TOPIC 1: LINEAR PANEL DATA MODELS

1. Panel Data

2. Static Panel Data Models

   2.1. Assumptions

   2.2. "Fixed effects" and "Random effects"

   2.3. Estimation with exogenous individual effects
2.4. Estimation with endogenous individual effects

2.5. Test of exogeneity of individual effects

2.6. Standard errors

2.7. Within-Groups with measurement error

2.8. Applications

3. Dynamic Panel Data Models

3.1. Introduction

3.2. Inconsistency of OLS-FD

3.3. Inconsistency of WG
3.4. Anderson-Hsiao (JE, 1982)

3.5. Arellano-Bond (REStud, 1991)


3.7. Specification tests
1. PANEL DATA

- We have panel data when we observe a group of individuals (e.g., households, firms, countries) over several periods of time.

\[ \text{Data} = (x_{it} : i = 1, 2, \ldots, N; \quad t \in \{1, 2, \ldots, T\}) \]

\[ x_{it} : \quad K \times 1 \text{ vector of observed variables} \]

\[ i : \quad \text{subindex for individuals} \]
\[ t : \quad \text{subindex for time} \]

\[ N = \quad \# \text{ of individuals} \]
\[ T_i = \quad \# \text{ of periods of time the individual } i \text{ is observed} \]
• Balanced Panels: When $T_i = T$ for every $i$.

• Unbalanced Panels: When $T_i$ can be different each $i$.

• Unbalanced panels can be such that individuals have different starting periods.

\[
\begin{align*}
  i = 1 & \quad \vdash - - - - - - - - \\
  i = 2 & \quad \vdash - - - - - \\
  i = 3 & \quad \vdash - - - - - - - - \\
  i = 4 & \quad \vdash - - - - - - - - - - - - - - - - \\
\end{align*}
\]

• Extension to more than 2 dimensions, e.g., firms over space and time; big data.
• **Remark 1:**
  - We consider panels with large $N$ and $T_i$ relatively small (asymptotics as $N \to \infty$). This is typically the case in micro-panels of households, individuals, or firms.
  - For static panel data models (i.e., strictly exogenous regressors) all the results that we present in these notes apply to the reverse case with $N$ small and $T$ large (but swapping $T$ and $N$ in the asymptotic formulas).

• **Remark 2:**
  - A panel dataset is much more than a sequence of cross-sections over time. In a sequence of cross-sections we do not observe the same individuals over time. In panel data, we follow individuals for several periods of time.
  - This important feature makes it possible to control for (potentially endogenous) unobserved individual heterogeneity.
Panel data (PD) have several advantages relative to cross-sectional (CS) and time-series (TS) data to control for spurious correlations due to omitted / unobservable variables.

Two sources of sample variation (time and cross-sectional) we can control for the potential endogeneity bias induced by unobservables that vary only over time but not over individuals (aggregate time effects) or that vary over individuals but are constant over time (time-invariant individual effects).
Example 1. Production function

- CS data from $N$ firms producing the same product. Consider the Cobb-Douglas production in logarithms:

$$y_i = \beta_0 + \beta_1 l_i + \beta_2 k_i + \varepsilon_i$$

- $\varepsilon_i$ represents the amount of unobserved inputs, e.g., manager ability, land quality, rainfall, etc.

- Unobserved inputs can be correlated with labor and capital: we expect $\text{cov}(l_i, \varepsilon_i) > 0$ and $\text{cov}(k_i, \varepsilon_i) > 0$.

- OLS estimation will provide inconsistent estimates of $\beta_1$ and $\beta_2$. 
Example 1 (cont.)

Suppose that we have PD such that:

\[ y_{it} = \beta_0 + \beta_1 l_{it} + \beta_2 k_{it} + \varepsilon_{it} \]

And we assume that:

\[ \varepsilon_{it} = \alpha_i + \gamma_t + u_{it} \]

where:

- \( \alpha_i \) represents time-invariant unobserved inputs, e.g., quality of land;
- \( \gamma_t \) common shocks for all the firms;
- \( u_{it} \) is a firm-specific productivity shock.
• Example 1 (cont.)

• Under some conditions we can use PD to control for endogeneity bias generated by the unobservables $\alpha_i$ and $\gamma_t$.

• The model in first differences is:

$$\Delta y_{it} = \beta_1 \Delta l_{it} + \beta_2 \Delta k_{it} + \Delta \gamma_t + \Delta u_{it}$$

• If regressors are not correlated with $\Delta u_{it}$ (e.g., rainfall), the OLS estimator of the equation in first differences is consistent.

• Using PD we can also obtain instruments [exploit Granger causality] for the consistent estimation of the model even when $(\Delta l_{it}, \Delta k_{it})$ are correlated with $\Delta u_{it}$. 
• **Example 2 (Dynamic labor demand).**

• Sargent (JPE, 1978) estimates the following dynamic model of labor demand (Euler equation) for a representative firm.

\[
\Delta L_t = \beta_0 + \beta_1 \Delta L_{t-1} + \beta_2 W_t + \beta_3 \frac{Y_t}{L_t} + \gamma_t + u_t
\]

where:

• \(\Delta L_t\) is employment change; \(W_t\) is the wage rate; \(Y_t\) is output;

• \(u_t\) is an expectational error orthogonal to information at \(t\) or before;

• \(\gamma_t\) is a component of the marginal profit unobserved to the researcher but observable to the firm;
Example 2 (cont.)

- The error term in this model is: \( \varepsilon_t = \gamma_t + u_t \). While \( u_t \) is not correlated with the regressors, the error \( \gamma_t \) could be. OLS estimation will be inconsistent.

- Suppose we have firm level PD \( \{L_{it}, W_{it}, Y_{it}\} \). And suppose that we assume that

\[
\varepsilon_{it} = \alpha_i + \gamma_t + u_{it}
\]

where \( u_{it} \) is not serially correlated.

- The Euler equation in differences:

\[
\Delta^2 L_{it} = \beta_1 \Delta^2 L_{it-1} + \beta_2 \Delta W_{it} + \beta_3 \Delta \frac{Y_{it}}{L_{it}} + \Delta \gamma_t + \Delta u_{it}
\]

We can estimate consistently \( \beta' \)'s using an IV/GMM method.
An General PD Model

• Consider the model:

\[ y_{it} = f(x_{it}, \varepsilon_{it}, \beta) \]

where:

- \( f(.) \) is a known function;
- \( \beta \) is a vector of unknown parameters.
- \( x_{it} \) is a vector of observable explanatory variables;
- \( \varepsilon_{it} \) is unobservable;

• A standard specification of the unobservable is:

\[ \varepsilon_{it} = \alpha_i + \gamma_t + u_{it} \]
INCIDENTAL PARAMETERS PROBLEM

- In PD model with an error structure \( \varepsilon_{it} = \alpha_i + \gamma_t + u_{it} \), we can treat \( \alpha \)'s and \( \gamma \)'s as parameters to estimate.

- We can include as explanatory variables \((N - 1)\) individual dummies associated to \( \alpha \)'s and \((T - 1)\) time dummies associated to \( \gamma \)'s.

- The number of parameters \( \alpha \)'s increases at the same rate as \( N \) as \( N \to \infty \). We cannot estimate these parameters consistently.

- Inconsistency of our estimator of \( \alpha \)'s may contaminate / bias our estimation of the parameters of interest \( \beta \).
INCIDENTAL PARAMETERS PROBLEM [more generally]

- Given a model with vector of parameters \( \theta \) and a random sample with size \( n \), we say that the model has an *incidental parameters problem* if:

  \[
  \text{As } n \to \infty, \text{ we have that } \dim(\theta) \to \infty
  \]

- Neyman and Scott (Econometrica, 1948) show that in general, there is not a consistent estimator of the whole vector \( \theta \).

- Suppose that \( \theta = (\alpha, \beta) \), such that as \( n \to \infty \),

  \[
  \dim(\alpha) \to \infty, \text{ but } \dim(\beta) = K \text{ for any value of } n
  \]

Can we obtain a consistent estimator of \( \beta \) despite our estimation of \( \alpha \) will be always inconsistent?
Two relevant classifications of PD Models

- There are two classifications of PD models that have important implications on the properties of different estimators (e.g., fixed effects estimators).

- (A) Classification according to the additivity of the unobservables:
  
  Additive separable unobservable: \[ y_{it} = h(x_{it}, \beta) + \varepsilon_{it} \]

  Non additive separable unobservable: e.g., \[ y_{it} = \max\{x_{it}' \beta + \varepsilon_{it}, 0\} \]

- (B) Classification according to Static and Dynamic PD models.
Additive vs. Nonadditive Unobservables

- The Incidental Parameters Problem has a very different nature depending on whether unobservables are additive or non-additive models.

- The properties of some estimators are very different in linear and in nonlinear PD models. For instance, the fixed-effects or within-groups estimator is consistent in static linear PD models but it is inconsistent in most nonlinear static PD models.
Static and Dynamic Panel Data Models

- The classification Static vs Dynamic PD models depends on the relationship between the observable variables \( \{x_{it}\} \) and the transitory shock \( \{u_{it}\} \).

- In a **static model** regressors are "strictly exogenous" in the sense that do not include lagged endogenous variables:

\[ E(u_{it} x_{it+s}) = 0 \quad \text{for any } s \in \{-2, -1, 0, +1, +2, \ldots\} \]

- **Dynamic models** include predetermined or lagged endogenous variables. For instance, \( y_{i,t-1} \) is included in \( x_{it} \). In these models, \( x_{it+1} \) includes \( y_{it} \), and therefore

\[ E(u_{it} x_{it+1}) \neq 0 \]
The properties of some estimators are very different in static and dynamic PD models. Estimators which are consistent and efficient in static models, are not even consistent in dynamic models.
2. STATIC PANEL DATA MODELS

- Consider the model:

\[ y_{it} = x_{it}' \beta + \varepsilon_{it} \]

where: \( x_{it} \) is a \( K \times 1 \) of regressors. \( \varepsilon_{it} \) is the error term.

- Most PD models assume the following structure for the error term:

\[ \varepsilon_{it} = \alpha_i + \gamma_t + u_{it} \]

\( \gamma_t \) is a common aggregate effect;

\( \alpha_i \) represents persistent unobserved heterogeneity is called "individual effect" or "unobserved heterogeneity"

\( u_{it} \) is called "transitory shock" or "time-variant unobservable", ...
Time dummies

- We can incorporate parameters $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_T\}$ using time dummies.

- Let $D_t^{(s)}$ be the time-dummy for period $s$ such that:
  $$D_{it}^{(s)} = 1 \text{ if } t = s; \quad \text{and} \quad D_{it}^{(s)} = 0 \text{ if } t \neq s$$

- Then,
  $$\gamma_t = \sum_{s=1}^{T} D_{it}^{(s)} \gamma_s = [D_{it}^{(1)}, \ldots, D_{it}^{(T)}] \gamma = D_{it}' \gamma$$

- Therefore, the model:
  $$y_{it} = [x_{it}, D_{it}]' \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \alpha_i + u_{it}$$
Time dummies  (cont.)

