London School of Economics
Department of Economics

Vassilis Hajivassiliou
Winter Term 2024

Further Topics in Econometrics (Ec485)
Answers to Problem Set #1 — Panel Data Models

1. Consider the "dummy variable" linear regression model:

$$y_i = i_T \alpha_i + X_i \beta + \epsilon_i$$

for a balanced panel data set with $i = 1, \cdots, N$ cross-sectional units, each observed for T time periods. y_i and ϵ_i are $T \times 1$ vectors, X_i a $T \times k$ matrix, β a $k \times 1$ vector of unknown slope parameters, and there is a different intercept α_i for each unit i. Define N dummy variable vectors (of dimension $NT \times 1$) indicating the different units, e.g., d_i is an $NT \times 1$ vector with typical element

$$d_{it} = \begin{cases} 1 & \text{if observation } it \text{ refers to individual unit } i \\ 0 & \text{otherwise} \end{cases}$$

Stacking the observations for all N units in the standard way and defining the vector α ($N \times 1$) conformably, gives the matrix formulation:

$$y = D\alpha + X\beta + \epsilon.$$

Define the usual projection matrix $M_d = I - D(D'D)^{-1}D'$. Use standard partitioned-regression results to show that:

(a) The OLS coefficient vector $\hat{\beta}$, known as the "fixed-effects" estimator, can be obtained by regressing $\{y_{it} - \bar{y}_{i.}\}$ on $\{x_{it} - \bar{x}_{i.}\}$, where $y_{i.}$ is the mean of the T observations of i for the y variable, and $x_{i.}$ is the $k \times 1$ vector of means of the x variables over the T observations of i.
ANSWER: *see Greene, 16.4.2*

(b) The OLS estimates for the N intercepts are:

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{x}'_{i.}\hat{\beta}.$$

ANSWER: *see Greene, 16.4.2*

(c) The disturbance variance estimator is:

$$s^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \hat{\alpha}_i - x'_{it}\hat{\beta})}{NT - N - k}$$

How does this expression differ from the one obtained by regressing $y_{it} - \bar{y}_{i.}$ on $x_{it} - \bar{x}_{i.}$?

ANSWER:

The FE/Within estimator corresponds to OLS estimation of the model

$$y_{it} - \bar{y}_{i\cdot} = (x_{it} - \bar{x}_{i\cdot}) + (\epsilon_{it} - \bar{\epsilon}_{i\cdot}) \tag{1}$$

say

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon}$$

where:

$$\tilde{\epsilon} = \{\alpha_i + \nu_{it} - \alpha_i - \bar{\nu}_{i\cdot}\} = \{\nu_{it} - \bar{\nu}_{i\cdot}\} = Q\epsilon$$

and:

$$Q \equiv I_{NT} - P \quad \text{and} \quad P \equiv I_N \otimes (\frac{1}{T} i_T i_T')$$

OLS on equation (1) gives sum of squared residuals having expectation

$$ERSS = \sigma_\nu^2 traceQ = \sigma_\nu^2 (NT - N(T-1)) = \sigma_\nu^2 N(T-1)$$

[NB: you may want to subtract the number of regressors, k, but that is irrelevant for $N \to \infty$]

But the package thinks that the degrees of freedom are $NT - k$, which is incorrect.

Thus, to obtain a consistent estimator for for σ_ν^2 we can use:

$$s_\nu^2 = \frac{NT - k}{N(T-1) - k} \cdot s^2_{FE\_by\_OLS}$$

Intuition: regular OLS ignores the fact that N dummy variables (with the $\alpha_i s$) are implicitly present in the model.

1. (a) & (b): See Greene 16.4.2 pp 466-8

(c) The FE/within estimator corresponds to OLS estimation of the model

$$y_{it} - \bar{y}_{i\cdot} = (x_{it} - \bar{x}_{i\cdot}) \cdot \beta + (\varepsilon_{it} - \bar{\varepsilon}_{i\cdot}) \qquad *$$

Say $\tilde{y} = \tilde{X} \beta + \tilde{\varepsilon}$

where $\tilde{\varepsilon} = \{ \alpha_i + v_{it} - \alpha_i - \bar{v}_{i\cdot} \} = \{ v_{it} - \bar{v}_{i\cdot} \} = Q \varepsilon$

where $Q \equiv I_{NT} - P$ and $P \equiv I_N \otimes \left(\frac{1}{T} i_T i_T' \right)$

OLS on * gives $\hat{\tilde{\varepsilon}}' \hat{\tilde{\varepsilon}}$ with $E = \sigma_v^2 \text{ trace } Q$
$= \sigma_v^2 (NT - NT \frac{1}{T}) = \sigma_v^2 N(T-1)$ [you may want to subtract K - irrelevant for $N \to \infty$]

But the package thinks the degrees of freedom are $NT - K$
(n.b. K = # regr)

So consistent $s_v^2 = \frac{NT-K}{N(T-1)-K} \cdot s^2_{FE\,by\,OLS}$
(with $N \to \infty$ T small)

Intuition: regular OLS ignores the fact that N dummy variables (with α_i's) are implicitly present in the model.

