51}$$

- Where F denotes the distribution function of $-\epsilon_i$
- Thus, depending on the distributional assumptions of ϵ_i, one can use either of the binary choice models
- A normalization on the distribution of ϵ_i is needed as the scale of utility is not identified
- Thus, we fix the variance at a given value and use the Maximum Likelihood method

- The likelihood contribution of observation i with $y_i = 1$ is given by $P\{y_i = 1|x_i\}$ as a function of β. We do the same for $y_i = 0$
- We can write the likelihood function to be maximized for the entire sample as

$$L(\beta) = \prod_{i=1}^{N} P\{y_i = 1|x_i;\beta\}^{y_i} P\{y_i = 0|x_i;\beta\}^{1-y_i} \tag{52}$$

and the corresponding log-likelihood function (which is convenient to work with) will be given by

$$logL(\beta) = \sum_{i=1}^{N} y_i logF(x_i'\beta) + \sum_{i=1}^{N}(1-y_i)log(1-F(x_i'\beta)). \tag{53}$$

- Substitute the appropriate function for F to get an expression to be maximized w.r.t β
- The values of β and their interpretation depends on the functional form used

- A goodness-of-fit measure is a summary statistic indicating the accuracy with which the model approximates the observed data, like R^2 in OLS
- In binary choice models, the dependent variable is qualitative and accuracy can be judged either in terms of the fit between the calculated probabilities and observed response frequencies or in terms of the model's ability to forecast observed responses
- Unlike the linear regression model, there is no single measure of goodness-of-fit for binary choice models
- Often, goodness-of-fit measures are based on comparison with a model that contains only a constant as explanatory variable

Cross-sectional Binary Choice Models

Goodness-of-fit Cont.

- Let $logL_1$ represent the maximum loglikelihood value of the model of interest and let $logL_0$ denote the maximum value of the loglikelihood function when all parameters, except the intercept are zero
- Obviously, $logL_1 \geq logL_0 \implies$ the larger the difference between the two likelihood values, the more the extended model adds to the very restrictive model
- A formal likelihood ratio test can be based on the difference between the two values
- One popular measure goodness-of-fit measure is given by

$$pseudo - R^2 = 1 - \frac{1}{1 + 2(logL_1 - logL_0)/N} \qquad (54)$$

- A panel data set contains repeated observations over the same units (individuals, households, firms, countries) collected over a number of periods
- Can be micro or macro level
  1. Allows estimation of more complicated and more realistic models than a single cross-section or a single timeseries data
  2. Panel data are larger than cross-sectional and time-series data with variation over two dimensions (individuals and time) $\Longrightarrow$ estimators are more accurate.
  3. Even with identical sample sizes, the use of panel data will often yield more efficient estimators than a series of independent cross-sections,
  4. Panel data reduces identification problems.
     - Identification in the presence of endogenous regressors or measurement error
     - Robustness to omitted variables
     - Identification of individual dynamics

- A linear regression model in which the intercept terms very over the individual units i, i.e.

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \qquad u_{it} \sim IID(0, \sigma_u^2), \& E(x'_{it}u_{it}) = 0. \tag{55}$$

- One can rewrite this model in the usual way by introducing a dummy variable for each unit i,

$$y_{it} = \sum_{j=1}^{n} \alpha_j d_{ij} + x'_{it}\beta + u_{it}, \tag{56}$$

- Where $d_{ij} = 1$ if $i = j$ & 0 otherwise

- We thus have N dummy variables in the model. The parameters $\alpha_1, ..., \alpha_N$ & β can be estimated by OLS
- The implied estimator for β is referred to as the *Least Square Dummy Variable* (LSDV) estimator.
- It may however be numerically inconvenient to have so many dummy variables in a regression model. Why?
- It is possible to estimate β in a simpler way! This involves transformation of the data to eliminate α_i
- To do so, note that

$$\overline{y}_i = \alpha_i + \overline{x}_i'\beta + \overline{u}_i, \tag{57}$$

- Where $\overline{y}_i = T^{-1}\sum_t y_{it}$ & $\overline{x}_i$ & $\overline{u}_i$ are defined in a similar way. Consequently, one can write