- For notational simplicity, unless state otherwise, I represent the model as:

\[ y_{it} = x'_{it} \beta + \alpha_i + u_{it} \]

But it should be understood that \( x_{it} \) includes the time-dummy variables, and \( \beta \) includes the time-effect parameters \( \gamma \).

- This notation applies to large \( N \) and \( T \) small. In this case, \( \gamma \) and \( \beta \) are not subject to the incidental parameters problem [in contrast to \( \alpha \)].

- When \( T \) is large and \( N \) is small we can include individual dummies in \( x_{it} \). In that case, \( \alpha \) and \( \beta \) are not subject to the incidental parameters problem [in contrast to \( \gamma \)].
More notation ...

- We can see a PD model as a system of $T$ equations (an eq. for $t = 1$, for $t = 2$, ...) such that we have a CS of $N$ observations for each of these $T$ equations.

$$
y_{i1} = x'_{i1} \beta + \alpha_i + u_{i1}
$$
$$
y_{i2} = x'_{i2} \beta + \alpha_i + u_{i2}
$$
$$
\vdots 
$$
$$
y_{iT} = x'_{iT} \beta + \alpha_i + u_{iT}
$$

- We can represent this system in vector form as:

$$
Y_i = X_i \beta + 1 \alpha_i + U_i
$$

where:

$Y_i$ and $U_i$ are $T \times 1$ vectors; and $1$ is a $T \times 1$ vector of $1's$;

$X_i$ is $T \times K$ matrix.
• For **unbalanced panels** we can use a similar notation.

• Let $d_{it} \in \{0, 1\}$ be the indicator of the event "individual $i$ is observed in the cross-section at period $t$". Then:

$$Y_i = X_i \beta + d_i \alpha_i + U_i$$

where:

$$Y_i = \begin{bmatrix} d_{i1} & y_{i1} \\ d_{i2} & y_{i2} \\ \vdots \\ d_{iT} & y_{iT} \end{bmatrix}_{T \times 1} ; \quad X_i = \begin{bmatrix} d_{i1} x'_{i1} \\ d_{i2} x'_{i2} \\ \vdots \\ d_{iT} x'_{iT} \end{bmatrix}_{T \times K} ; \quad d_i = \begin{bmatrix} d_{i1} \\ d_{i2} \\ \vdots \\ d_{iT} \end{bmatrix}_{T \times 1}$$

The expressions of the different estimators we will see below apply to this definition of $Y_i$, $X_i$, and $d_i$. 
• **Assumption:** Strict exogeneity (Static Model)

\[ \mathbb{E}[x_{it} u_{is}] = 0 \quad \text{for any } (t, s) \in \{1, \ldots, T\} \]

• **Comment:** This assumption does not hold for models where the set of regressors includes predetermined endogenous variables: e.g., \( y_{i,t-1}, y_{i,t-2} \).

• For instance, suppose that \( y_{i,t-1} \subset x_{it} \). The model establishes that \( y_{i,t-1} \) depends on \( u_{i,t-1} \). Therefore, it is clear that:

\[ \mathbb{E}[x_{it} u_{i,t-1}] \neq 0 \]
FIXED EFFECTS vs RANDOM EFFECTS

- One of the most important issues in the estimation of panel data models is endogeneity due to correlation between the regressors and the individual effect.

\[ \mathbb{E}(x_{it} \alpha_i) \neq 0 \]

- There are two main approaches to control for this endogeneity problem:

  1. the **Fixed effects** approach;

  2. the **Correlated Random Effects** approach.
Fixed effects (FE) models / methods

- This approach does not impose any restriction on the joint distribution of \((x_{i1}, x_{i2}, ..., x_{iT})\) and \(\alpha_i\).

- \(CDF(\alpha_i \mid x_{i1}, x_{i2}, ..., x_{iT})\) is completely unrestricted. In this sense, the FE model is nonparametric with respect the distribution \(CDF(\alpha_i \mid x_i)\).

- Typically, fixed effects methods are based on some transformation of the model that eliminates the individual effects, or that make them redundant in a conditional likelihood function.
Correlated Random Effects (CRE) models / methods

- The CRE model imposes some restrictions on the distribution $CDF(\alpha_i \mid x_{i1}, x_{i2}, \ldots, x_{iT})$.

- The stronger restriction is that $\alpha_i$ is independent of $(x_{i1}, x_{i2}, \ldots, x_{iT})$ and iid$(0, \sigma^2_{\alpha})$. Some textbooks define RE in this restrictive way.

- However, there are more general RE models. For instance, Chamberlain’s CRE model:

\[ \alpha_i = \lambda_0 + x_{i1}' \lambda_1 + \ldots + x_{iT}' \lambda_T + e_i \]

where $e_i$ is independent of $(x_{i1}, x_{i2}, \ldots, x_{iT})$. Based on this assumption, we estimate the parameters $\beta$ and $\lambda'$s. It is a parametric approach because it depends on a parametric assumption on the distribution of $\{x_{i1}, x_{i2}, \ldots, x_{iT}\}$ and $\alpha_i$. 
Relative advantages and limitations of FE and CRE models

(a) FE is more robust because it does not depend on additional assumptions. If the assumption of the CRE is not correct the CRE estimator may be inconsistent.

(b) The FE transformation may eliminate sample variability of the regressors that is exogenous and useful to estimate the model. Therefore, the FE estimator may be less precise or efficient than the CRE estimator (provided the CRE assumption is consistent).

(c) For some models (e.g., some nonlinear dynamic models) there is not a consistent FE method.

Chamberlain (ECMA, 2010 on dynamic probit models).
ESTIMATION WHEN $\mathbb{E}[x_{it} \alpha_i] = 0$

- We start with the simpler case in which the individual effect is not correlated with the regressors.

- In this model, the OLS estimator in the equation in levels is consistent:

$$\hat{\beta}_{OLS} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} x_{it}' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} y_{it} \right)$$

$$= \left( \sum_{i=1}^{N} X_i' X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' Y_i \right)$$

This estimator is also called the **Pooled OLS estimator** because we pool together all the observations as if we had pure cross-sectional data.
GLS ESTIMATOR WHEN $\mathbb{E}[x_{it} \alpha_i] = 0$ (cont.)

- The pooled OLS estimator is not efficient because the error term $\varepsilon_{it}$ is serially correlated:

$$
\mathbb{E} \left( \varepsilon_{it} \varepsilon_{i, t-j} \right) = \mathbb{E} \left( \left[ \alpha_i + u_{it} \right] \left[ \alpha_i + u_{i, t-j} \right] \right) = \mathbb{E} \left( \alpha_i^2 \right) = \sigma^2_{\alpha}
$$

- The asymptotically efficient estimator is the GLS.

- To obtain the efficient GLS estimator of $\beta$ it is convenient to write the model in matrix form:

$$
Y = X \beta + \varepsilon
$$
GLS ESTIMATOR WHEN $\mathbb{E}[x_{it} \alpha_i] = 0$ (cont.)

where:

$$
Y = \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_N
\end{bmatrix}_{NT \times 1} ;
X = \begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_N
\end{bmatrix}_{NT \times K} ;
\varepsilon = \begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_N
\end{bmatrix}_{NT \times 1}
$$

- Let $V_\varepsilon$ be the variance matrix of $\varepsilon$. We know that the GLS estimator can be defined as the OLS estimator of the following transformed model:

$$
\left( V_\varepsilon^{-1/2} Y \right) = \left( V_\varepsilon^{-1/2} X \right) \beta + \left( V_\varepsilon^{-1/2} \varepsilon \right)
$$

We now derive the expressions of $V_\varepsilon$, $V_\varepsilon^{-1/2}$, and of the transformed variables $V_\varepsilon^{-1/2} Y$ and $V_\varepsilon^{-1/2} X$. 
GLS ESTIMATOR WHEN $\mathbb{E}[x_{it} \alpha_i] = 0$ (cont.)

- We have

$$V_\varepsilon = \mathbb{E}[\varepsilon \varepsilon'] = \mathbb{E} \begin{bmatrix} \varepsilon_1 \varepsilon'_1 & \varepsilon_1 \varepsilon'_2 & \cdots & \varepsilon_1 \varepsilon'_N \\ \varepsilon_2 \varepsilon'_2 & \cdots & \varepsilon_2 \varepsilon'_N \\ \vdots & \ddots & \vdots \\ \varepsilon_N \varepsilon'_N \end{bmatrix}$$

- If $\varepsilon_i$ is homocedastic over $i$ and $\mathbb{E}[\varepsilon_i \varepsilon'_j] = 0$ for $i \neq j$,

$$V_\varepsilon = \begin{bmatrix} \Omega & 0 & \cdots & 0 \\ 0 & \Omega & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \Omega \end{bmatrix} = I_N \otimes \Omega$$

where $\Omega$ is the $T \times T$ matrix $\mathbb{E}[\varepsilon_i \varepsilon'_i]$. 
• **GLS and FGLS when u is not i.i.d.** \( \mathbf{V}_\varepsilon = \mathbf{I}_N \otimes \Omega \) where for any individual \( i \) the variance-covariance matrix

\[
\Omega = \mathbb{E} \left[ \begin{array}{cccc}
\varepsilon_{i1}^2 & \varepsilon_{i1}\varepsilon_{i2} & \cdots & \varepsilon_{i1}\varepsilon_{iT} \\
\varepsilon_{i2} & \varepsilon_{i2}^2 & \cdots & \varepsilon_{i2}\varepsilon_{iT} \\
\vdots & \cdots & \ddots & \vdots \\
\varepsilon_{iT} & \cdots & \cdots & \varepsilon_{iT}^2
\end{array} \right]
\]

is an unrestricted.

• The GLS estimator is equivalent to the OLS estimator of the transformed model:

\[
(\mathbf{I}_N \otimes \Omega^{-1/2}) \mathbf{Y} = (\mathbf{I}_N \otimes \Omega^{-1/2}) \mathbf{X} \beta + \varepsilon^*
\]

• This FGLS estimator can be obtained without any restriction on the structure of the variance matrix \( \Omega \).
Let $\hat{\varepsilon}_i = (\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2}, ..., \hat{\varepsilon}_{iT})'$ be the vector of OLS residuals for individual $i$. Then, a consistent estimator of $\Omega$ is:

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i \hat{\varepsilon}_i'$$

The FGLS estimator uses this estimator of $\Omega$. 
ESTIMATION WHEN $\mathbb{E}[x_{it} \alpha_i] \neq 0$

• Now, the Pooled OLS estimator is inconsistent because $\mathbb{E}[x_{it} (\alpha_i + u_{it})] \neq 0$.

• We now examine the following estimators:

(a) OLS in first differences

(b) Within-groups (Fixed Effects) estimator

(c) Least squares dummy variables estimator

(d) Chamberlain’s CRE
(a) OLS in first-differences.

- The OLS-FD estimator is a FE estimator based on a first-differencing transformation of the model:

\[ \Delta y_{it} = \Delta x_{it}' \beta + \Delta u_{it} \]

where \( \Delta \) represents a first time difference, e.g., \( \Delta y_{it} = y_{it} - y_{i,t-1} \).