2. Consider the linear regression model:

$$y_{it} = X_{it}\beta + \epsilon_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T_i.$$

for an *unbalanced* panel-data set N cross-sectional units, observed for possibly different numbers T_i of time periods. The regressors are fixed in repeated samples.
The disturbance term is believed to have the *one-factor random-effects* structure:
$\epsilon_{it} = \alpha_i + \nu_{it}$ with α_i independent of ν_{jt} for any i, j, t, $\alpha_i \sim (0, \sigma^2_\alpha)$ i.i.d. over i, and $\nu_{it} \sim (0, \sigma^2_\nu)$ i.i.d. over both i and t. Define

$$\theta_i \equiv 1 - \sqrt{\frac{\sigma^2_\nu}{T_i\sigma^2_\alpha + \sigma^2_\nu}}$$

(a) Show that the transformed error term: $\epsilon^*_{it} \equiv \epsilon_{it} - \lambda_i\bar{\epsilon}_{i\cdot}$ where $\epsilon_{it} \equiv \frac{1}{T_i}\sum_t \epsilon_{it}$, satisfies the Gauss-Markov conditions. Specifically, you should show that ϵ^*_{it} is homoskedastic and serially uncorrelated.

ANSWER:

Consider $\epsilon^*_{it} \equiv \epsilon_{it} - \lambda_i\bar{\epsilon}_{i\cdot}$. *Let:*

$$\rho \equiv \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\nu} = \frac{\sigma^2_\alpha}{\sigma^2_\epsilon} \quad \text{and} \quad \Omega = \sigma^2_\epsilon \left(\begin{array}{cccc|c} 1 & \rho & \cdots & \rho & \\ & \ddots & \ddots & \rho & \\ & & \ddots & \rho & 0 \\ & & & 1 & \\ \hline & & & & \\ & & 0 & & 0 \\ & & & & \end{array} \right)$$

In other words,

$$corr(\epsilon_{it}, \epsilon_{is}) = \begin{cases} \rho & for \quad t \neq s \\ 1 & for \quad t = s \end{cases} \quad \text{and} \quad cov(\epsilon_{it}, \epsilon_{is}) = \begin{cases} \rho \cdot \sigma^2_\epsilon & for \quad t \neq s \\ \sigma^2_\epsilon & for \quad t = s \end{cases}$$

Assume a balanced Panel Data Set with $T_i = T$ for all i. We also use I to denote the identity matrix I_{NT} of dimension $NT \times NT$.

Then:

$$\begin{aligned}
E\epsilon^*\epsilon^{*\prime} &= (I-\lambda P)E\epsilon\epsilon'(I-\lambda P) \\
&= (I-\lambda P)(\sigma_\nu^2 I + \sigma_\alpha^2 P)(I-\lambda P) \\
&= (\sigma_\nu^2 I + \sigma_\alpha^2 TP - \sigma_\nu^2\lambda P - \sigma_\alpha^2 T\lambda P)(I-\lambda P) \\
&= [\sigma_\nu^2 I + (\sigma_\alpha^2 T(1-\lambda) - \sigma_\nu^2\lambda)P](I-\lambda P) \\
&= \sigma_\nu^2 I + P[\sigma_\alpha^2 T(1-\lambda) - \sigma_\nu^2\lambda - \sigma_\nu^2\lambda - \lambda(\sigma_\alpha^2 T(1-\lambda) - \sigma_\nu^2\lambda)] \\
&= \sigma_\nu^2 I + P[\sigma_\alpha^2 T(1-\lambda)^2 + \sigma_\nu^2\lambda^2 - 2\sigma_\nu^2\lambda] \\
&= \sigma_\nu^2 I + P[\sigma_\alpha^2 T(1-\lambda)^2 + \sigma_\nu^2(1-\lambda)^2 - \sigma_\nu^2] \\
&= \sigma_\nu^2 I + P[(\sigma_\alpha^2 T + \sigma_\nu^2)(1-\lambda)^2 - \sigma_\nu^2] \\
&= \sigma_\nu^2 I
\end{aligned}$$

if the term in square brackets after the P matrix vanishes, i.e., if

$$\begin{aligned}
(\sigma_\alpha^2 T + \sigma_\nu^2)(1-\lambda)^2 - \sigma_\nu^2 &= 0 \\
(1-\lambda)^2 &= \frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2} \\
\lambda &= 1 \pm \sqrt{\frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2}}
\end{aligned}$$

The $1+\sqrt{}$ root is discarded because it would lead to a non-convex combination between the two base analysis-of-covariance transformations, namely the Within and the Between — see Question 3 below.

Hence we have established that for

$$\lambda \equiv 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2}}$$

the transformed error ϵ^* with typical element $\epsilon_{it}^* \equiv \epsilon_{it} - \lambda_i \bar{\epsilon}_{i\cdot}$ has a scalar VCov matrix $E\epsilon^*\epsilon^{*\prime} = \sigma_\nu^2 I_{NT}$. Hence it is homoskedastic and not autocorrelated, a pure Gauss-Markov error over both i and t.

[NB: a more tedious derivation follows without linear algebra, by establishing the condition which makes:

$$cov(\epsilon_{it}^*, \epsilon_{is}^*) = 0$$

for $t \neq s$]

In this context, the convex combination idea fits nicely: Consider the convex combination of the Within and Between transformations:

$$\theta\{z_{it} - \bar{z}_{i\cdot}\} + (1-\theta)\bar{z}_{i\cdot}$$

with $0 \leq \theta \leq 1$ At one extreme, $\theta = 0$ we obtain the Between transformation, while for $\theta = 1$ we obtain the Within/FE transformation.

