Lent Term 2023

Further Topics in Econometrics
(Ec485/Ec518)

Panel Data Models

1 Static Models: Linear or Nonlinear?

The basic linear model is:

$$y_{it} = x_{it}'\beta + \epsilon_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T_i \tag{1}$$

Typically N is very large (several hundreds or even thousands) while T_i is quite small (ranging from $2 - 10$ in most cases, and very rarely exceeding 20). If $T_i = T$ for every i, the panel data set is said to be "balanced" — otherwise it is "unbalanced."

1.1 Preliminary Issues

Endogenously vs. Exogenously Unbalanced Data Sets An important issue to determine at the outset is whether a panel data set is unbalanced due to endogenous causes, i.e., causes related to the economic mechanism we are trying to model. For example, if y_{it} is earnings and the richer people are more likely to drop out of the sample as time goes by because the value of their time is higher than others', that would be a case of an endogenously unbalanced data set. In such a case, though the basic model we are trying to fit is the linear regression (1), to take correct account of the fact that in such a case the relevant expression would be the conditional expectation

$$E(y_{it}|x_{it} \text{ and individual i stays in the sample at period t})$$

we would need *non-linear* sample-selectivity methods. In other words, we would need to model the discrete mechanism characterizing the dummy variable:

$$d_{it} = \begin{cases} 1 & \text{if individual } i \text{ is in the sample in period } t \\ 0 & \text{otherwise} \end{cases}$$

and the joint linear conditional model:

$$y_{it} = x_{it}'\beta + \epsilon_{it} \quad \text{observed iff } d_{it} = 1.$$

Types of Explanatory Variables The set of explanatory variables x_{it} may include:

1. variables that vary across individuals and time periods, e.g., wage, age, and years of experience;
2. variables that are time-invariant, i.e., vary only across individuals, e.g., race and sex; and
3. variables that vary only over time but not across individuals, e.g., economy-wide unemployment, minimum-wage level, and other macroeconomic factors.

Stacking of Vectors and Matrices In matrix notation, the model can be written as:

$$y = X\beta + \epsilon$$

where y and ϵ are $\sum_{i=1}^{N} T_i \times 1$ vectors ($= NT \times 1$ for balanced data sets), and X is a $\sum_{i=1}^{N} T_i \times k$ ($NT \times k$ respectively) matrix. The convention is to stack observations in groups of all time observations for each individual, e.g.,
$y = (y_{11}, y_{12}, \cdots, y_{1T_1}, y_{21}, y_{22}, \cdots, y_{2T_2}, \cdots, y_{N1}, y_{N2}, \cdots, y_{NT_N})'$

1.2 Unobserved Persistent Heterogeneity/Error Components

The simple *one-factor error components model* is:

$$\epsilon_{it} = \alpha_i + \nu_{it}, \quad \alpha_i \sim i.i.d. \text{over i}, \quad \nu_{it} \sim i.i.d. \text{ over i and t},$$

$$\alpha_i \text{ independent of every } \nu_{jt}, \forall i, j, t,$$

with $E\alpha_i = 0$, $E\nu_{it} = 0$, $E\alpha_i^2 = \sigma_\alpha^2$, and $E\nu_{it}^2 = \sigma_\nu^2$. The presence of the time-invariant random-effect α_i implies the presence of persistent unobserved heterogeneity and the variance-covariance structure:

$$E\epsilon_{it}\epsilon_{js} = \begin{cases} \sigma_\alpha^2 + \sigma_\nu^2 & \text{for } i = j \text{ and } t = s \\ \sigma_\alpha^2 & \text{for } i = j \text{ and } t \neq s \\ 0 & \text{if } i \neq j. \end{cases}$$

Consequences Error-components structures imply violations of A4:$E\epsilon\epsilon = \sigma^2 I$ in the form of (at least) serial correlation in the error terms. Hence OLS estimation in such models will not be BLUE and will have a variance-covariance matrix not equal to $\sigma^2(X'^{-1}$.