- Strict exogeneity of \( x_{it} \) implies that \( \mathbb{E}[\Delta x_{it} \Delta u_{it}] = 0 \):

\[
\mathbb{E}[\Delta x_{it} \Delta u_{it}] = \mathbb{E}[x_{it} u_{it}] - \mathbb{E}[x_{it} u_{i,t-1}] - \mathbb{E}[x_{i,t-1} u_{it}] + \mathbb{E}[x_{i,t-1} u_{i,t-1}]
\]

\[ = 0 + 0 + 0 + 0 \]

- Therefore, OLS in FD is a consistent estimator.

\[
\hat{\beta}_{FD} = \left( \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it} \Delta x_{it}' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta x_{it} \Delta y_{it} \right)
\]
The estimator is not efficient (unless $u_{it}$ is a random walk), but we can define a FGLS-FD estimator.
(b) Within-Groups (or FE) estimator

- In general, Fixed Effect estimators of $\beta$ are based either on a transformation of the model or on some sufficient statistic that eliminates the "incidental" or "nuisance" parameters $\{\alpha_i\}$.

- The WG estimator is a FE estimator based on the WG transformation of the model:
  \[
  (y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)'\beta + (u_{it} - \bar{u}_i)
  \]

- The WG estimator is just the OLS estimator of the WG transformation.
  \[
  \hat{\beta}_{WG} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \left( \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \right)
  \]
Consistency of WG estimator in Static-Linear PD

- This estimator is consistent if there is not perfect collinearity in \((x_{it} - \bar{x}_i)\) and \(\mathbb{E}((x_{it} - \bar{x}_i) (u_{it} - \bar{u}_i)) = 0\).

- Note that:
\[
\mathbb{E}((x_{it} - \bar{x}_i) (u_{it} - \bar{u}_i)) = \mathbb{E}[x_{it}u_{it}] - \mathbb{E}[x_{it}\bar{u}_i] - \mathbb{E}[ar{x}_i u_{it}] + \mathbb{E}[\bar{x}_i \bar{u}_i] = 0 + 0 + 0 + 0
\]

Strict exogeneity of the regressors imply that all these expectations are zero and therefore the WG estimator is consistent.

- However, the estimator of \(\alpha_i\) is not consistent for fixed \(T\):
\[
p \lim_{N \to \infty} \alpha_i = \frac{1}{T} \sum_{t=1}^{T} \left( y_{it} - x'_{it} \left[ p \lim_{N \to \infty} \beta \right] \right) = \frac{1}{T} \sum_{t=1}^{T} (y_{it} - x'_{it} \beta) = \alpha_i + \bar{u}_i \neq \alpha_i
\]
(c) Least squares dummy variables (LSDV) estimator. Suppose that we treat the individual effects \(\{\alpha_1, \alpha_2, ..., \alpha_N\}\) as parameters to be estimated together with \(\beta\).

- We can write the model in vector form as:

\[
Y = X\beta + D\alpha + U = (X : D) \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + U
\]

- \(D\) is a \(NT \times N\) matrix of dummy variables, one for each individual.

- The \(i - th\) column of matrix \(D\) contains the observations of the dummy variable for individual \(i\), i.e., 1 if the observation belongs to \(i\) and 0 otherwise.

- \(\alpha\) is the vector of "parameters" \((\alpha_1, \alpha_2, ..., \alpha_N)'\).
LSDV estimator (cont.)

• The LSDV is simply the OLS estimator of \( \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \) in this model:

\[
\begin{pmatrix}
\hat{\beta}_{LSDV} \\
\hat{\alpha}_{LSDV}
\end{pmatrix}
= \left[ (X : D)'(X : D) \right]^{-1} (X : D)'Y
= \left[ \begin{array}{cc}
X'X & X'D \\
D'X & D'D
\end{array} \right]^{-1} \begin{bmatrix}
X'Y \\
D'Y
\end{bmatrix}
\]

• This way of implementing this estimator can computationally demanding if we the panel contains many individuals. We nee to keep in memory and to invert a matrix \((N + K) \times (N + K)\)

• e.g., panel from the CPS with \(N = 120,000\) individuals ...
LSDV estimator (cont.)

• Fortunately, we can obtain the LSDV estimator without having to invert "directly" (or by "brute force") this matrix. We can use properties of partitioned matrix together with the particular structure of the matrices $D'D$ and $D'X$.

• By Frish-Waugh Theorem, the OLS estimator of $\beta$ can be written:

$$\hat{\beta}_{LSDV} = [X' \ M_D \ X]^{-1} [X' \ M_D \ Y]$$

where $M_D$ is the idempotent matrix:

$$M_D = I_{NT} - D(D'D)^{-1}D'$$

$$= (I_N \otimes I_T) - \frac{1}{T} (I_N \otimes 11')$$

$$= I_N \otimes \left( I_T - \frac{1}{T} 11' \right)$$
LSDV estimator  (cont.)

- When we pre-multiply $Y$ and $X$ by $M_D$, these variables are transformed in deviations with respect to individual means: $M_D Y = Y^*$ and $M_D X = X^*$:

$$y_{it}^* = y_{it} - \bar{y}_i \quad ; \quad x_{it}^* = x_{it} - \bar{x}_i$$

and $\bar{y}_i = T^{-1} \sum_{t=1}^{t} y_{it}$ and $\bar{x}_i = T^{-1} \sum_{t=1}^{t} x_{it}$.

- Since $M_D$ is an idempotent matrix, we can write:

$$\hat{\beta}_{LSDV} = [X^* X^*]^{-1} [X^* Y^*]$$

- Therefore, the least squares dummy variables estimator of $\beta$ is EQUIVALENT to the OLS estimator of the transformed model:

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + \varepsilon_{it}^*$$
The LSDV and the Within-Groups are the numerically equivalent (i.e., the same estimator). The WG approach is the efficient algorithm, from a computational point of view, of obtaining the LSDV.

It is simple to show that the estimator of $\alpha_i$ is:

$$\alpha_{i, LSDV} = T^{-1} \sum_{t=1}^{T} \left( y_{it} - x_{it}' \hat{\beta} \right)$$
Is the LSDV estimator consistent? Incidental parameters problem

- In general, the LSDV is **NOT consistent** as \( N \to \infty \) and \( T \) is fixed.

- The reason is the incidental parameters problem: the number of parameters in \( \alpha \) grow at the same rate as \( N \), and we have only \( T \) observations for each \( \alpha_i \). Therefore, \( \widehat{\alpha_{LSDV}} \) is NOT consistent.

- In general, the inconsistency of \( \widehat{\alpha_{LSDV}} \) affects the estimation of \( \beta \) and implies that \( \widehat{\beta_{LSDV}} \) is also inconsistent.

- However, we show below that in STATIC & LINEAR PD models, \( \widehat{\beta_{LSDV}} \) is consistent.

- In general, the consistency property of \( \widehat{\beta_{LSDV}} \) does NOT extend to dynamic PD models and to nonlinear PD models.
• **Remark:** It is interesting that we can get a consistent estimator of $\beta$ despite our estimator of $\alpha$ is inconsistent as $N$ goes to infinity and $T$ is fixed. In the LSDV estimator we try to control for the endogeneity of the unobservables $\alpha_i$ by introducing individual dummies. However, the parameter estimates associated with these dummies are asymptotically biased (as $N$ goes to infinity and $T$ is fixed) such that these dummies are not fully capturing the individual effects $\alpha_i$. Then, how is it possible that we get a consistent estimator of $\beta$?

• The answer is that the estimator of $\alpha_i$, though inconsistent, is such that the estimation error $(\hat{\alpha}_i - \alpha_i)$ is not correlated with the regressors when these are strictly exogenous. Notice that asymptotically $\hat{\alpha}_i$ is equal to $\alpha_i + \bar{u}_i$. Therefore, the estimation error $\hat{\alpha}_i - \alpha_i$ is equal to $\bar{u}_i$. When the regressors are strictly exogenous this estimation error is not correlated with $x_{it}$.

• When the transitory shock $u_{it}$ is i.i.d., the WG estimator is also asymptotically efficient. WHY?
ROBUST STANDARD ERRORS

- Consider the static linear PD model:

\[ \tilde{y}_{it} = \tilde{x}_{it} \beta + \tilde{\varepsilon}_{it} \]

\( \tilde{y}_{it}, \tilde{x}_{it}, \tilde{\varepsilon}_{it} \) are transformations of the original variables \( y_{it}, x_{it}, \varepsilon_{it} \).

- This representation includes as particular cases:

  - Model in levels: \( \tilde{y}_{it} = y_{it} \)

  - Model in FD: \( \tilde{y}_{it} = \Delta y_{it} \)

  - Within Groups: \( \tilde{y}_{it} = y_{it} - \bar{y}_i \)

  - Balestra-Nerlove: \( \tilde{y}_{it} = y_{it} - \theta \bar{y}_i \)
ROBUST STANDARD ERRORS (cont.)

- In vector form:
  \[
  \begin{align*}
  \widetilde{Y}_i &= \widetilde{X}_i \beta + \widetilde{\varepsilon}_i \\
  (T \times 1) &= (T \times K) (K \times 1) (T \times 1)
  \end{align*}
  \]

- The OLS estimator in this transformed model is:
  \[
  \hat{\beta} = \left( \sum_{i=1}^{N} \widetilde{X}_i \widetilde{X}_i \right)^{-1} \left( \sum_{i=1}^{N} \widetilde{X}_i \widetilde{Y}_i \right)
  \]

  Depending on the transformation, the estimator can be the Pooled-OLS, the WG, the OLS-FD, the FGLS, ...

- Under the assumption that the vectors \( \widetilde{\varepsilon}_i \) are independent over individuals, the variance matrix of this estimator is:
  \[
  V(\hat{\beta}) = \left( \sum_{i=1}^{N} \widetilde{X}_i \widetilde{X}_i \right)^{-1} \left( \sum_{i=1}^{N} \widetilde{X}_i \mathbb{E} [\widetilde{\varepsilon}_i \widetilde{\varepsilon}_i | \widetilde{X}_i] \widetilde{X}_i \right) \left( \sum_{i=1}^{N} \widetilde{X}_i \widetilde{X}_i \right)^{-1}
  \]
ROBUST STANDARD ERRORS (cont.)

- Therefore, the panel-robust (or clustered over individuals) estimator of $\mathbf{V}(\hat{\beta})$ is:

\[
\hat{\mathbf{V}}(\hat{\beta}) = \left( \sum_{i=1}^{N} \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i \right)^{-1} \left( \sum_{i=1}^{N} \tilde{\mathbf{X}}'_i \hat{\varepsilon}_i \hat{\varepsilon}'_i \tilde{\mathbf{X}}_i \right) \left( \sum_{i=1}^{N} \tilde{\mathbf{X}}'_i \tilde{\mathbf{X}}_i \right)^{-1}
\]

where $\hat{\varepsilon}_i = \tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \hat{\beta}$.

- Note that $\text{cov}(u_{it}, u_{is})$ and $\text{var}(u_{it})$ are completely unrestricted to vary over $i, t, s$. Observations within individual $i$ are correlated in an unrestricted way.

- These standard errors are denoted **Clustered-over individuals Standard Errors**.