The optimal θ defines the GLS/RE transformation, and is somewhere in between the two extremes. I.e.,

$$0 \leq \theta^{optimal} = \lambda \equiv 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2}} \leq 1$$

*Since *both* the lambda roots for GLS/RE give a scalar VCov $\sigma_\nu^2 I$ for the GLS-transformed error vector, either of them makes this error have a spherical distribution with Homoskedasticity and no autocorrelation. To understand the issue of the convex combination, recall that the matrix P transforms a vector (e.g., the residual vector epsilon) into time-averages for each individual (repeated for all time observations for each individual so the transformed vector will be of the same dimension as the original one). Similarly, its orthogonal matrix Q=(I-P) will change a vector into *deviations* from the individual time-averages. So Py and PX would be the transformed dep.var. and regressors for "Between" analysis, while Qy and QX would be the transformed dependant variable and regressors suitable for Within or Fixed-Effects analysis. The GLS/RandomEffects estimator is the optimal convex combination of Between and Within variation. To see this, note that the GLS transformation is: I - lambda*P = I - lambda*(I-Q) = (1-lambda)*I + lambda*Q Consider the convex combination of the P and Q transformations:*

*lambda*P + (1-lambda)*Q*

Note that both formulae reduce to Q (=Within/FE) if lambda=1; and they reduce to P (=Between) if lambda=0.

The convex combination fraction is between 0 and 1 depending on T, σ_ν^2 and σ_α^2 . So $\lambda \equiv 1 + \sqrt{\frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2}}$ will exceed 1, while $\lambda \equiv 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\alpha^2 T + \sigma_\nu^2}}$ will be between 0 and 1 as needed. This is why we reject the 1+... root and keep the 1-... one.

(b) Use the previous result to define a convenient implementation of the GLS estimator for this model.

ANSWER: the GLS transformation with the optimal λ changes the error term to a Gauss-Markov one with Homoskedasticity and No autocorrelation. Hence the BLUE estimator, which is the GLS, is identical to plain OLS on the transformed model.

(c) How would you obtain a consistent estimator for λ_i which you would need to define the feasible GLS estimator?

ANSWER: see lecture notes.

(2) $\varepsilon_{it}^* = \varepsilon_{it} - \lambda_i \bar{\varepsilon}_{i\cdot}$

Let $\rho \equiv \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_v^2} = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$ $\quad \Omega = \sigma_\varepsilon^2 \begin{pmatrix} \begin{matrix} 1 & \rho & \cdots & \rho \\ & \ddots & & \\ & & 1 & \rho \end{matrix} & \underset{\sim}{O} \\ \underset{\sim}{O} & \end{pmatrix}$

i.e., $\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) = \rho \quad t \neq s$
$= 1 \quad t = s$

& $\text{Cov}(\varepsilon_{it}, \varepsilon_{is}) = \rho \cdot \sigma_\varepsilon^2 \quad t \neq s$
$= \sigma_\varepsilon^2 \quad t = s$

$$\text{Cov}(\varepsilon_{it}^*, \varepsilon_{is}^*) = E(\varepsilon_{it} - \lambda_i \bar{\varepsilon}_{i\cdot})(\varepsilon_{is} - \lambda_i \bar{\varepsilon}_{i\cdot})$$

$$= \sigma_\varepsilon^2 \rho + \lambda_i^2 E \frac{1}{T_i^2} \left(\sum_t \varepsilon_{it} \sum_s \varepsilon_{is} \right) - 2 \frac{\lambda_i}{T_i} E\, \varepsilon_{it} \sum_s \varepsilon_{is}$$

$$= \sigma_\varepsilon^2 \rho + \frac{\lambda_i^2}{T_i^2} \left[T_i \sigma_\varepsilon^2 + T_i (T_i - 1) \rho \sigma_\varepsilon^2 \right] - 2 \lambda_i \frac{1}{T_i} \left[\sigma_\varepsilon^2 + (T_i - 1) \rho \sigma_\varepsilon^2 \right]$$

$$= \frac{\sigma_\varepsilon^2}{T_i} \Big[\rho T_i + \lambda_i^2 \underbrace{\left[1 + (T_i - 1)\rho \right]}_{\delta_i} - 2 \lambda_i \underbrace{\left[1 + (T_i - 1)\rho \right]}_{\delta_i} \Big]$$

$$= \frac{\sigma_\varepsilon^2}{T_i} \left[\rho T_i + \lambda_i^2 \delta_i - 2 \lambda_i \delta_i \right] = 0$$

iff $\rho T_i + \lambda_i^2 \delta_i - 2 \lambda_i \delta_i = 0$, a quadratic in λ_i

with roots $\dfrac{\delta_i \pm \sqrt{\delta_i^2 - T_i \rho \delta_i}}{\delta_i} = 1 \pm \sqrt{1 - \dfrac{T_i \rho}{\delta_i}}$

$$= 1 \pm \sqrt{\frac{\sigma_v^2}{T_i \sigma_\alpha^2 + \sigma_v^2}} \; \begin{cases} + & \text{inadmissible} \\ - & \lambda_i \end{cases}$$

Similarly for $\text{Var}(\varepsilon_{it}^*) = V(\varepsilon_{it} - \lambda_i \bar{\varepsilon}_{i\cdot})$.