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)'\beta + (u_{it} - \overline{u}_i) \tag{58}$$

- You can see that α_i is gone!
- The transformation that produces equation (7) is called the *within transformation*
- The OLS estimator for β obtained from this transformed model is often called the *within or fixed effects estimator* and is exactly identical to the LSDV estimator described above.

- The within estimator is given by,

$$\hat{\beta}_{FE} = (\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)')^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(y_{it} - \overline{y}_i) \tag{59}$$

- For consistency, it is required that

$$E\{(x_{it} - \overline{x}_i)u_{it}\} = 0 \tag{60}$$

- Sufficient for this is that x_{it} is uncorrelated with u_{it}, and that $\overline{x}_i$ has no correlation with the error term! $\Longrightarrow$

$$E\{x_{it}u_{is}\} = 0 \quad for \quad all \quad s, t. \tag{61}$$

- $\Longrightarrow$ x_{it} are *strictly exogenous*.

- An alternative way to illuminate α_i is to first-difference Eq(55) as

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})'\beta + (u_{it} - u_{i,t-1})$$

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta u_{it} \tag{62}$$

- Applying OLS to this equation gives the First Differences (FD) estimator

$$\hat{\beta}_{FD} = (\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}\Delta x_{it}')^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}\Delta y_{it}. \tag{63}$$

- Consistency requires: $E\{\Delta x_{it} \Delta u_{it}\} = 0$ or

$$E\{(x_{it} - x_{it-1})(u_{it} - u_{it-1})\} = 0 \qquad (64)$$

- This is a weaker condition than the strict exogeneity condition in Eq (61).
- For eg. it would allow correlation between x_{it} and u_{it-2}
- Computation of standard errors for $\hat{\beta}_{FD}$ requires taking in to account serial correlation in Δu_{it}
- Since the conditions for consistency of the FD estimator are slightly weaker than those for the FE estimator, it is in general, somewhat less efficient.
- For $T = 2$, both estimators are identical.
- If the two estimators provide very different results $\Longrightarrow$ assumption (61) is problematic!

- In this model, α_i is assumed to be random factors, independently and identically distributed over individuals. The model is specified as

$$y_{it} = \beta_0 + x'_{it}\beta + \alpha_i + u_{it}, \quad u_{it} \sim IID(0, \sigma_u^2); \quad \alpha_i \sim IID(0, \sigma_\alpha^2) \tag{65}$$

$$\varepsilon_{it} = \alpha_i + u_{it} \tag{66}$$

- Also referred to as a *one-way error components model*.
- All correlation of ε_{it} is attributed to α_i
- It is assumed that α_i and u_{it} are mutually independent and independent of x_{js}, $\forall j \& s$
- $\implies$ estimating Eq(65) by OLS results in unbiased and consistent parameter estimates.

- However, ε_{it} exhibits a particular form of autocorrelation (unless in a special case where $\sigma_\alpha^2 = 0$)
- Consequently, the standard errors computed using OLS are incorrect and a more efficient estimator that exploits the structure of the error covariance matrix can be computed using GLS.
- Remember:

$$\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2} \tag{67}$$

- Note that $I_T - (1/T)l_T l_T'$ transforms the data in deviations from individual means and $(1/T)l_T l_T'$ takes individual means.
- The GLS estimator for β can be given by

$$\hat{\beta}_{GLS} = (\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)' + \psi T \sum_{i=1}^{N}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})')^{-1}$$

$$\times(\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(y_{it} - \overline{y}_i) + \psi T \sum_{i=1}^{N}(\overline{x}_i - \overline{x})(\overline{y}_i - \overline{y})) \quad (68)$$

- Where $\overline{x}$ refers to overall average of x_{it}
- If $T \Longrightarrow \infty$, $\psi \Longrightarrow 0 \Longrightarrow$, the FE estimator arises (FE & RE become equivalent).
- If $\psi = 1$, the GLS estimator is just the OLS estimator (and Ω is diagonal).