Macro or Time Effects The *two-factor error components model* is:

$$\epsilon_{it} = \alpha_i + \delta_t + \nu_{it}, \quad \alpha_i \sim i.i.d. \text{ over i}, \quad \delta_t \sim i.i.d. \text{ over t }, \quad \nu_{it} \sim i.i.d. \text{ over i and t},$$

$$\alpha_i, \nu_{jt}, \text{ and } \delta_t \text{ mutually statistically independent } \forall i, j, t,$$

with $E\alpha_i = E\nu_{it} = E\delta_t = 0$, $E\alpha_i^2 = \sigma_\alpha^2$, $E\delta_t^2 = \sigma_\delta^2$, and $E\nu_{it}^2 = \sigma_\nu^2$. The presence of the individual-invariant δ_t random effect is meant to capture unobservable macroeconomic or time-effects.

1.3 Alternative Estimators

OLS $\hat{\beta}_{ols}$ is obtained by an OLS regression of y_{it} on x_{it}. The true error from this model is $\epsilon_{it} = \alpha_i + \nu_{it}$, which is serially correlated within each individual block of observations. Hence, this estimator will not be BLUE since A4 is violated.

Fixed Effects or Within This is obtained by either (a) a regression of y_{it} on x_{it} and a set of N dummy variables for each individual or, equivalently (proved using partitioned regression theory) (b) a regression of $y_{it} - \bar{y}_{i\cdot}$ on $x_{it} - \bar{x}_{i\cdot}$ to get $\hat{\beta}_{fe}$ followed by the calculations $\hat{\alpha}_{i,fe} = \bar{y}_{i\cdot} - \bar{x}'_{i\cdot}\hat{\beta}_{fe}$.

The true error term corresponding to this regression is $\epsilon_{it} - \bar{\epsilon}_{i\cdot} = \nu_{it} - \bar{\nu}_{i\cdot}$.

Assuming A1-A3 hold, $\hat{\beta}_{fe}$ is unbiased and consistent as long as $NT \to \infty$, while the $\hat{\alpha}_i$'s are unbiased but do not converge to the true α_i's *unless* $T_i \to \infty$ for every i (since they are each based only on the T_i observations of individual i).

Random Effects or GLS This is obtained by an OLS regression of $y_{it} - \lambda_i \bar{y}_{i\cdot}$ on $x_{it} - \lambda_i \bar{x}_{i\cdot}$, where $\lambda_i = 1 - \sqrt{\frac{\sigma_\nu^2}{T_i\sigma_\alpha^2 + \sigma_\nu^2}}$. In practice, we have to work with the feasible GLS estimator that uses a consistent estimator $\hat{\lambda}_i$ for λ_i based on consistent estimators for σ_α^2 and σ_ν^2. Assuming A1-A3 are satisfied and the error has indeed a one-factor RE structure, this estimator will be asymptotically (as $NT \to \infty$) the best linear consistent asymptotically normal estimator.

The true error term corresponding to this GLS regression is:
$\epsilon_{it} - \lambda_i\bar{\epsilon}_{i\cdot} = (1 - \lambda_i)\alpha_i + \nu_{it} - \lambda_i\bar{\nu}_{i\cdot}$. This error satisfies A4, i.e., it is homoskedastic and serially uncorrelated.

Between This investigates the variability *between* individuals and is calculated through an OLS regression of $\bar{y}_{i\cdot}$ on $\bar{x}_{i\cdot}$. The true error of this equation is $\bar{\epsilon}_{i\cdot} = \alpha_i + \bar{\nu}_{i\cdot}$.

The main attractive feature of this approach is to allow us to obtain a consistent estimator for σ_α^2 — see below.

1.4 Comparisons of Methods – Fixed vs. Random Effects

The Case of $E(\alpha_i|x_{it}) = 0$ Among the four estimators defined above, the best linear estimator asymptotically as $NT \to \infty$ is the feasible GLS/RE one, since the implied error of the GLS regression satisfies A4.

The Case of $E(\alpha_i|x_{it}) \neq 0$ This implies that A3R3 is violated because of the correlation between α_i and the regressors. Hence, all methods with the exception of FE/Within will be inconsistent. Since $\hat{\beta}_{fe}$ is obtained having eliminated the α_i's through the FE transformation, the consistency of this estimator only requires $NT \to \infty$ irrespective of any possible correlation between α's and x's. Thus in such a case, FE is the only available estimator that is consistent. Some notes follow:

1. There is no need to carry out GLS as opposed to plain OLS estimation of the FE-transformed model (i.e., $y_{it} - \bar{y}_{i\cdot}$ on $x_{it} - \bar{x}_{i\cdot}$) to "correct" the induced serial correlation in the resulting error $\nu_{it} - \bar{\nu}_{i\cdot}$. This is because this error has an idempotent covariance matrix, which implies that GLS and OLS will give numerically the same answer in such a case.