Alternative derivation using matrixalgebra:

Assume balanced PDS for simplicity

$$E\varepsilon^*\varepsilon^{*\prime} = (I-\lambda P)E\varepsilon\varepsilon'(I-\lambda P) = (I-\lambda P)(\sigma_v^2 I + \sigma_\alpha^2 TP)(I-\lambda P)$$

$$= (\sigma_v^2 I + \sigma_\alpha^2 TP - \sigma_v^2 \lambda P - \sigma_\alpha^2 T\lambda P)(I-\lambda P)$$

$$= [\sigma_v^2 I + (\sigma_\alpha^2 T(1-\lambda) - \sigma_v^2 \lambda)P](I-\lambda P)$$

$$= \sigma_v^2 I + P[\sigma_\alpha^2 T(1-\lambda) - \sigma_v^2\lambda - \sigma_v^2\lambda - \lambda(\sigma_\alpha^2 T(1-\lambda) - \sigma_v^2\lambda)]$$

$$= \sigma_v^2 I + P[\sigma_\alpha^2 T(1-\lambda)^2 + \lambda^2\sigma_v^2 - 2\lambda\sigma_v^2]$$

$$= \sigma_v^2 I + P[\sigma_\alpha^2 T(1-\lambda)^2 + (1-\lambda)^2\sigma_v^2 - \sigma_v^2]$$

$$= \sigma_v^2 I + P[(\sigma_\alpha^2 T + \sigma_v^2)(1-\lambda)^2 - \sigma_v^2] = \sigma_v^2 I$$

Since $(1-\lambda)^2 \equiv \dfrac{\sigma_v^2}{\sigma_\alpha^2 T + \sigma_v^2}$

3. For a balanced panel data set, recall the transformations:

$$\begin{array}{ll} \{z_{it} - \bar{z}_{i\cdot}\} & \text{“}Within''\\ \{\bar{z}_{i\cdot}\} & \text{“}Between''\\ \{z_{it} - \lambda\bar{z}_{i\cdot}\} & \text{“}GLS'' \end{array}$$

Running OLS on the “within-”, “between-”, and “GLS-” transformed models defines the $\hat{\beta}_W$, $\hat{\beta}_B$ and $\hat{\beta}_{GLS}$ respectively. It can be shown (see Greene, sections 14.3–14.4) that $\hat{\beta}_{GLS}$ is a matrix-weighted average of $\hat{\beta}_W$ and $\hat{\beta}_B$. Specifically,

$$\hat{\beta}_{GLS} = F^W\hat{\beta}_W + (I - F^W)\hat{\beta}_B,$$

where $F^W \equiv [S^W_{XX} + (1-\lambda)^2 S^B_{XX}]^{-1} S^W_{XX}$, θ was defined above, and $S^{W,B}_{XX}$ are sample-moment matrices of the X variables from the W, B transformations respectively.
Define three alternative Wu-Hausman statistics based on the three difference vectors:

$$\hat{d}_1 = \hat{\beta}_B - \hat{\beta}_W, \quad \hat{d}_2 = \hat{\beta}_{GLS} - \hat{\beta}_W, \quad \hat{d}_3 = \hat{\beta}_{GLS} - \hat{\beta}_B.$$

(a) Show that if λ is known exactly (i.e., does not need to be estimated) the three Wu-Hausman tests will be *algebraically* equivalent.
ANSWER:

$$\begin{aligned} \hat{d}_1 &= \hat{\beta}_B - \hat{\beta}_W\\ \hat{d}_2 &= \hat{\beta}_{GLS} - \hat{\beta}_W\\ &= F\hat{\beta}_W + (I-F)\hat{\beta}_B - \hat{\beta}_W = (I-F)(\hat{\beta}_B - \hat{\beta}_W)\\ &= (I-F)\hat{d}_1\\ \hat{d}_3 &= \hat{\beta}_{GLS} - \hat{\beta}_B\\ &= F\hat{\beta}_W + (I-F)\hat{\beta}_B - \hat{\beta}_B = -F\hat{\beta}_B + F\hat{\beta}_W\\ &= -F\hat{d}_1 \end{aligned}$$

Hence we see that:

$$\begin{array}{lll} \hat{d}_2 = A_2\hat{d}_1\ , & A2 = I - F & \text{nonsingular}\\ \hat{d}_3 = A_3\hat{d}_1\ , & A_3 = -F & \text{nonsingular} \end{array}$$

and

$$\begin{aligned} \hat{q}_1 &= \hat{d}_1'[V(\hat{d}_1)]^{-1}\hat{d}_1\\ \hat{q}_2 &= \hat{d}_2'[V(\hat{d}_2)]^{-1}\hat{d}_2 = \hat{d}_1'A_2'[A_2V(\hat{d}_1)A_2']^{-1}A_2\hat{d}_1\\ &= \hat{d}_1'A_2'(A_2')^{-1}V(\hat{d}_1)^{-1}(A_2)^{-1}A_2\hat{d}_1\\ &= \hat{d}_1'V(\hat{d}_1)^{-1}\hat{d}_1\\ &= \hat{q}_1 \end{aligned}$$

Similarly for $\hat{q}_3 = \hat{q}_1$.
General result: Quadratic form of z in $V(z)^{-1}$ is the same as quadratic form of Az in $V(Az)^{-1}$ for any fixed nonsingular matrix A.