- With panel data, it is very important to correct standard errors for serial correlation.
ROBUST STANDARD ERRORS: And Example (Labor Supply)

- We have a sample of 532 males in US over 10 years: 1979-1988 (from Ziliak, JBES 1997). The labor supply equation is:

\[ \log H_{it} = \beta \log W_{it} + \alpha_i + u_{it} \]

- \( H_{it} \) represents the individual’s annual hours of work

- \( W_{it} \) represents the individual’s after-tax hourly wage

- We are interested in the estimation of the elasticity parameter \( \beta \). Using cross-sectional data, estimates of \( \beta \) are very small. We are concerned with \( \text{cov}(W_{it}, \alpha_i) \neq 0 \).
And Example (Labor Supply) [Cont.]

- $\text{cov}(W_{it}, \alpha_i) > 0$ [upward biased in $\beta$] taste for working positively correlated with wages.

- $\text{cov}(W_{it}, \alpha_i) < 0$ [downward biased in $\beta$] omitted sources of wealth and nonlabor income (with negative effect on hours) can be positively correlated with wage.

Main results:

- Substantial downward bias in $\beta$ because $\text{cov}(W_{it}, \alpha_i) \neq 0$;

- Very important to use robust s.e.;

- Substantial loss of precision in FE estimation
• Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pool OLS</th>
<th>Within</th>
<th>FD</th>
<th>FGLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.083</td>
<td>0.168</td>
<td>0.109</td>
<td>0.120</td>
</tr>
<tr>
<td>Robust s.e.</td>
<td>0.030</td>
<td>(0.085)</td>
<td>(0.084)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>Default s.e.</td>
<td>0.009</td>
<td>(0.020)</td>
<td>(0.021)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\sigma_\alpha$</td>
<td>0.181</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>0.232</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.624</td>
<td></td>
<td></td>
<td>0.586</td>
</tr>
<tr>
<td>Number Obs.</td>
<td>5,320</td>
<td>5,320</td>
<td>4,788</td>
<td>5,320</td>
</tr>
</tbody>
</table>
Some issues with FE estimators in static PD models

- Under the strict exogeneity assumption, FE estimators are consistent and robust because the consistency does not depend on any assumption of the joint distribution of $\alpha_i$ and $x$.

- However, there are two relevant issues with FE estimators:
  
  (a) Variance of FE can be substantially larger than Pooled-OLS;

  (b) FE estimator may exacerbate measurement error bias.
(a) Variance of FE can be substantially larger than Pooled-OLS

- We can decompose the variance of the variables \( y_{it} \) and \( x_{it} \) in the sum of Within Groups and Between Groups variance:

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x})^2 = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (\lbrack x_{it} - \bar{x}_i \rbrack + \lbrack \bar{x}_i - \bar{x} \rbrack)^2
\]

\[
= \sum_{i=1}^{N} \sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)^2 + \sum_{i=1}^{N} T_i (\bar{x}_i - \bar{x})^2
\]

\[
SS_{TOTAL} = SS_{WG} + SS_{BG}
\]

- Typically, in micro-level datasets of individuals or firms, most of the variability is BG: individuals are very heterogeneous and this heterogeneity is very persistent over time:

\[
SS_{BG} >> SS_{WG}
\]
(a) Variance of FE can be substantially larger than Pooled-OLS

- Note that for in the Pooled OLS estimator we are using $SS_{TOTAL}$ of the regressors, and this typically implies a very small variance.

- In contrast, in the WG estimator (and similarly in the OLS-FD) we exploit only the $SS_{WG}$ of the regressors. Typically, this represents a small fraction of $SS_{TOTAL}$.

- Therefore, $\text{Var}(WG) \gg \gg \text{Var}(\text{Pooled-OLS})$. 
(b) FE estimator may exacerbate measurement error bias

- When regressors are measured with error the WG estimator still eliminates the bias associated with the individual effect $\alpha_i$, but it may amplify the bias associated with the measurement error.

- To illustrate this, consider the following simple model:

$$y_{it} = \beta x_{it}^* + \alpha_i + u_{it}$$

where $x_{it}^*$ satisfies the strict exogeneity assumption, but it is correlated with $\alpha_i$, $\text{cov}(x_{it}^*, \alpha_i) = \gamma > 0$.

- There researcher does not observe the "true" regressor $x_{it}^*$ but the variable $x_{it}$ that is measured with error:

$$x_{it} = x_{it}^* + e_{it}$$
(b) FE estimator may exacerbate measurement error bias

- The model in levels is:

\[ y_{it} = \beta x_{it} + \varepsilon_{it} \]

where \( \varepsilon_{it} = \alpha_i + u_{it} - \beta e_{it} \). The asymptotic bias of the OLS estimator of this equation is:

\[
Bias(\text{OLS}) = \frac{\text{cov}(x, \varepsilon)}{\text{var}(x_{it})} = \frac{\gamma - \beta \sigma^2_e}{SS_{\text{TOTAL}}/NT}
\]

- The model in WG transformation is:

\[
(y_{it} - \bar{y}_i) = \beta (x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)
\]

where \( (\varepsilon_{it} - \bar{\varepsilon}_i) = (u_{it} - \beta e_{it}) - (\bar{u}_i - \beta \bar{e}_i) \). The asymptotic bias of the WG estimator of this equation is:

\[
Bias(\text{WG}) = \frac{\text{cov}((x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i))}{\text{var}((x_{it} - \bar{x}_i))} = \frac{-\beta \sigma^2_e}{SS_{\text{WG}}/NT}
\]
WG AND MEASUREMENT ERROR  (cont.)

\[ Bias(OLS) = \frac{\gamma - \beta \sigma_e^2}{SS_{TOTAL}/NT} \]

\[ Bias(WG) = \frac{-\beta \sigma_e^2}{SS_{WG}/NT} \]

- **Remark 1:** The sign of the bias may go from positive to negative.

- **Remark 2:** If most of the variability in \( x \) is cross-sectional (i.e., \( SS_{TOTAL} \gg SS_{WG} \)), then the absolute value of the bias can be larger with the WG.
Remarks

- Given these issues, it is clear that there is a "price" the researcher should pay for the robustness of the FE estimator.

- In this context, it is useful to have:

  (a) A test of the null hypothesis \( \mathbb{E}(x_{it} \alpha_i) = 0 \). We do not want to get rid of the BG variation of the regressors if this variation is not correlated with \( \alpha_i \). [Hausman test]

  (b) An estimator that allows for \( \mathbb{E}(x_{it} \alpha_i) \neq 0 \) but does not eliminate the whole BG variability of the regressors (trades-off robustness for efficiency) [Chamberlain RE estimator].
HAUSMAN TEST OF ENDOGENOUS INDIVIDUAL EFFECTS

- $H_0: \mathbb{E}(x_{it} \alpha_i) = 0$ and $H_1: \mathbb{E}(x_{it} \alpha_i) \neq 0$

- Let $\hat{\beta}_{FGLS}$ be the optimal (FGLS) estimator under the assumption that $\mathbb{E}(x_{it} \alpha_i) = 0$.

- Under $[H_0: \mathbb{E}(x_{it} \alpha_i) = 0] \Rightarrow \left\{ \begin{array}{l} \hat{\beta}_{FGLS} \text{ is consistent and efficient} \\
\hat{\beta}_{WG} \text{ is consistent (but not efficient)} \end{array} \right.$

- Under $[H_1: \mathbb{E}(x_{it} \alpha_i) \neq 0] \Rightarrow \left\{ \begin{array}{l} \hat{\beta}_{FGLS} \text{ is inconsistent} \\
\hat{\beta}_{WG} \text{ is still consistent} \end{array} \right.$
HAUSMAN TEST OF ENDOGENOUS INDIVIDUAL EFFECTS

• Therefore, under $H_0$: $\text{plim } (\hat{\beta}_{FGLS} - \hat{\beta}_{WG}) = 0$ and it is possible to prove that:

$$H = (\hat{\beta}_{FGLS} - \hat{\beta}_{WG})' [\text{Var}(\hat{\beta}_{FGLS} - \hat{\beta}_{WG})]^{-1} (\hat{\beta}_{FGLS} - \hat{\beta}_{WG}) \sim \chi^2_K$$

• And, under $H_1$: $\text{plim } (\hat{\beta}_{FGLS} - \hat{\beta}_{WG}) \neq 0$ and $H \sim \text{noncentral } \chi^2$

• Note that the covariance between an efficient estimator and an inefficient estimator is equal to the variance of the efficient estimator:

$$\text{Cov}(\hat{\beta}_{eff}; \hat{\beta}_{ineff}) = \text{Var}(\hat{\beta}_{eff})$$

• Therefore,

$$H = (\hat{\beta}_{FGLS} - \hat{\beta}_{WG})' [\text{Var}(\hat{\beta}_{WG}) - \text{Var}(\hat{\beta}_{FGLS})]^{-1} (\hat{\beta}_{FGLS} - \hat{\beta}_{WG})$$
HAUSMAN TEST OF ENDOGENOUS INDIVIDUAL EFFECTS

- This is a very useful test in PD econometrics.

- However, note that the potential problem of not enough time-variability in the regressors appears also here and affects the power of this test.

- If $x$’s do not have enough time variability then $Var(\hat{\beta}_{WG})$ is "large" relative to $Var(\hat{\beta}_{FGLS})$ such that, and therefore the statistic $H$ can be small even when $||\hat{\beta}_{WG} - \hat{\beta}_{FGLS}||$ is quite large.

- In other words, if there is not enough time variability in $x$ the test may have low power and we may not be able to reject the $H_0$ even if it is false.
CHAMBERLAIN’S CORRELATED RANDOM EFFECTS ESTIMATOR

• Consider the PD model

\[ y_{it} = x'_{it} \beta + \alpha_i + u_{it} \]

• Our main concern is the correlation between \( \alpha_i \) and the regressors \( x \).

• Suppose that we make the following assumption about the joint distribution of \( \alpha_i \) and \( x \):

\[ \alpha_i = x'_{i1} \lambda_1 + \ldots + x'_{iT} \lambda_T + e_i \]

where \( e_i \) is independent of \( x_i \equiv (x_{i1}, x_{i2}, \ldots, x_{iT}) \).

\[ e_i \perp x_i \]
Based on this assumption, we have:

\[ y_{it} = x'_{i1} \lambda_1 + \ldots + x'_{i1} [\lambda_t + \beta] + \ldots x'_{iT} \lambda_T + e_i + u_{it} \]

\[ = \begin{bmatrix} x'_{i1} \\
\vdots \\
x'_{iT} \end{bmatrix} \begin{bmatrix} \lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_T \end{bmatrix} + e_i + u_{it} \]

\[ = x'_i \Pi_t + u^*_{it} \]

where \( x'_i \equiv (x'_{i1}, x'_{i2}, \ldots, x'_{iT}) \) is a vector of regressors, \( \Pi_t \) is a matrix, and \( u^*_{it} = e_i + u_{it} \).
CHAMBERLAIN’S CORRELATED RANDOM EFFECTS ESTIMATOR

- Then, we have a system of $T$ equations with the same regressors but different dependent variables and parameters (a SURE):

\[
\begin{align*}
y_{i1} &= x_i' \Pi_1 + u_{i1}^* \\
y_{i2} &= x_i' \Pi_2 + u_{i2}^* \\
    &\vdots \\
y_{iT} &= x_i' \Pi_T + u_{iT}^*
\end{align*}
\]

- We can estimate each of these $T$ equations separately by OLS (or by FGLS to account for the serial correlation in $u_{it}^*$) to obtain consistent estimates of the parameters.