2. Measurement errors in x_{it}'s may be exacerbated through the FE transformation in case the measurement error fluctuates substantially over time, while the explanatory variables do not. This is because in such a case the FE transformation will result in regressors that contain a lot of noise and very little signal.

3. The fixed-effects transformation will eliminate from the regression any variables that are time-invariant (just like it does to the α's). In many situations in practice, such variables are the ones of primary focus. Alternative methods that combine Instrumental Variable Estimation with the Fixed-Effects transformation can overcome this problem.

1.5 First-Differences and Within/FE

Carrying out OLS on the first-differenced model:

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta\epsilon_{it}$$

also provides an estimator for β that does not rely on the properties of the α_i's, since the Δ transformation eliminates them. This approach, however, will generally be dominated by the Fixed-Effects approach since the latter, in addition to eliminating the α's, also implies error terms with an idempotent covariance matrix, which eliminates any need for carrying out GLS estimation in preference to OLS.

In contrast to this, the error $\Delta\epsilon_{it}$ is a Moving-Average(1) error process with MA parameter equal to -1; therefore, it has a (non-singular) variance-covariance structure which would require either GLS estimation or at least correcting the standard errors of the OLS estimator applied to the Δ model.

The only exception occurs with a balanced set with $T = 2$ for all individuals. In that case, the Δ model and the Fixed-Effects model are *numerically* identical models, since $y_{i2} - y_{i1} = y_{i2} - \frac{1}{2}(y_{i1} + y_{i2})$.

1.6 Estimation of σ^2_α and σ^2_ϵ

Assuming a balanced sample,[1] the OLS s^2_{fe} obtained from the Fixed-effects regression is a consistent estimator for the variance of $\nu_{it} - \bar{\nu}_{i\cdot}$, which is $\sigma^2_\nu + \frac{1}{T}\sigma^2_\nu - 2\frac{1}{T}\sigma^2_\nu = \frac{T-1}{T}\sigma^2_\nu$. Hence, a consistent estimator for σ^2_ν is: $\hat{\sigma}^2_\nu \equiv \frac{T}{T-1}s^2_{fe}$.

Note that practically the same answer (for large NT) can be obtained by recognizing that the fixed-effect regression has $NT - N - k$ degrees of freedom instead of the $NT - k$ assumed by the OLS package. Hence, a consistent (and unbiased) estimator for σ^2_ν is: $\frac{NT-k}{N(T-1)-k}s^2_{fe}$.

A consistent estimator for σ^2_α can then be obtained using the $s^2_{ols,B}$ from the Between regression by recognizing that the Between regression error term $\bar{\epsilon}_{i\cdot} = \alpha_i + \bar{\nu}_{i\cdot}$ has variance $\sigma^2_\alpha + \frac{1}{T}\sigma^2_\nu$. Hence, we define $\hat{\sigma}^2_\alpha \equiv \sigma^2_B - \frac{1}{T}\hat{\sigma}^2_\nu$. (Note that this estimate might turn out to be negative, since it is obtained from two separate regressions. In such a case, one should conclude that the best guess for σ^2_α is then 0, i.e., the error term is a pure i.i.d. innovation satisfying A4.)

1.7 Wu-Hausman Tests

Suppose we have two alternative estimators, $\hat{\theta}_I$ and $\hat{\theta}_{II}$, for a true parameter vector θ. Further suppose that if a particular hypothesis H_0 is correct, both estimators are consistent and asymptotically normal with variance-covariance matrices V_I and V_{II}, and matrix of covariances between the two estimators $V_{I,II}$. Finally suppose that if the null hypothesis is false the two estimators converge to different answers — for example, one of them might remain consistent while the other one becomes inconsistent, or both of them might become inconsistent but idiosyncratically so. Then the Wu-Hausman quadratic form:

$$Q = (\hat{\theta}_I - \hat{\theta}_{II})'(V_I + V_{II} - V_{I,II} - V_{II,I})^{-1}(\hat{\theta}_I - \hat{\theta}_{II})$$

[1] If it is not, we can artificially balance it solely for the purposes of obtaining the σ estimates by working only with $\min_i T_i$ for each individual. Of course, estimation of the β's should use the complete, unbalanced data set.

under H_0 converges in distribution to a $\chi^2(k)$, where k is the number of elements in θ.