(b) What types of hypotheses can these statistics be used to test? When would these test procedures have high power?
ANSWER:
If $\epsilon_{it} = \alpha_i + \nu_{it}$ with $E(\nu_{it}|X) = 0$ but $E(\alpha_i|X) \neq 0$, GLS/RE and B will be inconsistent, while FE/W will be consistent. Hence can consider Hausman tests based on three differences vectors:

$$\hat{d}_1 \; : \; V(\hat{d}_1) = V_B + V_W - 2 \cdot 0 \qquad \text{since } P \text{ and } Q = I - P \text{ are orthogonal}$$
$$\hat{d}_2 \; : \; V(\hat{d}_2) = V_W - V_{GLS} \qquad \text{by the asymptotic Rao-Blackwell theorem since GLS is efficient un}$$
$$\hat{d}_3 \; : \; V(\hat{d}_3) = V_B - V_{GLS} \qquad \text{again by Rao-Blackwell}$$

A HT will have high power if:
(i) the two estimators that are differenced have different probability limits under the alternative H_1.
[For $\hat{d}_1$, there is inconsistency due to $\hat{\beta}_B$; for $\hat{d}_2$, there is inconsistency due to $\hat{\beta}_{GLS}$; and for $\hat{d}_3$, there is inconsistency due to both $\hat{\beta}_{GLS}$ and $\hat{\beta}_B$.]
And:
(ii) the asymptotic VCov $AV(V(\hat{d}_j))$, $j = 1, 2, 3$, is "low".

(c) Define a fourth Wu-Hausman statistic based on the difference vector: $\hat{d}_4 = \hat{\beta}_{GLS} - \hat{\beta}_{OLS}$ where $\hat{\beta}_{OLS}$ is the OLS estimator from the untransformed data.

i. Explain how you would calculate the variance-covariance matrix of $\hat{d}_4$.
NB: You do *not* need to calculate the precise expression — simply explain what the issues are.
ANSWER:
Note that:
$$V(\hat{d}_4) = \hat{V}_{OLS} - \hat{V}_{GLS}$$
since GLS is efficient under H_0. But of course
$$\hat{V}_{OLS} \neq \sigma^2_\epsilon (X'X)^{-1}$$
since $\epsilon_{it} = \alpha_i + \nu_{it}$ implies serial correlation in ϵ_{it}. So need formula
$$(X'X)^{-1} X' \Omega X (X'X)^{-1}$$

ii. Would such a test have good power properties?
ANSWER:
Power issues as above. NOTE: if both GLS and OLS remain consistent under H_1 (e.g., $E\alpha|X = 0$ and $E\nu|X = 0$) then the test would have zero power.

③ (a) $\hat{d}_1 = (\hat{\beta}_B - \hat{\beta}_W)$

$$\hat{d}_2 = (\hat{\beta}_{GLS} - \hat{\beta}_W) = F\hat{\beta}_W + (I-F)\hat{\beta}_B - \hat{\beta}_W = (I-F)(\hat{\beta}_B - \hat{\beta}_W)$$
$$= (I-F)\hat{d}_1$$

$$\hat{d}_3 = (\hat{\beta}_{GLS} - \hat{\beta}_B) = F\hat{\beta}_W + (I-F)\hat{\beta}_B - \hat{\beta}_B = -F\hat{\beta}_B + F\hat{\beta}_W$$
$$= -F\hat{d}_1.$$

So $\hat{d}_2 = A_2\hat{d}_1$, $A_2 = I - F$, non singular

and $\hat{d}_3 = A_3\hat{d}_1$, $A_3 = -F$, nonsingular

$$\hat{q}_1 = \hat{d}_1'[V(\hat{d}_1)]^{-1}\hat{d}_1$$

$$\hat{q}_2 = \hat{d}_2'[V(\hat{d}_2)]^{-1}\hat{d}_2 = \hat{d}_1' A_2' [A_2 V(\hat{d}_1) A_2']^{-1} A_2 \hat{d}_1$$
$$= \hat{d}_1' A_2' (A_2')^{-1} V(\hat{d}_1)^{-1} A_2^{-1} A_2 \hat{d}_1 = \hat{d}_1' V(\hat{d}_1)^{-1}\hat{d}_1 = \hat{q}_1$$

Similarly for $\hat{q}_3 = \hat{q}_1$.

<u>General result</u>: Quadratic form of Z in $V(Z)^{-1}$ is the <u>same</u> as Q.F. of AZ in $V(AZ)^{-1}$ for any fixed nonsingular A.

(b) If $\varepsilon_{it} = v_{it} + \alpha_i$ with $E(v_{it} | x_{it}) = 0$ but $E\alpha_i | x_{it} \neq 0$, GLS/RE of B will be <u>inconsistent</u>, while FE/W will be consistent.

Hence: HT based on $\hat{d}_1$: $V(\hat{d}_1) = V_B + V_W - 2 \cdot 0$
since $\leftarrow$ P & Q = I − P are orthogonal.

$\hat{d}_2$: $V(\hat{d}_2) = V_W - V_{GLS}$
(Rao Blackwell since GLS efficient under H_0).

$\hat{d}_3$: $V(\hat{d}_3) = V_B - V_{GLS}$
(again Rao-Blackwell).

High Power iff $\text{plim}\, \hat{d}' [AV(V(\hat{d}))]^{-1} \text{plim}\, \hat{d}$ is high under H_1.