- But we want to estimate $\beta$, not $\Pi$’s.
Note that we can represent the relationship between $\Pi$’s and $\lambda$’s and $\beta$ as a linear regression-like equation:

$$
\Pi \equiv \begin{bmatrix}
\Pi_1 \\
\Pi_2 \\
\vdots \\
\Pi_T
\end{bmatrix} = \begin{bmatrix}
I_K & I_K & 0 & \cdots & 0 \\
I_K & 0 & I_K & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
I_K & 0 & 0 & \cdots & I_K
\end{bmatrix} \begin{bmatrix}
\beta \\
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_T
\end{bmatrix}
$$

or in a compact form:

$$
\Pi = A \begin{bmatrix}
\beta \\
\lambda
\end{bmatrix}
$$

where $A$ is a $T^2K \times (T + 1)K$ matrix of 0’s and 1’s.
CHAMBERLAIN’S CORRELATED RANDOM EFFECTS ESTIMATOR

• Given the estimate of $\hat{\Pi}$, we can obtain a consistency estimator of $\lambda$’s and $\beta$ as:

$$
\begin{bmatrix}
\hat{\beta} \\
\hat{\lambda}
\end{bmatrix} = (A'A)^{-1} A' \hat{\Pi}
$$

• We can use the estimated variance of $\hat{\Pi}$, i.e., $\hat{\mathbf{V}}(\hat{\Pi})$, to obtain a more efficient estimator (the Optimal Minimum Distance estimator):

$$
\begin{bmatrix}
\hat{\beta} \\
\hat{\lambda}
\end{bmatrix} = (A' \hat{\mathbf{V}}(\hat{\Pi})^{-1} A)^{-1} (A' \hat{\mathbf{V}}(\hat{\Pi})^{-1} \hat{\Pi})
$$

This estimator is asymptotically efficient under the assumptions of the Correlated RE model.
CHAMBERLAIN’S CORRELATED RANDOM EFFECTS ESTIMATOR

- A very interesting property of the CRE model/estimator is that we can include time-invariant regressors.

- Consider the model:

\[ y_{it} = \beta_1 x_{it} + \beta_2 z_i + \alpha_i + u_{it} \]

- Under the CRE assumption:

\[ \alpha_i = \lambda_1^x x_{i1} + \ldots + \lambda_T^x x_{iT} + \lambda^z z_i + e_i \]

- It is simple to show that we can estimate/identify the parameter \( \beta_2 + \lambda^z \).
Hausman tests of endogenous heterogeneity using Correlated RE

- Suppose that the $SS_{WG}$ variation in the data is small such that the WG estimator is imprecise and the Hausman test of endogenous heterogeneity based on $\| \hat{\beta}_{WG} - \hat{\beta}_{FGLS} \|$ has very little power.

- We can use the correlated RE estimator to construct a more powerful test.

$$H = (\beta_{FGLS} - \beta_{CRE})'[Var(\beta_{CRE}) - Var(\beta_{FGLS})]^{-1}(\beta_{FGLS} - \beta_{CRE})$$

- Of course, we need to assume that CRE estimator is always consistent under the null and under the alternative.

- Note that under the null, the FGLS is more efficient than the CRE because it imposes the (correct) restrictions $\lambda = 0$. 
APPLICATION: Crime in Norway

• We have data on crime for 53 districts and 2 years (1972 and 1978) in Norway. We are interested in the estimation of the following model:

\[ \log(crime_{it}) = \alpha_0 + \alpha_1 dum_{78t} + \beta avgpsolv_{it} + \alpha_i + u_{it} \]

\( crime_{it} \) is the crime rate in district \( i \) at year \( t \);

\( dum_t \) is the dummy for year 1978;

\( avgpsolv_{it} \) is the percentage of crimes which were solved in district \( i \) during the previous two-years.

• We are particularly interested in the estimation of the deterrence effect of solving crimes (i.e., parameter \( \beta \)).
• OLS in levels

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>4.182</td>
<td>0.193</td>
</tr>
<tr>
<td>dummy 78</td>
<td>-0.056</td>
<td>0.088</td>
</tr>
<tr>
<td>avgpsov</td>
<td>-0.036</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Number obs. 106
R-square 0.47

According to these estimates, police’s resolution of crimes has a very significant deterrence effect. If the proportion of solved cases increases in 1 percentage point, the crime rate decreases 3.6%. This effect is statistically significant.
• However, the consistency of this OLS estimator is based on the assumption that $E(\text{avgpsolv}_{it} \cdot \alpha_i) = 0$.

• Suppose that it is easier to solve crimes in communities where citizens have a low propensity to criminal activities. For instance, good citizens can be more collaborative with police in the solution of crimes. If that is the case, we have that $E(\text{avgpsolv}_{it} \cdot \alpha_i) < 0$ and the OLS estimator of $\beta$ will be downward biased.

• Therefore, the OLS may be a asymp. biased estimator of the causal effect of $\text{avgpsolv}$ on crime.
The deterrence effect is still significantly different to zero, but much smaller in magnitude. Though we should make a formal test, it seems that $\alpha_i$ is correlated

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>3.316</td>
<td>0.231</td>
</tr>
<tr>
<td>dummy 78</td>
<td>0.099</td>
<td>0.062</td>
</tr>
<tr>
<td>avgpsolv</td>
<td>-0.017</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Within-Groups

Dependent variable: log(crime)

Standard errors robust of heterocedasticity

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>3.316</td>
<td>0.231</td>
</tr>
<tr>
<td>dummy 78</td>
<td>0.099</td>
<td>0.062</td>
</tr>
<tr>
<td>avgpsolv</td>
<td>-0.017</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Number obs. 106

R-square (within) 0.408

R-square (all) 0.436

$\hat{\sigma}_\alpha$ 0.471

$\hat{\sigma}_u$ 0.244
with the regressors. There may be unobserved district characteristics (i.e., $\alpha_i$) which have a positive effect on crime and are negatively correlated with the percentage of cases solved. That is, there are "good districts" (i.e., low $\alpha_i$) where crime is low and most crimes are easy to solve, and "bad districts" (i.e., high $\alpha_i$) where crime is high and crimes are more difficult to solve. If we do not control for these unobserved district characteristics, we get a downward bias estimate of $\beta$, i.e., an estimate of the deterrence effect that is partly spurious.
FGLS estimator

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>3.799</td>
<td>0.186</td>
</tr>
<tr>
<td>dummy 78</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>avgpsolv</td>
<td>-0.027</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Number obs. 106
R-square (within) 0.384
R-square (all) 0.468
\( \hat{\sigma}_\alpha \) 0.371
\( \hat{\sigma}_u \) 0.244

This FGLS estimate of \( \beta \) is between the OLS and the WG. Still, it seems that \( \alpha_i \) is correlated with the regressors and OLS overestimates the deterrence effect.
Hausman test of the null hypothesis that the regressors are not correlated with $\alpha_i$.

- The Hausman statistic is equal to 14.85. Under the null hypothesis that $E(\alpha_i X_i) = 0$, it is distributed as Chi-square with 2 degrees of freedom (i.e., number of regressors, other than the constant term).

- The p-value of the test is 0.0006, which means that we can clearly reject the null hypothesis under the typical choice of significance level (e.g., 5%, 1%) and even with much smaller significance levels.
Which estimate of $\beta$ do you find more reasonable?

- There is clear evidence that the unobserved district effect $\alpha_i$ is correlated with the regressors.

- Therefore, both the OLS and the Balestra-Nerlove estimators are inconsistent. The most reasonable estimator in this context seems the Within-Groups.
3. DYNAMIC PANEL DATA MODELS

3.1. INTRODUCTION

- **Dynamic models**: lagged values the variables (dependent and/or explanatory) have a causal effect on the current value of the dependent variable.

- **Multiple sources of dynamics**: adjustment costs, irreversibilities, habits, storability, etc.

- Short-run and long-run responses are different.

- The assumption of **strict exogeneity of the regressors does not hold**. This has very important implications on the properties of estimators.
Example 1: Firms' dynamic labor demand. Consider the dynamic labor demand equation (Sargent, JPE 1978; Arellano and Bond, REStud 1991):

\[ n_{it} = \beta_0 + \beta_1 n_{i,t-1} + \beta_2 w_{it} + \beta_3 y_{it} + \gamma_t + \alpha_i + u_{it} \]

\( n = \ln(\text{Employment}) \), \( w = \ln(\text{Wage}) \), and \( y = \ln(\text{Output}) \); \( \gamma_t \) represents aggregate shocks, and \( \alpha_i \) represents firm specific and time invariant factors that affect productivity and or input prices and which are unobservable to the econometrician. The unobservable \( u_{it} \) represents idiosyncratic shocks. The parameter \( \beta_1 \) is associated with labor adjustment costs.

In this model, it is clear that \( n_{i,t-1} \) is not strictly exogenous with respect to \( u_{it} \). It is not correlated with \( u_{it} \) or with future values of \( u \), but by construction \( n_{i,t-1} \) depends on \( u_{i,t-1}, u_{i,t-2}, ... \)
• Parameter $\beta_1$ has important economic implications. Short-run elasticity = $\beta_2$ : Long-run elasticity = $\frac{\beta_2}{1 - \beta_1}$. 
Example 2. [Cont]

- Employment has a strong positive serial correlation (after controlling for wages and output). Two possible explanations.

- (A) **Structural state dependence.** Labor adjustment costs.

- (B) **Persistent unobserved heterogeneity.** $\alpha_i$ represents firm heterogeneity in the steady-state optimal level of employment (due to productivity, other labor costs). $\alpha_i$ generates positive serial correlation in employment.

- The two explanations have very different economic implications (e.g., LR elasticity).
• **Example 2 (Investment in R&D):** Consider the following dynamic model for firm’s investment in R&D:

\[
RD_{it} = \beta_0 + \beta_1 RD_{it-1} + \varepsilon_{it}
\]

\(RD_{it}\) is the firm investment during a year, and \(RD_{it-1}\) is the same firm investment at previous year.

• The term \(\beta_1 RD_{it-1}\) captures the fact that previous investment in R&D increases the marginal profit (e.g., reduces the marginal cost) of investment today, e.g., know-how.

• OLS estimation of this equation typically provides a positive, large, and significant estimate of the parameter \(\beta_1\).
Two possible explanations for the OLS estimate of $\beta_1$.

(A) **Structural state dependence.** Previous investment in R&D causes an increase in the marginal profit of investment today.

(B) **Persistent unobserved heterogeneity.** $\varepsilon_{it}$ represents firm heterogeneity in the profitability of investment in R&D. If $\varepsilon_{it}$ is persistent over time, then firms with higher values of $\varepsilon_{it}$ tend to have also higher past levels of R&D: $\text{cov}(RD_{it-1}, \varepsilon_{it}) > 0$. OLS estimation will provide an upward biased estimate of $\beta_1$. 
Example 3: AR(1) Panel Data Model. It is the simplest example of dynamic PD model.

\[ y_{it} = \beta y_{i,t-1} + \alpha_i + u_{it} \]

For the moment, we maintain assumption \( u_{it} \sim iid \) over \((i, t) (0, \sigma_u^2)\).