In the case that one of the estimators, say $\hat{\beta}_I$, is efficient under H_0, from the Rao-Blackwell theorem it follows that $V_{I,II} = V_{II,I} = V_I$. Hence, the variance-covariance expression in the middle of Q simplifies to $V_{II} - V_I$.

Applying this approach to the linear panel data problem, we can use the Q statistic based on any pair of estimators for β out of $\hat{\beta}_{fe}$, $\hat{\beta}_{gls}$, and $\hat{\beta}_B$. For large NT, the three alternative Q statistics can be shown to be *numerically* equivalent, in small samples differing only due to imperfections in estimating σ^2_ν and σ^2_α. Given that $\hat{\beta}_{gls}$ is efficient under
H_0 : α_i and ν_{it} mean-independent of x_{it}
and that $V_{B,FE} = 0$ (since the variability in $y_{it} - \bar{y}_{i\cdot}$ is orthogonal to the variability in $\bar{y}_{i\cdot}$), it follows that:

$$\begin{aligned} V(\hat{\beta}_B - \hat{\beta}_{fe}) &= V(\hat{\beta}_B) + V(\hat{\beta}_{fe}), \\ V(\hat{\beta}_B - \hat{\beta}_{gls}) &= V(\hat{\beta}_B) - \\ V(\hat{\beta}_{gls}), \quad V(\hat{\beta}_{gls} - \hat{\beta}_{fe}) &= V(\hat{\beta}_f e) - V(\hat{\beta}_{gls}). \end{aligned}$$

1.8 Estimation with Time Effects

The two-factor error components model:

$$\epsilon_{it} = \alpha_i + \delta_t + \nu_{it}$$

can be analyzed in one of two ways:
(1) We can treat both α and δ as random effects and work with the GLS transformation:

$$y_{it} - \lambda_i \bar{y}_{i\cdot} - \theta_t \bar{y}_{\cdot i} - \zeta y_{\cdot\cdot}$$

where λ, θ, and ζ are functions of T_i, N, σ^2_α, σ^2_δ, and σ^2_ν. Alternatively:
(2) We can treat α_i as a random effect while the δ_t's as fixed effects to be estimated using a dummy variable for each time-period. This approach is very straightforward and has the additional advantage that few degrees of freedom are lost by the fixed-effects assumption for the δ's, since there are only T of them. **The fixed-δ_t, random-α_i approach is particularly useful in non-linear panel data models.**

2 Linear Dynamic Models

The basic model (with a single lagged dependent variable) is:

$$y_{it} = \theta y_{i,t-1} + x'_{it}\beta + \epsilon_{it} \tag{2}$$

The main issues can be illustrated with the simple one-factor error-components model
$\epsilon_{it} = \alpha_i + \epsilon_{it}$.

2.1 Consequences for OLS, RE/GLS, Between, and FE/W Estimators

In the dynamic case, *all* four estimation approaches we have discussed above suffer from endogeneity or a violation of A3R3, which leads to inconsistency. This is because:

1. The regressor $y_{i,t-1}$, which depends on the error $\epsilon_{i,t-1} = \alpha_i + \nu_{i,t-1}$, is correlated with the current error $\epsilon_{it} = \alpha_i + \nu_{i,t}$ because of the presence of α_i in both. Hence OLS will be inconsistent.

2. The regressor $y_{i,t-1} - \hat{\lambda}\bar{y}_{i\cdot}$, which depends on the error $\epsilon_{i,t-1} - \hat{\lambda}\bar{\epsilon}_{i\cdot} = (1-\hat{\lambda})\alpha_i + \nu_{i,t-1} - \hat{\lambda}\bar{\nu}_{i\cdot}$, is correlated with the current error $\epsilon_{i,t} - \hat{\lambda}\bar{\epsilon}_{i\cdot} = (1-\hat{\lambda})\alpha_i + \nu_{it} - \hat{\lambda}\bar{\nu}_{i\cdot}$, because of the presence of α_i in both, and because $E\nu_{is}\bar{\nu}_{i\cdot} \neq 0$ and $E\bar{\nu}_{i\cdot}^2 \neq 0$ (unless if $T_i \to \infty$). Hence GLS will be inconsistent.