I.e., iff the two estimators are different under H_1

($\hat{d}_1$: inconsistency due to $\hat{\beta}_B$
$\hat{d}_2$: — — — — — $\hat{\beta}_{GLS}$
$\hat{d}_3$: — — — — — $\hat{\beta}_{GLS}$ & $\hat{\beta}_B$).

and $AV(V(\hat{d}))$ is "low"

(c) $V(\hat{d}_4) = \hat{V}_{OLS} - \hat{V}_{GLS}$ since GLS efficient under H_0.

But (i) $\hat{V}_{OLS} \neq \sigma_\varepsilon^2 (X'X)^{-1}$ since $\varepsilon_{it} = \alpha_i + v_{it}$ implies serial correlation in ε_{it}. So need formula $(X'X)^{-1} X' \Omega X (X'X)^{-1}$

(ii) Power issues as above. Note: if both GLS & OLS remain consistent under H_1, (e.g. $E\alpha/x = 0$, $Ev/x = 0$) then test would have 0 power.

4. The classic *linear dynamic balanced panel data model*:

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + z'_i\gamma + \alpha_i + \nu_{it}, \qquad i = 1, \cdots, N, \quad t = 1, \cdots, T$$

(a) The Bargava and Sargan approach:
Step 1 – write out explicitly as a separate equation for each t, so a cross-section on a System of T equations for the T endogenous variables $y_1, y_2, \cdots, y_T$ and the exogenous variables $x_{iT}, x_{i,T-1}, \cdots, x_{i1}, z_i$
Step 2 – write out a linear quasi-reduced form equation for y_{i1} in terms of full exogenous information available to the econometrician (but not actually available to the economic agents at time $t = 1$):

$$\begin{array}{rcl} y_{iT} & = & \delta y_{i,T-1} + x'_{iT}\beta + z'_i\gamma + \alpha_i + \nu_{iT} \\ y_{i,T-1} & = & \delta y_{i,T-2} + x'_{i,T-1}\beta + z'_i\gamma + \alpha_i + \nu_{i,T-1} \\ \vdots & \vdots & \vdots \\ y_{i2} & = & \delta y_{i1} + x'_{i2}\beta + z'_i\gamma + \alpha_i + \nu_{i2} \\ y_{i1} & = & x'_{i1}\theta_1 + x'_{i2}\theta_2 + \cdots + x'_{iT}\theta_T + z'_i\zeta + u_{i1} \end{array} \qquad i = 1, \cdots, N$$

i. Optimal estimation assuming normality of the errors is achieved through *Full Information MLE* of $(\delta, \beta', \gamma', \theta'_1, \cdots, \theta'_T, \zeta')'$ and $(\sigma^2_\alpha, \sigma^2_\nu, \sigma^2_0)$ implied by the cross-equation restrictions of the above system and the variance-covariance restrictions of the structure:

$$\begin{pmatrix} \omega^2_T & \omega_{T,T-1} & \omega_{T,T-2} & \cdots & \omega_{T2} & \omega_{T1} \\ \omega_{T-1,T} & \omega^2_{T-1} & \omega_{T-1,T-2} & \cdots & \omega_{T-1,2} & \omega_{T-1,1} \\ \omega_{T-2,T} & \omega_{T-2,T-1} & \ddots & \vdots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \omega_{32} & \omega_{31} \\ \omega_{2T} & \omega_{2,T-1} & \cdots & \omega_{23} & \omega^2_2 & \omega_{21} \\ \omega_{1T} & \omega_{1,T-1} & \cdots & \omega_{13} & \omega_{12} & \omega^2_1 \end{pmatrix} =$$

$$\begin{pmatrix} \sigma^2_\nu + \sigma^2_\alpha & \sigma^2_\alpha & \sigma^2_\alpha & \cdots & \sigma^2_\alpha & A \\ \sigma^2_\alpha & \sigma^2_\nu + \sigma^2_\alpha & \sigma^2_\alpha & \cdots & \sigma^2_\alpha & A \\ \sigma^2_\alpha & \sigma^2_\alpha & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma^2_\alpha & A \\ \sigma^2_\alpha & \sigma^2_\alpha & \cdots & \sigma^2_\alpha & \sigma^2_\nu + \sigma^2_\alpha & A \\ A & A & \cdots & A & A & B \end{pmatrix}$$

where $A = \frac{\sigma^2_\alpha}{1-\delta}$ and $B = \frac{\sigma^2_\alpha}{(1-\delta)^2} + \frac{\sigma^2_\nu}{1-\delta^2} + \sigma^2_0$.
Without assuming normality, the optimal linear system estimator is $3SLS$

ii. If one wants to test the one-factor analytic structure, one can carry out FIML $(\delta, \beta', \gamma', \theta'_1, \cdots, \theta'_T, \zeta')'$ with an unrestricted Ω cross-equation variance-covariance, and compare the results to those of the first FIML through, say, a Likelihood Ratio statistic.

iii. If one believes that the initial condition y_{i1} is exogenous, then one applies $FIML$ (under Normality) or $3SLS$ (without Normality) on the system with $T-1$ equations in the $T-1$ endogenous variables $y_2, \cdots, y_T$ and the exogenous variables $y_1, x_{iT}, x_{i,T-1}, \cdots, x_{i1}, z_i$. To test the exogeneity of y_{i1}, one would need to carry out a *non-nested* test, since the null hypothesis that y_{i1} is exogenous implies $A = B = 0$ *simultaneously.*

(b) Applying first differencing to the original model gives:

$$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta x_{it}' \beta + \Delta \nu_{it}, \qquad i = 1, \cdots, N, \quad t = 2, \cdots, T$$

Since the error term $\Delta \nu_{it}$ is a MA(1) with known parameter -1, valid instrumental variables for the lagged dependent variable term are: $\Delta y_{i,t-m}$ and $y_{i,t-m}$ for $m \geq 2$. Note that there is a triangular structure in the set of optimal instruments, since the further along one moves in time, the greater the number of valid instruments.