We will use this model as a benchmark. We will derive simple analytical expressions for the asymptotic bias and variance of different estimators of \( \beta \) in this model.

These expressions will help us to understand the problems and merits of different estimators.
• Solving backwards, we have that

\[ y_{it} = \alpha_i (1 + \beta + \beta^2 + \ldots) + u_{it} + \beta u_{i,t-1} + \beta^2 u_{i,t-2} + \ldots \]

• If \( \beta \neq 0 \):

(1) The regressor \( y_{i,t-1} \) is correlated with \( \alpha_i \): OLS is never consistent.

(2) The regressor \( y_{i,t-1} \) is correlated with \( u_{i,t-1}, u_{i,t-2}, \ldots \), and therefore it is not strictly exogenous.

• Point (2) has important implications on the properties of different estimators. More specifically, for dynamic PD models the WG estimator and the OLS estimator in first differences are not consistent.
3.2. INCONSISTENCY OF OLS-FD

- Consider the model in FD:

\[ \Delta y_{it} = \beta \Delta y_{i,t-1} + \Delta u_{it} \]

- OLS estimation of this equation is consistent if \( \text{Cov}(\Delta y_{i,t-1}, \Delta u_{it}) = 0. \)

- However,

\[
\text{Cov}(\Delta y_{i,t-1}, \Delta u_{it}) = \\
= \mathbb{E}(y_{t-1}u_t) - \mathbb{E}(y_{t-1}u_{t-1}) - \mathbb{E}(y_{t-2}u_t) + \mathbb{E}(y_{t-2}u_{t-1}) \\
= 0 - \sigma_u^2 0 0 \\
= -\sigma_u^2 \neq 0
\]
• **Exercise:** Show that $\text{Var}(\Delta y_{i,t-1}) = \frac{2\sigma^2_u}{1 + \beta}$.

• Therefore

$$\lim_{N \to \infty} \hat{\beta}_{OLS-FD} = \beta + \frac{\text{Cov}(\Delta y_{i,t-1}, \Delta u_{it})}{\text{Var}(\Delta y_{i,t-1})}$$

$$= \frac{\beta - 1}{2}$$

• And the **bias is** $\frac{-(1+\beta)}{2}$. It is clear that the bias of this estimator can be very large. For instance, when the true $\beta$ is $+0.5$ the estimator converges in probability to $-0.25$.

• Note that the bias does not depend on $T$. Therefore, the **bias does not go to zero as $T$ goes to infinity.** The bias of the OLS-FD is as bad for $T = 3$ as for $T = 500$. 
Remark 1: What if \( u_{it} \) is serially correlated?

Suppose \( u_{it} \sim AR(1) \):

\[
    u_{it} = \rho \ u_{it-1} + a_{it}
\]

with \( a_{it} \) not serially correlated and \(|\rho| < 1\).

Then, it is possible to show that [Good Exercise!]

\[
    \text{Bias} \left( \hat{\beta}_{OLS-FD} \right) = \frac{\text{Cov}(\Delta y_{i,t-1}, \Delta u_{it})}{\text{Var}(\Delta y_{i,t-1})} = \left( \frac{1 + \beta}{2} \right) (\rho - 1)
\]

For \( \rho = 1 \), \( \hat{\beta}_{OLS-FD} \) is consistent. For \( |\rho| < 1 \), \( \hat{\beta}_{OLS-FD} \) is downward biased.
Remark 2: Bias reduction

- If $u_{it}$ is not serially correlated, we have that $\hat{\beta}_{OLS-FD} \xrightarrow{p} \frac{\beta - 1}{2}$.

- This provides a very simple approach to correct for the bias and obtain a consistent estimator of $\beta$.

- Define the estimator:
  \[
  \hat{\beta}^* = 2 \hat{\beta}_{OLS-FD} + 1
  \]

  For this model, $\hat{\beta}^*$ is consistent and asymptotically normal (and very simple to calculate).

- Can we generalize this approach to more general DPD models?
• **Remark 2: Bias reduction (2)**

- In general, the models that we have in empirical applications include more explanatory variables (not only $y_{it-1}$) and the error term $u_{it}$ may be serially correlated, e.g., when $u_{it}$ is AR(1), the bias correction depends on $\rho$ that is unknown.

- In that model, we do not have a simple expression for the bias (it depends on the joint stochastic process of the exogenous regressors) and on unknown parameters in the stochastic process for $u_{it}$ (e.g., see the previous example for the bias when $u_{it}$ follows an AR(1)).

- Still in the absence of better methods (e.g., in dynamic nonlinear PD models), bias reduction techniques can be useful.
3.3. INCONSISTENCY OF WG

- Consider the WG transformed equation:

\[
(y_{it} - \bar{y}_i) = \beta \left( y_{i,t-1} - \bar{y}_i(-1) \right) + (u_{it} - \bar{u}_i)
\]

- The WG estimator is consistent if \( \text{Cov}((y_{i,t-1} - \bar{y}_i), (u_{it} - u_i)) = 0 \). However, this covariance is

\[
\text{Cov}((y_{i,t-1} - \bar{y}_i), (u_{it} - u_i))
\]

\[
= \mathbb{E}(y_{i,t-1}, u_{it}) - \mathbb{E}(y_{i,t-1}, \bar{u}_i) - \mathbb{E}(\bar{y}_i, u_{it}) + \mathbb{E}(\bar{y}_i, \bar{u}_i)
\]

\[
= 0 \neq 0 \neq 0 \neq 0
\]

\[
< 0
\]
Nickell (Econometrica, 1981) derived the expression for the asymptotic bias of the WG estimator in an AR(1) panel data model (for fixed $T$):

$$bias(\hat{\beta_{WG}}) = \frac{-(1 - \beta^2) \cdot h(T, \beta)}{(T - 1) - 2\beta \cdot h(T, \beta)}$$

where:

$$h(T, \beta) = \frac{1}{1 - \beta} \left(1 - \frac{1 - \beta^T}{T(1 - \beta)}\right)$$

1. For $|\beta| < 1$, it is always negative.

2. It increases in absolute when $\beta$ goes to one.

3. It decreases with $T$ and becomes zero as $T$ goes to infinity.

4. If $\beta$ is not small, the bias can be important even for $T$ greater than 20.
• Remark: WG is consistent as $T \to \infty$

• In contrast to the case of the OLS-FD, the bias of the WG estimator declines monotonically with $T$. Why?

• As we have seen, $WG$ is numerically equivalent to LSDV. In the LSDV, we use $T$ observations to estimate each individual effect. As $T$ increases, the bias in the estimates of the individual effects becomes smaller and this contributes to reduce the bias of $\beta$.

• In contrast, the bias of the OLS-FD comes from $Cov(\Delta y_{i,t-1}, \Delta u_{it}) \neq 0$, and this source is not affected by $T \to \infty$. 
Remark: Bias reduction

Again, we have obtained an expression: \( \text{plim} \hat{\beta}_{WG} = p(\beta, T) \), where the function \( p(.) \) is known, and it strictly monotonic in \( \beta \), and therefore it is invertible.

Define the estimator:

\[ \hat{\beta}^* = p^{-1}(\hat{\beta}_{WG}, T) \]

where \( p^{-1}(.) \) is the inverse function. By the continuous function theorem, \( \hat{\beta}^* \) is consistent and asymptotically normal.

Again, a practical problem with this approach is that when we generalize the DPD model, the expression of \( p(\beta, T) \) becomes complicated and depends on many unknown parameters.
3.4. ANDERSON-HSIAO ESTIMATOR (JoE, 1982)

- Consider the equation in FD:

\[ \Delta y_{it} = \beta \Delta y_{i,t-1} + \Delta u_{it} \]

- Suppose that \( u_{it} \) is iid. Then,

(a) \( \Delta u_{it} \) is NOT correlated with \( y_{i,t-2}, y_{i,t-3}, \ldots \)

(b) If \( \beta \neq 0 \), then \( \Delta y_{i,t-1} \) is correlated with \( y_{i,t-2}, y_{i,t-3}, \ldots \)

- Therefore, under these conditions, we can use \( y_{it-2} \) (and also further lags of \( y \)) as an instrument for \( \Delta y_{i,t-1} \) in the equation in FD.
ANDERSON-HSIAO ESTIMATOR  (cont)

- For the AR(1)-PD model without other regressors, the Anderson-Hsiao estimator is defined as:

\[
\hat{\beta}_{AH} = \frac{\sum_{t=3}^{T} \left[ \sum_{i=1}^{N} y_{i,t-2} \Delta y_{it} \right]}{\sum_{t=3}^{T} \left[ \sum_{i=1}^{N} y_{i,t-2} \Delta y_{i,t-1} \right]}
\]

Notice that we need \( T \geq 3 \) to implement this estimator.

- In the AR(1)-PD model that includes also exogenous regressors, Anderson-Hsiao estimator is an IV estimator in the equation in FD where the FD of the lagged endogenous variable is instrumented with \( y_{it-2} \).
Asymptotic Properties of Anderson-Hsiao Estimator

To derive this asymptotic variance, notice that:

\[
\sqrt{N} \left( \hat{\beta}_{AH} - \beta \right) = \sum_{t=3}^{T} \left[ \frac{1}{N} \sum_{i=1}^{N} y_{i,t-2} \Delta y_{i,t-1} \right]^{-1} \sum_{t=3}^{T} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,t-2} \Delta u_{it} \right]
\]

As \( N \to \infty \),

\[
\frac{1}{N} \sum_{i=1}^{N} y_{i,t-2} \Delta y_{i,t-1} \to_p \mathbb{E} (y_{t-2} \Delta y_{t-1})
\]

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} y_{i,t-2} \Delta u_{it} \to_d N \left(0, \text{Var} \left( y_{i,t-2} \Delta u_{it} \right) \right)
\]

By Mann-Wald Theorem, as \( N \to \infty \),

\[
\sqrt{N} \left( \hat{\beta}_{AH} - \beta \right) \to_d N \left(0, V_{AH} \right)
\]

with \( V_{AH} \) = \[
\frac{\sum_{t=3}^{T} \text{Var} \left( y_{t-2} \Delta u_{t} \right)}{\left[ \sum_{t=3}^{T} \mathbb{E} (y_{t-2} \Delta y_{t-1}) \right]^2}
\]
• Asymptotic Variance Anderson-Hsiao Estimator

• **Exercise:** Show that \( \mathbb{E}(y_{t-2} \Delta y_{t-1}) = \frac{-\sigma_u^2}{1 + \beta} \) and \( Var(y_{t-2} \Delta u_{it}) = 2\sigma_u^2 \ Var(y_t) \).