3. The between estimator is not useful in this case because $\bar{y}_{i\cdot}$ would appear both as a dependent variable and as a regressor.

4. The regressor $y_{i,t-1} - \bar{y}_{i\cdot}$, which depends on the error $\epsilon_{i,t-1} - \bar{\epsilon}_{i\cdot} = \nu_{i,t-1} - \bar{\nu}_{i\cdot}$, is correlated with the current error $\epsilon_{it} - \bar{\epsilon}_{i\cdot} = \nu_{it} - \bar{\nu}_{i\cdot}$ because of the serial correlation in $\nu_{it} - \bar{\nu}_{i\cdot}$. Hence Fixed Effects will also be inconsistent unless $T_i \to \infty$ for all i, in which case the transformed error becomes just ν_{it} since $\bar{\nu}_{i\cdot}$ will converge in probability to $E\nu_{it} = 0$.

2.2 Treatment of Initial Conditions

If we consider all observations $i = 1, \cdots, N$, $t = 1, \cdots, T_i$, we see that given the lagged dependent variable that appears as a regressor we are missing one initial condition y_{i0} for each of the N individuals. Given this, we can drop all $t = 1$ observations and work instead with $t = 2, \cdots, T_i$. This treatment still leaves open the question as to how to hadnle the endogeneity of the lagged dependent variables appearing as regressors.

2.3 MLE/IVE + RE Approaches

Assuming for simplicity we have a balanced panel data set, we can view it as a cross-sectional data set with N observations on a system of $T-1$ equations:

$$\begin{aligned} y_{iT} &= \theta y_{i,T-1} + x'_{iT}\beta + \epsilon_{iT} \\ y_{i,T-1} &= \theta y_{i,T-2} + x'_{i,T-1}\beta + \epsilon_{i,T-1} \\ &\cdots \\ y_{i2} &= \theta y_{i1} + x'_{i2}\beta + \epsilon_{i2} \end{aligned} \tag{3}$$

This system exhibits (1) cross-equation restrictions (θ and β constant throughout); (2) correlations across the errors of different equations at a given i (as implied by, say, the one-factor error structure assumed for ϵ_{it}); and (3) common parameter restrictions between the RHS expressions and the variance-covariance matrix of the ϵ's.

This system is not complete because it contains T y's $(y_{iT}, \cdots, y_{i1})$ but has only $T-1$ equations to determine them. To correct this we can adopt one of two approaches:
Approach 1: we assume that y_{i1} is predetermined/exogenous for each i and hence think of this as a system of $T-1$ equations determining the $T-1$ endogenous variables $y_{iT}, \cdots, y_{i2}$, where the set of exogenous variables are $x_{iT}, \cdots, x_{i2}$ and y_{i1}.
Approach 2: we assume that y_{i1} is an endogenous variable and add a reduced-form equation to determine it without introducing earlier exogenous variables, by writing:

$$y_{i1} = x_{i1}\zeta_1 + \cdots + x_{iT}\zeta_T + u_{i1} \tag{4}$$

This is not a behavioural equation since it includes x's not available at time 1 but an equation describing the best linear way of explaining y_{i1} in terms of all available exogenous variables in the system. Thus we get a system of T equations determining the T endogenous variables $y_{iT}, \cdots, y_{i1}$, where the set of exogenous variables is $x_{iT}, \cdots, x_{i2}$. The error u_{i1} also contains the error we commit in trying to explain y_{i1} not through the true mechanism that generated it but through the linear prediction given by (4).

Once we adopt either approach, we can apply suitable MLE and/or IVE estimation of the system of equations like Full Information MLE and Three-Stage-

Least-Squares.[2]

2.4 First Differencing + IVE Approaches

We have seen that the Fixed Effects transformation will not overcome the endogeneity of the lagged dependent variable with respect to the error term because even though it eliminates the time-invariant α_i, it induces a serial correlation structure in the ν's in the form of $\nu_{it} - \bar{\nu}_{i\cdot}$. But we have seen earlier that an alternative way of eliminating the α_i is to use the first-differencing transformation Δ to obtain:

$$\Delta y_{it} = \theta \Delta y_{i,t-1} + \Delta x_{it}' \beta + \Delta \nu_{it} \tag{5}$$

This equation has a MA(1) error that does not contain α_i. Given this structure, the regressor $\Delta y_{i,t-1}$ is correlated with the error since it depends on $\Delta \nu_{i,t-1}$. But earlier lags $\Delta y_{i,t-\ell}$, $\ell > 1$ are good instrumental variables in that they are correlated with $\Delta y_{i,t-1}$ but are not correlated with the MA(1) error $\Delta \nu_{it}$. Similarly earlier lags of y levels, $y_{i,t-\ell}$, $\ell > 1$ are also valid instruments.