NB: the regular $y_{it} - \bar{y}_{i.}$ transformation is not useful for this model since in that case, no valid instruments can be obtained by lagging the ys and Δys any number of times, because $y_{i,t-1} - \bar{y}_{i.}$ and the implied error $\nu_{i,t-1} - \bar{\nu}_{i.}$ are serially correlated with y_{is} and Δy_{is} for *every* s.

(c) Comparing methods (a) and (b) when the following additional complications are present in the linear dynamic panel data model:

i. One of the x_{it} regressors is correlated with ν_{it}:
This for example could be caused by that regressor being measured with error. Then both (a) and (b) will be inconsistent. But there the typical problem of FE-type estimators (e.g., Arellano-Bond Difference+IVE) may occur here, in that such estimators may exacerbate the inconsistency because the $\Delta(\cdot)$ (and the general FE) transformation usually reduces significantly the signal-to-noise ratio. This is because typically the true signal may be changing more slowly over time compared to the measurement error.

ii. All of the x_{it} regressors are correlated with α_i:
The FE-type estimators eliminate the α_i and hence their consistency is not affected by such correlations. But it may be preferable to apply the generalization to RE-type of estimators that we discussed in the lectures, where we model explicitly the correlation between the x regressors and the α_i terms through a model of the form: $E(\alpha_i | X, Z) = \bar{x}_{i\cdot}\xi$. Hence, entering explicitly the time-averages of the xs into the original equation:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + \bar{x}_{i\cdot}\xi + z_i'\gamma + \alpha_i^* + \nu_{it}$$

yields a new α_i^* persistent heterogeneity random effect that is uncorrelated from all regressors. Hence, the B-S RE-type approach can then proceed normally.

iii. One of the x_{it} regressors is measured with error, ξ_{it}:
See point i. above.

iv. One of the z_i regressors is measured with error, ζ_i:
Since FE-type estimators eliminate time-invariant regressors, the Difference+IVE method will not be affected by this type of errors.
Defining $z_i^j = z_i^{*j} + w_i^j$ where the true z_i^{*j} is only observed through the error w_i^j:

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + z_i^{\prime,-j}\gamma^{-j} + z_i^{\prime j}\gamma^j + \alpha_i - w^j\gamma^j + \nu_{it}$$

If we could think of an instrumental variable for z_i^j, we could modify the B-S approach by adding a reduced-form equation for the observed z_i^j in terms of all the time-invariant variables plus the additional instrumental variable for it.
Otherwise, the random-effects-type estimators, e.g., B-S would be inconsistent.

5. Consider the dynamic linear regression model for balanced data:

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + z'_i\gamma + \epsilon_{it} \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T$$

where ϵ_{it} follows the one factor error components model: $\epsilon_{it} = \alpha_i + \nu_{it}$ with α_i modelling individual unobserved persistent heterogeneity.

(a) Describe two estimation approaches for this model: the first should rely on the "Fixed Effects" principle of eliminating the unobserved persistent heterogeneity term α_i and carrying out estimation conditional on it.The second should rely on the "Random Effects" principle of deriving the (possibly optimal) estimator that considers either the full p.d.f.. or the first tow moments of the disturbance vector $(\epsilon_{11}, \cdots, \epsilon_{1T}, \cdots, \epsilon_{i1}, \cdots, \epsilon_{iT}, \cdots, \epsilon_{N1}, \cdots, \epsilon_{NT})'$, i.e., $pdf(\epsilon|X, Z)$ or $E(\epsilon|X, Z)$ and $VCov(\epsilon|X, Z)$.

Arellano-Bond estimation by applying (optimal) IVE to the first-differences model:

Arellano-Bond along First-Differencing+IVE lines using the fact that since the first-differenced model is:

$\Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta x'_{it}\beta + \Delta\nu_{it}$

in view of the non-invertible MA(1) error, the endogeneity of $\Delta y_{i,t-1}$ *can be overcome by using information from* $t = 2$ *and/or earlier and doing IVE.*

Barghava-Sargan estimation by writing panel data model as a cross-sectional system of equations, one for each time period and treating it as a Simultaneous equations system. Additional trick: complete model by adding the quasi-reduced form for the first observation as $y_{i1} = z_i\zeta + x_{i1}\xi_1 + \cdots + x_{iT}\xi_T + u_{i1}$