• Given these expressions, we have that the variance of the limiting distribution is:

\[
V_{AH} = \frac{(T - 2) \ 2\sigma_u^2 \left( \frac{\sigma_\alpha^2}{(1 - \beta)^2} + \frac{\sigma_u^2}{1 - \beta^2} \right)}{\left[ (T - 2) \sigma_u \ (1 + \beta)^{-1} \right]^2}
\]

\[
= \frac{2}{(T - 2)} \left( \frac{1 + \beta}{1 - \beta} \right) \left( 1 + \left( \frac{1 + \beta}{1 - \beta} \right) \frac{\sigma_\alpha^2}{\sigma_u^2} \right)
\]
- Asymptotic Variance AH Estimator [Cont]

- **Comment 1:** The variance increases more than exponentially with $\beta$. It can be pretty large if $\beta$ is close to 1.

- For $\beta = 1$, the variance is infinite, i.e., the instrument $y_{i,t-2}$ is not correlated with $\Delta y_{i,t-1}$. 
• **Comment 2:** The variance also increases with $\sigma_\alpha^2/\sigma_u^2$. Typically, in many PD models with household or firm level data, the variance of the individual effect is large relative to the variance of the transitory shock. Therefore, the AH estimator can be quite imprecise.

• **Example:** Suppose that:

$$\beta = 0.6 ; \quad \sigma_\alpha/\sigma_u = 6 ; \quad N (T - 2) = 3600$$

You can verify that for this example $sd(\hat{\beta}_{AH}) = 0.567$, and the 95% confidence interval for $\beta$ is:

$$95\% \ CI \ for \ \beta = [-0.511 , \ 1.711]$$

That basically does not provide any information about the true value of $\beta$. 
Comment 3: There are several reasons for the inefficiency of this estimator.

First, it does not exploit all the moment conditions that the model implies. Relatedly, for every instrument that we include (i.e., \( y_{it-2}, y_{it-3}, \ldots \)) we lose a complete cross-section of data.

The Arellano-Bond GMM estimator deals with these problems.

Second, even if we exploit all the moment conditions and all the data in an efficient way, lagged levels \( y \) can be weakly correlated with current \( \Delta y \) when \( \beta \) is close to one. The Arellano-Bover and the Blundell-Bond estimators deal with this second problem.
Comment 4: The consistency of the AH estimator depends crucially on the assumption of no-serial correlation of \( u_{it} \). For instance, if \( u_{it} \) follows an AR process, AH is inconsistent.

In contrast, if \( u_{it} \) follows a random walk, OLS-FD is consistent and it does not suffer of "weak instruments" problem.

This motivates two important questions:

(a) Can we test for serial correlation in \( u_{it} \)? [YES]

(b) Can we extend AH idea to serially correlated \( u_{it} \)? [YES]
Extension of AH-IV estimator to AR(1) transitory shock

- Suppose that $u_{it} = \rho u_{i,t-1} + a_{it}$, where $a_{it}$ is not serially correlated.

- We can take quasi-first-differences:

$$y_{it} - \rho y_{i,t-1} = \beta y_{it-1} - \beta \rho y_{i,t-2} + \alpha_i(1 - \rho) + u_{it} - \rho u_{i,t-1}$$

Or

$$y_{it} = (\beta + \rho) y_{i,t-1} + (-\beta \rho) y_{i,t-2} + \alpha_i(1 - \rho) + a_{it}$$

- This equation in first differences is:

$$\Delta y_{it} = (\beta + \rho) \Delta y_{i,t-1} + (-\beta \rho) \Delta y_{i,t-2} + a_{it}$$

- Note that $\mathbb{E}(y_{t-2} a_{t-2}) = 0$ and $\mathbb{E}(\Delta y_{t-2} a_{t-2}) = 0$; $\Delta y_{i,t-1}$ depends on $y_{i,t-2}$; and $y_{t-2}$ and $\Delta y_{t-2}$ are not collinear. Consistent IV estimator of $\pi_1 \equiv \beta + \rho$ and $\pi_1 \equiv -\beta \rho$. 
3.5. ARELLANO-BOND GMM (RESTUD, 1991)

- The AH estimator is based on the moment condition:

$$
E \left( \sum_{t=3}^{T} y_{i,t-2} \Delta u_{it} \right) = E \left( \sum_{t=3}^{T} y_{i,t-2} (\Delta y_{it} - \beta \Delta y_{i,t-1}) \right) = 0
$$

The estimator pools together $T - 2$ cross-sections.

- But the model implies other moment conditions which are not exploited by the AH estimator.

- Let’s represent the model in FD as a system of linear equations, one for each time period.

$$
\begin{align*}
\text{Equation for } t = 3 : & \quad \Delta y_{i3} = \beta \Delta y_{i2} + \Delta u_{i3} \\
\text{Equation for } t = 4 : & \quad \Delta y_{i4} = \beta \Delta y_{i3} + \Delta u_{i4} \\
& \quad \vdots \\
\text{Equation for } t = T : & \quad \Delta y_{iT} = \beta \Delta y_{iT-1} + \Delta u_{iT}
\end{align*}
$$
• If \( \{u_{it}\} \) is iid, we have the following moment conditions:

For \( t = 3 \): \( \mathbb{E}[y_{i1} \triangle u_{i3}] = 0 \)

For \( t = 4 \):
\[
\begin{align*}
\mathbb{E}[y_{i1} \triangle u_{i4}] &= 0 \\
\mathbb{E}[y_{i2} \triangle u_{i4}] &= 0
\end{align*}
\]

For \( t = 5 \):
\[
\begin{align*}
\mathbb{E}[y_{i1} \triangle u_{i5}] &= 0 \\
\mathbb{E}[y_{i2} \triangle u_{i5}] &= 0 \\
\mathbb{E}[y_{i3} \triangle u_{i5}] &= 0
\end{align*}
\]

• The total number of moment conditions is:
\[
q \equiv 1 + 2 + \ldots + (T - 2) = \frac{(T - 2)(T - 1)}{2}
\]
We can represent these $q$ moment conditions in vector form as:

\[
\mathbb{E} \left[ \begin{array}{c}
    y_{i1} \triangle u_{i3} \\
    y_{i1} \triangle u_{i4} \\
    y_{i2} \triangle u_{i4} \\
    \ldots
\end{array} \right] = \mathbb{E} \left[ Z_i \left( \Delta Y_i - \beta \Delta Y_{i(-1)} \right) \right]
\]

where $\Delta Y_i$ and $\Delta Y_{i(-1)}$ are $(T - 2) \times 1$ vectors

\[
\Delta Y_i = (\Delta y_{i3}, \Delta y_{i4}, \ldots, \Delta y_{iT})'
\]

and $Z_i$ is the $q \times (T - 2)$ matrix of instruments.
- And $Z_i$ is the $q \times (T - 2)$ matrix of instruments:

$$Z_i = \begin{bmatrix}
y_{i1} & 0 & 0 & \ldots & 0 \\
0 & y_{i1} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & y_{i1} \\
0 & y_{i2} & 0 & \ldots & 0 \\
0 & 0 & y_{i2} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & y_{iT-2} \\
0 & 0 & 0 & \ldots & y_{iT-2}
\end{bmatrix}$$
To understand the structure of the $q \times (T - 2)$ matrix of instruments $Z_i$, note that:

- Each row of $Z_i$ represents an instrument (moment condition) for one of the $T - 2$ equations;

- For an instrument (moment condition) in equation $t$, we set equal to zero all the elements of that row except the element corresponding to period $t$.

- For instance, row $[y_{i1} \ 0 \ 0 \ \ldots \ 0]$ represents $y_{i1}$ as instrument in the first equation, $t = 3$.

- Row $[0 \ 0 \ y_{i1} \ \ldots \ 0]$ represents $y_{i1}$ as instrument in the third equation, $t = 5$.

- Row $[0 \ y_{i2} \ 0 \ \ldots \ 0]$ represents $y_{i2}$ as instrument in the second equation, $t = 4$. 
Let $m_N(\beta)$ be the $q \times 1$ vector of sample moment conditions, i.e., the sample counterpart of $\mathbb{E}\left[ Z_i \left( \Delta Y_i - \beta \Delta Y_{i(-1)} \right) \right]$

$$m_N(\beta) = N^{-1} \sum_{i=1}^{N} Z_i \left( \Delta Y_i - \beta \Delta Y_{i(-1)} \right)$$

Note that we use the $T-2$ cross-sections to construct these moment conditions.

The Arellano-Bond GMM estimator is:

$$\hat{\beta}_{AB} = \arg \min_{\{\beta\}} m_N(\beta)' \hat{\Omega}^{-1} m_N(\beta)$$

where $\hat{\Omega}$ is a consistent estimator of the optimal weighting matrix, $\Omega = \mathbb{E}\left( Z_i \Delta U_i \Delta U_i' Z_i \right)$, with $\Delta U_i = (\Delta u_{i3}, \Delta u_{i4}, \ldots, \Delta u_{iT})'$. 
Deriving the first order conditions and solving for $\hat{\beta}_{AB}$ we get:

$$\hat{\beta}_{AB} = \frac{\left(\sum_{i=1}^{N} z_i \Delta Y_{i,-1}\right)'}{\left(\sum_{i=1}^{N} z_i \Delta Y_{i,-1}\right)'} \hat{\Omega}^{-1} \left(\sum_{i=1}^{N} z_i \Delta Y_i\right)$$

$$\frac{\left(\sum_{i=1}^{N} z_i \Delta Y_{i,-1}\right)'}{\left(\sum_{i=1}^{N} z_i \Delta Y_{i,-1}\right)'} \hat{\Omega}^{-1} \left(\sum_{i=1}^{N} z_i \Delta Y_{i,-1}\right)$$
• This estimator is obtained in **two steps**.

• In the first stage, we choose the weighting matrix under the assumption that $u_{it}$ is iid over time. Under this assumption $\Omega = \sigma_u^2 \mathbb{E} \left( Z_i \mathbf{H} Z_i' \right)$ where $\mathbf{H}$ is the $(T - 2) \times (T - 2)$ matrix with $2'$s in the main diagonal, $-1'$s in the two sub-main diagonals, and zeros otherwise.

\[
\mathbf{H} = \begin{bmatrix}
2 & -1 & \ldots & 0 & 0 \\
-1 & 2 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & 2 & -1 \\
0 & 0 & \ldots & -1 & 2 \\
\end{bmatrix}
\]
Then, the first stage GMM estimator is:

\[
\hat{\beta}_{AB}^{(1)} = \frac{\left( \sum_{i=1}^{N} Z_i \Delta Y_i(-1) \right)'}{\left( \sum_{i=1}^{N} Z_i \Delta Y_i(-1) \right)'} \left( \sum_{i=1}^{N} Z_i H Z_i' \right)^{-1} \left( \sum_{i=1}^{N} Z_i \Delta Y_i \right)
\]
• In the second stage GMM estimator, given \( \hat{\beta}_A^{(1)} \) we can get the residuals
\[
\Delta \hat{u}_{it} = \Delta y_{it} - \hat{\beta}_A^{(1)} \Delta y_{i,t-1}.
\]
And we can use these residuals to obtain a consistent estimator of \( \Omega \) that is robust to heterocedasticity.

\[
\hat{\Omega} = N^{-1} \sum_{i=1}^{N} Z_i \Delta \hat{U}_i \Delta \hat{U}_i' Z_i'
\]

where \( \Delta \hat{U}_i = (\Delta \hat{u}_{i3}, \Delta \hat{u}_{i4}, \ldots, \Delta \hat{u}_{iT})' \).