Hence, an appropriate estimation approach in this context is Instrumental Variables estimation that pays proper attention to the MA(1) nature of the $\Delta \nu_{it}$ error term. An important point made by Arellano and Bond is that to improve the efficiency of the IV estimator one should recognize that the list of valid instruments grows for later observations for a given individual, since additional lags become available as valid instruments.

To understand the properties of the "First Differencing + IVE Approach," it is useful to consider the following three scenarios about the ν_{it} error term:

1. $\nu_{it} \sim i.i.d.$ over both i and t;

   This is the classic case analyzed by Arellano and Bond. 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.

2. $\nu_{it} = \xi_{it} + \lambda \xi_{i,t-1}$ with $\xi_{it} \sim i.i.d.$ i.i.d. over both i and t;

   This is an easy generalization: now valid instruments for AB are values of $y_{i,t-q}$, $q \geq 3$.

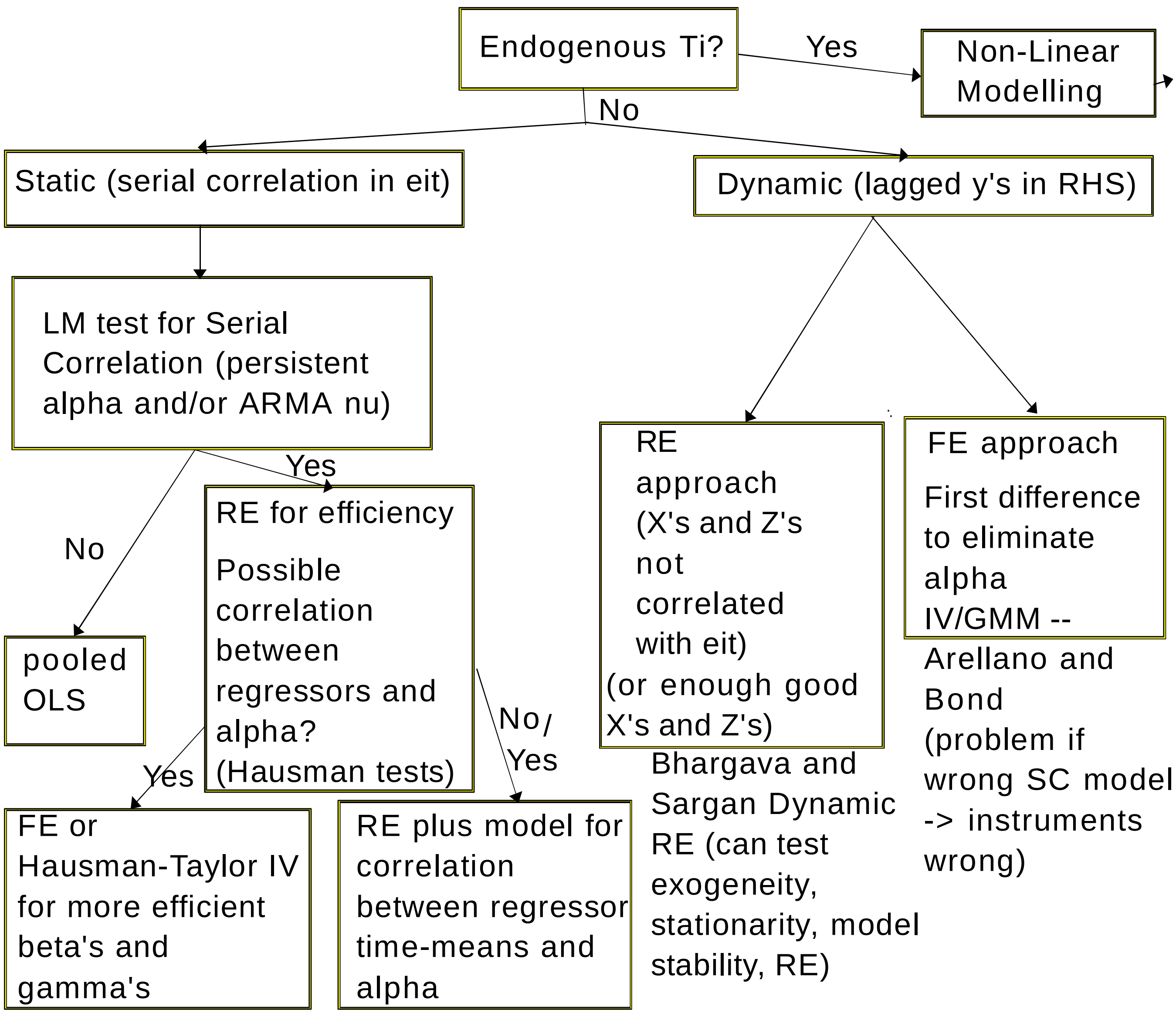
[2] The seminal work here is by Barghava and Sargan, who show how to develop useful tests using this approach of several interesting hypotheses, namely (a) the random effects assumption; (b) stationarity of the error terms; (c) correct specification of the model in terms of stability over time; and (d) possible exogeneity of the initial conditions.