You should discuss the properties of the two estimation approaches under the following three scenariios about the ν_{it} error term:

i. $\nu_{it} \sim N(0, \sigma^2_\nu)$ i.i.d. over both i and t;
Classic case. Valid instruments for AB are values of $y_{i,t-q}$, $q \geq 2$. *As we move further down the time periods of a given individual, additional lags of y become valid instruments. Hence, the efficient IV/GMM estimator uses instruments with a triangular structure, resetting once a new individual is considered.*

ii. $\nu_{it} = \xi_{it} + \lambda\xi_{i,t-1}$ with $\xi_{it} \sim N(0, \sigma^2_\xi)$ i.i.d. over both i and t;
Now valid instruments for AB are values of $y_{i,t-q}$, $q \geq 3$.

iii. $\nu_{it} = \rho\nu_{i,t-1} + \xi_{it}$ with $|\rho| < 1$ and $\xi_{it} \sim N(0, \sigma^2_\xi)$ i.i.d. over both i and t.
AB will not provide consistent estimates since no valid instruments exist in this case: The variables $y_{i,t-q}$, *are correlated with the regressor* $y_{i,t-1}$ *for *any* q.*

*In all three cases (i)-(iii), the B-S approach remains valid, since the system estimation approach used (3SLS or FIML) allows for *any* valid correlation structure among the equation errors,*

*$(u_{i1}, \epsilon_{i2}, \epsilon_{i3}, \cdots, \epsilon_{it}, \cdots, \epsilon_{iT})'$. Hence in fact the particular error structures (i)-(iii) can be *tested* using classical tests (Wald, LR, LM), with the Restricted model imposing the particular correlation structure in the estimation vs. the Unrestricted model with allowing 3SLS or FIML to estimate the variance-cov structure of the errors.*

(b) Now assume the simplest $\alpha_i + \nu_{it}$ structure and consider how the two estimation approaches you described above will need to be modified to analyze the alternative models:

$$y_{it} = g(x_{it}, \beta, z_i, \gamma) + \delta y_{i,t-1} + \epsilon_{it} \qquad \text{(Model 1)}$$

where the non-linear function $g(.)$ is known up to parameter vectors β and γ;

and

$$y_{it} = h(x_{it}, \beta, z_i, \gamma, \delta y_{i,t-1}) + \epsilon_{it} \qquad \text{(Model 2)}$$

and where the non-linear function $h(.)$ is known up to parameter vectors β and γ and parameter δ.

The first model is additive in the errors, so it can be analysed completely analogously by combining RE and FE or Δ transformations with NLLS instead of OLS, or GMM in place of IV as necesary. Key thing to remember: the FE and RE operators must be applied to the non-linear function exp(.) *and *not* the non-linear function evaluated at the FE- or RE-transformed data.*

I.e., using $g((x_{it} - \lambda_i \bar{x}_{i\cdot})'\beta + (1 - \lambda_i) z_i'\gamma)$ would be wrong for RE, while we should use instead:

$g(x_{it}'\beta + z_i'\gamma) - \lambda_i g(x_{i,t-1}'\beta + z_i'\gamma)$ for the non-linear term.

Since Model 1 contains the additive dynamic term $+\delta y_{i,t-1}$ it is not appropriate to combine the usual RE or FE transformations together with NLLS to account for the presence of the g(.) term, just like the linear case where OLS to the transformed models would lead to inconsistency because of the endogeneity of all transformations of the $+\delta y_{i,t-1}$ term. For example, applying first differencing to eliminate the alpha term, gives:

$$y_{it} - y_{i,t-1} = g(x_{it}, \beta, z_i, \gamma) - g(x_{i,t-1}, \beta, z_i, \gamma) + \delta(y_{i,t-1} - y_{i,t-2}) + \nu_{it} - \nu_{i,t-1}$$

Hence, one cannot apply NLLS to this model because of the MA(1) of the resulting error term. Instead, one should use NLIV/GMM based on $y_{i,t-2}, y_{i,t-3}, \ldots$ terms as valid instrumental variables.

In Model 2, there is a very significant additional complication: the non-linearity encompasses also the $y_{i,t-1}$ part. The presence of the lagged term under the non-linear function makes this model non-additive in the error term (at least with the α_i present in all periods).

*Hence RE- or FE- plus NLLS will *not* work for this model, but we need to use instead MLE that takes into account correctly the non-trivial Jacobian of the $y \longrightarrow error$ transformation.*

*Another possibility for estimating this model consistently (though not efficiently, as is the case of MLE) would be as follows: assuming, as with MLE, that the regressors are *strongly* exogenous w.r.t. the error term, implies that lagged Xs are valid instruments for the (endogenous) lagged ys that appear as regressors. Hence NLIV/GMM could be used instead.*

(c) Finally suppose that in part (b), the δ parameter equals 0. What happens to Models 1 and 2 in such case? Discuss estimation when (i) all regressors are measured without error; and (ii) when one or more regressor(s) contain(s) errors of measurement. In such case (ii), does it make a difference whether the mismeasured regressors are among the Xs or the Zs?

If δ is 0, then Models 1 and 2 have exactly the same structure. So let us focus on Model 1.

If all the regressors are measured correctly (case (i)), then this is the classic case (easy) with the error term appearing additively outside the nonlinear function g(.). Solution: FE or RE transformations plus NLLS.

*In case (ii) however, the problem becomes essentially one with *non*additive error terms, since the measurement errors are inside the nonlinear function. It is equally hard to deal with this problem whether it is the Xs or Zs that are mismeasured — there is no simplification afforded by the within or first-differencing transformations, since the Zs are underneath the g(.) function, and hence such transformations will not eliminate the time-invariant Zs.*