- From the general formula for the GLS, one can derive:

$$\hat{\beta}_{GLS} = W\hat{\beta}_B + (I_K - W)\hat{\beta}_{FE}, \tag{69}$$

- Where,

$$\hat{\beta}_B = (\sum_{i=1}^{N}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})')^{-1}\sum_{i=1}^{N}(\overline{x}_i - \overline{x})(\overline{y}_i - \overline{y}), \tag{70}$$

- Is called the *between estimator* for β.
- Is nothing but the OLS estimator in the model for individual means

$$\overline{y}_i = \beta_0 + \overline{x}_i'\beta + \alpha_i + \overline{u}_i, \quad i = 1, ..., N. \tag{71}$$

- The choice between the two is not easy, and in many applications, particularly when T is small, the differences in the estimates for β can be substantial.
- The common approach is to use *Hausman's test* - a test for $H_O : E\{x_{it}, \alpha_i\} = 0$
- A significant difference in the two estimators implies that the null hypothesis is unlikely to hold.
- Assume that $E\{u_{it}x_{is} = 0\}$ for all s, t, so that $\hat{\beta}_{FE}$ is consistent for β whether $x_{it}\&\alpha_i$ are uncorrelated, whereas $\hat{\beta}_{RE}$ is consistent and efficient only if x_{it} and α_i are not correlated.
- Hausma's test evaluates the difference between the two estimators $(\beta_{FE} - \beta_{RE})$.
- It is easy to show that under the null,

$$V\{\hat{\beta}_{FE} - \hat{\beta}_{RE}\} = V\{\hat{\beta}_{FE}\} - V\{\hat{\beta}_{RE}\}. \tag{72}$$

- Thus, the Hausman statistic is given by

$$\xi_H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\hat{V}\{\hat{\beta}_{FE}\} - \hat{V}\{\hat{\beta}_{RE}\}]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \tag{73}$$

- Where the $\hat{V}s$ denote estimates of the true covariance matrices
- Under the null hypothesis, $(plim(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = 0$, the statistic ξ_H has an asymptotic Chi-squared distribution with K degrees of freedom, where K is the number of elements in β

- The FE model eliminates anything that is time invariant (E.g, Gender) from the model
- Imposing the assumption that $E\{x_{it}, \alpha_i\} = 0$ to deal with the above problem might be unjustifiable
- The way out is to use an IV method which is considered to be in between the FE & RE approaches
- To elaborate the general approach of the HT estimator, consider a linear model with four groups of explanatory variables:

$$y_{it} = \beta_0 + x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + w'_{1i}\gamma_1 + w'_{2i}\gamma_2 + \alpha_i + u_{it}, \quad (74)$$

- Where the x vars are time varying and the w vars are time invariant

- The variables with index 1 are assumed to be uncorrelated with both α_i & u_{it}, while the ones with index 2 are correlated with α_i but not with u_{it}
- Hausman and Taylor show that equation (74) can be estimated by instrumental variables using the following variables as instruments: $x_{1,it}, w_{1i}$ & $x_{2,it} - \overline{x}_{2i}, \overline{x}_{1i}$.
- Note:
    - The exogenous variables serve as their own instruments
    - $x_{2,it}$ is instrumented by its deviation from individual means (as in the FE approach)
    - w_{2i} is instrumented by the individual average of $x_{1,it}$
    - We don't need to use external instruments! (an attractive advantage of the HT model)
- Identification requires that the number of variables in $x_{1,it}$ is at least as large as that in w_{2i}.

- **The End!**