3. $\nu_{it} = \rho \nu_{i,t-1} + \xi_{it}$ with $|\rho| < 1$ and $\xi_{it} \sim i.i.d.$ over both i and t.

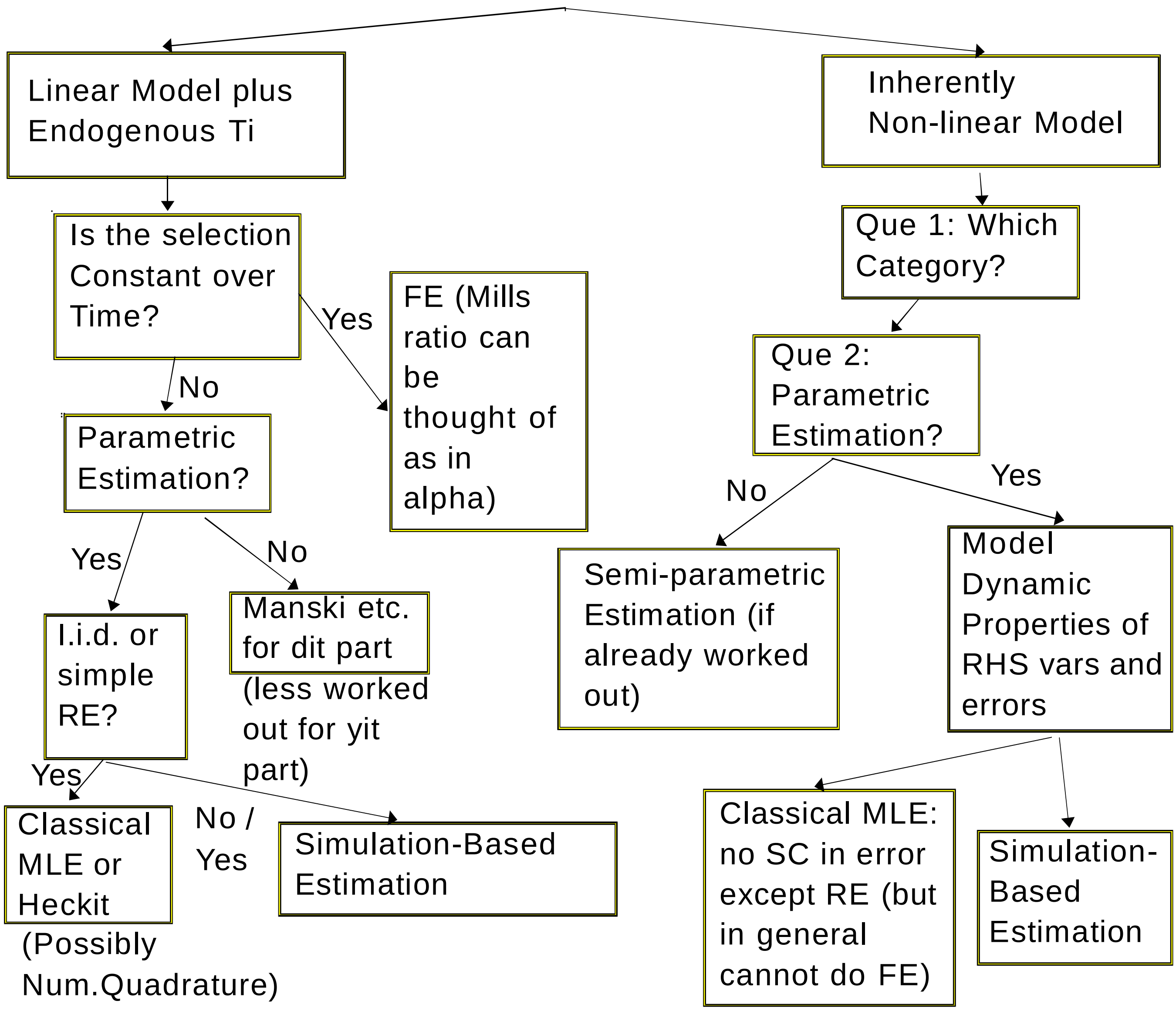
   This case renders the AB approach inconsistent in general, because all earlier lags of the dependent variable y are no longer valid instruments: The variables $y_{i,t-q}$, are correlated with the regressor $y_{i,t-1}$ for *any* q. IVE may be salvaged in this case, provided one uses suitable lags of the *explanatory* variables as instruments. Doing so, however, is not in the spirit of the original AB approach.

Panel Data Estimation: A Road Map

Part I: Linear Panel Data Models



Part II: Non-Linear Panel Data Models



2.5 Dynamic Nonlinear TSCS Models

Consider the dynamic linear regression model for balanced data:

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + z'_i\gamma + \epsilon_{it} \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T \tag{6}$$

where ϵ_{it} follows the one factor error components model: $\epsilon_{it} = \alpha_i + \nu_{it}$ with α_i modelling individual unobserved persistent heterogeneity. Two analogous nonlinear models can be considered that are additive in the error ϵ_{it}:

$$y_{it} = \delta y_{i,t-1} + g(x'_{it}\beta + z'_i\gamma) + \epsilon_{it} \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T \tag{7}$$

and

$$y_{it} = g(\delta y_{i,t-1} + x'_{it}\beta + z'_i\gamma) + \epsilon_{it} \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T \tag{8}$$

A more general model that is non-additive in the error term is:

$$y_{it} = h(\delta y_{i,t-1} + x'_{it}\beta + z'_i\gamma + \epsilon_{it}) \quad , \quad i = 1, \cdots, N \ , \quad t = 1, \cdots, T \tag{9}$$

It is important to realize two facts: first, the nonlinearity of model (7) is quite straightforward to analyze, it being a benign generalization of the linear dynamic model (6). This is because the lagged term $\delta y_{i,t-1} = \cdots + \epsilon_{i,t-1}$ resulting in a model that is additive in both current as well as past error terms.

The second fact is that the nonlinearity in model (8) is not straighforward, making the model closer in difficulty to the nonadditive nonlinear model (9). As explained already, the lagged term $\delta y_{i,t-1} = \cdots + \epsilon_{i,t-1}$. Consequently, the third model is additive in the current error but non-additive in earlier error(s), making it similar to the fourth model which is non-additive in the current error.

Given these two facts, we proceed by starting with the non-dynamic case, i.e., by setting $\delta = 0$. Since models (7) and (7) are now equivalent and have no lagged terms, they are additive in the error term. Hence, the models can be analyzed *by combining RE and FE or Δ transformations with NLLS instead of OLS, or GMM in place of IV as necessary.* The key thing to remember is that the FE and RE operators must be applied to the non-linear function $g(.)$ and *not* the nonlinear 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.

We now consider the dynamic models for $\delta \neq 0$. The simpler model (7) is additive in both the current as well as past error terms. We must be careful, however, because of the presence of 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.

Proceeding to Model (8), we now encounter a very significant additional complication: the non-linearity now 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.

As noted already, the final Model (9) is also non-additive in the (current) error term. In a similar vein, we need MLE because NLLS+RE or +FE will not work.

Finally, please note that another possibility exists for estimating consistently the "non-additive in error(s)" Models (8) and (9) 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. Such methods would be consistent but not fully efficient — efficiency is reserved for full MLE.