• Finally, the two-stage GMM estimator is:

\[
\hat{\beta}_A^{(2)} = \frac{\left( \sum_{i=1}^{N} Z_i \Delta Y_i^{(-1)} \right)' \left( \sum_{i=1}^{N} Z_i \Delta \hat{U}_i \Delta \hat{U}_i' Z_i' \right)^{-1} \left( \sum_{i=1}^{N} Z_i \Delta Y_i \right)}{\left( \sum_{i=1}^{N} Z_i \Delta Y_i^{(-1)} \right)' \left( \sum_{i=1}^{N} Z_i \Delta \hat{U}_i \Delta \hat{U}_i' Z_i' \right)^{-1} \left( \sum_{i=1}^{N} Z_i \Delta Y_i^{(-1)} \right)}
\]

• We can generalize this estimator to any PD model with predetermined explanatory variables.
• We can generalize this estimator to any PD model with predetermined explanatory variables.

• Consider the model in vector form:

\[ Y_i = X_i \beta + \alpha_i + U_i \]

• Let \( Z_i \) be the matrix of instruments constructed in the way described above, such that:

\[ \mathbb{E}[Z_i (\Delta Y_i - X_i \beta)] = 0 \]
• The two-stage GMM estimator of $\beta$ is:

$$\hat{\beta}_{AB}^{(2)} = \left[ \left( \sum_{i=1}^{N} Z_i \Delta X_i \right)' \left( \sum_{i=1}^{N} Z_i \Delta \hat{U}_i \Delta \hat{U}_i' Z_i' \right)^{-1} \left( \sum_{i=1}^{N} Z_i \Delta Y_i \right) \right]^{-1}$$

$$\left[ \left( \sum_{i=1}^{N} Z_i \Delta X_i \right)' \left( \sum_{i=1}^{N} Z_i \Delta \hat{U}_i \Delta \hat{U}_i' Z_i' \right)^{-1} \left( \sum_{i=1}^{N} Z_i \Delta X_i \right) \right]^{-1}$$
EXAMPLE. Employment equation: Arellano & Bond (REStud, 91)

- Panel of 140 quoted manufacturing firms in UK. 611 observations (max $T_i = 6$; average $T_i = 4.3$). Dynamic employment equation (in logs):

$$ n_{it} = \beta_1 n_{i,t-1} + \beta_2 w_{it} + \beta_3 k_{it} + \beta_4 q_{it} + \gamma_t + \alpha_i + u_{it} $$

- GMM estimation of

$$ \Delta n_{it} = \beta_1 \Delta n_{i,t-1} + \beta_2 \Delta w_{it} + \beta_3 \Delta k_{it} + \beta_4 \Delta q_{it} + \gamma_t + \alpha_i + u_{it} $$

using moment conditions:

$\{n_{i1}, w_{i1}, k_{i1}, q_{i1}\}$ in eq. at $t = 3, 4, 5, 6$

$\{n_{i2}, w_{i2}, k_{i2}, q_{i2}\}$ in eq. at $t = 4, 5, 6$

$\{n_{i3}, w_{i3}, k_{i3}, q_{i3}\}$ in eq. at $t = 5, 6$
EXAMPLE. Employment equation [Cont.]

- Note that the regressors \((w_{it}, k_{it}, y_{it})\) are also endogenous in the equations in levels and in first-differences.

\[
E[\Delta w_{it} \Delta u_{it}] \neq 0, E[\Delta k_{it} \Delta u_{it}] \neq 0, E[\Delta q_{it} \Delta u_{it}] \neq 0
\]

- The AB-GMM can deal also with this endogeneity.

- \(\{w_{it-2}, k_{it-2}, q_{it-2}\}\) are valid instruments for \(\{\Delta w_{it}, \Delta k_{it}, \Delta q_{it}\}\) because:

  - Not correlated with \(\Delta u_{it}\) (if \(u_{it}\) not serially correlated)

  - Correlated with \(\{\Delta w_{it}, \Delta k_{it}, \Delta q_{it}\}\) because they capture the serial correlation in these variables because shocks other than \(u_{it}\), and because potential dynamics in these variables.
Employment equation: Arellano & Bond (REStud, 91). Tables 4 and 5

140 firms and 611 observations
Dependent variable: ln(employment)

<table>
<thead>
<tr>
<th>Regressor</th>
<th>OLS-Levels</th>
<th>WG</th>
<th>AH</th>
<th>AB-1step</th>
<th>AB-2step</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln(\text{emp}) [t - 1] )</td>
<td>1.045</td>
<td>0.734</td>
<td>2.308</td>
<td>0.686</td>
<td>0.629</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.058)</td>
<td>(1.055)</td>
<td>(0.145)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>( \ln(\text{wage}) [t] )</td>
<td>-0.524</td>
<td>-0.557</td>
<td>-0.810</td>
<td>-0.608</td>
<td>-0.526</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
<td>(0.155)</td>
<td>(0.283)</td>
<td>(0.178)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>( \ln(\text{wage}) [t - 1] )</td>
<td>0.477</td>
<td>0.326</td>
<td>1.422</td>
<td>0.393</td>
<td>0.311</td>
</tr>
<tr>
<td></td>
<td>(0.169)</td>
<td>(0.143)</td>
<td>(0.851)</td>
<td>(0.168)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>Sargan (d.o.f.)</td>
<td></td>
<td></td>
<td></td>
<td>65.8 (25)</td>
<td>31.4 (25)</td>
</tr>
<tr>
<td>m2</td>
<td></td>
<td></td>
<td></td>
<td>-0.516</td>
<td>-0.434</td>
</tr>
</tbody>
</table>
SPECIFICATION TESTS IN DPD MODELS

- Consistency of the previous Arellano-Bond (or for that matter Anderson-Hsiao) estimator is based on the assumption that $u_{it}$ is not serially correlated.

- Serial correlation in $u_{it}$ implies that $y_{it-2}$ is no longer a valid instrument in equation in FD.

- We would like to test for this assumption.

- Two important specification tests:
  
  (a) Hansen-Sargan test of over-identifying restrictions

  (b) Test of second order correlation in $\Delta u_{it}$
(a) Test of over-identifying restrictions (Hansen-Sargan):

\[ H_0 : \mathbb{E} (Z_i \Delta U_i) = 0 \quad (q \text{ restrictions}) \]

- The model is based on the \( q \) restrictions \( \mathbb{E} (Z_i \Delta U_i) = 0 \). In principle, we would like to impose all these restrictions in the estimation of \( \beta \).

\[
m_N(\hat{\beta}) = \sum_{i=1}^{N} Z_i \left[ \Delta Y_i - \Delta X_i \hat{\beta} \right] = 0
\]

- However, if \( q > K \), there not exists a value of \( \hat{\beta} \) that solves this system. Therefore, we define \( \hat{\beta} \) as the value that minimizes a quadratic form, or equivalently that solves the \( K \) equations:

\[
\frac{\partial m_N(\hat{\beta})'}{\partial \beta} \hat{\Omega}^{-1} m_N(\hat{\beta}) = 0
\]
Hansen-Sargan Test (cont)

- That is, by construction, $\hat{\beta}$ satisfies the conditions:

$$\left( \sum_{i=1}^{N} Z_i \Delta X_i \right) \hat{\Omega}^{-1} \left( \sum_{i=1}^{N} Z_i \left[ \Delta Y_i - \Delta X_i \hat{\beta} \right] \right) = 0$$

- However, this is compatible with having that:

$$\left\| m_N(\hat{\beta}) \right\| = \left\| \sum_{i=1}^{N} Z_i \left[ \Delta Y_i - \Delta X_i \hat{\beta} \right] \right\| \ggggggggg 0$$

- The Hansen-Sargan test is concerned with whether $\left\| m_N(\hat{\beta}) \right\|$ is significantly greater than zero.
Hansen-Sargan Test (cont)

- Under $H_0 : \mathbb{E}(Z_i \Delta U_i) = 0$, we have that $\|m_N(\hat{\beta})\| \rightarrow_p 0$ and

$$HS \equiv N \ m_N(\hat{\beta})' \ \hat{\Omega}^{-1} \ m_N(\hat{\beta}) \sim_d \chi_{q-K}^2$$

If $\xi_i \equiv Z_i \Delta \hat{U}_i$, note that:

$$HS = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i \right) \ Var(\xi_i)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i \right)$$

- Under the alternative hypothesis $H_1 : \mathbb{E}(Z_i \Delta U_i) \neq 0$, we have that $\|m_N(\hat{\beta})\| \rightarrow_p c > 0$, and

$$HS \equiv m_N(\hat{\beta})' \ \hat{\Omega}^{-1} \ m_N(\hat{\beta}) \sim_d \text{non central } \chi_{q-K}^2 \ \text{with noncentrality } c$$
Example: Employment equation. Arellano & Bond (REStud, 91)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>AB-1step</th>
<th>AB-2step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sargan (d.o.f.)</td>
<td>65.8 (25)</td>
<td>31.4 (25)</td>
</tr>
</tbody>
</table>

- Why 25 degrees of freedom?

40 moment restrictions:

\[
\{n_{i1}, w_{i1}, k_{i1}, q_{i1}\} \text{ in eq. at } t = 3, 4, 5, 6 \quad [16]
\]

\[
\{n_{i2}, w_{i2}, k_{i2}, q_{i2}\} \text{ in eq. at } t = 4, 5, 6 \quad [12]
\]
\{n_{i3}, w_{i3}, k_{i3}, q_{i3}\} \text{ in eq. at } t = 5, 6 \quad [8]

and \( E(\Delta u_{it}) = 0 \quad t = 3, 4, 5, 6 \quad [4] 

• Parameters: 4 time dummies (4 parameters); model with two lags of all the variables (11 parameters); Total of 15 parameters.
Testing for serial correlation in $u_{it}$

- Our set of moment conditions has been obtained under the assumption that $u_{it} \sim iid$. If this error term is autocorrelated, some of the moment conditions are not valid and the GMM estimator will be inconsistent. Therefore, it is important to test this assumption.

- Given our estimator $\hat{\beta}$, we can not get consistent residuals for $u_{it}$, i.e., the residual in levels is a residual for $\alpha_i + u_{it}$. And we cannot estimate consistently $\alpha_i$.

- However, we can get residuals for $\Delta u_{it}$. Our test of serial correlation will test the serial correlation of $\Delta u_{it}$.
Testing for serial correlation in $u_{it}$ (cont)

- Note that if $u_{it}$ is not serially correlated:
  
  $\mathbb{E}(\Delta u_{it} \Delta u_{it-1}) = -\sigma^2_u(t) < 0$

  $\mathbb{E}(\Delta u_{it} \Delta u_{it-j}) = 0$ for any lag $j \geq 2$

- Therefore, we can (indirectly) test for no-serial correlation in $u_{it}$ by testing for zero second-order (or higher) serial correlation in $\Delta u_{it}$.

- Testing for serial correlation of order 1 in $\Delta u_{it}$ is also useful. If we find that $\mathbb{E}(\Delta u_{it} \Delta u_{it-1}) = 0$, then OLS-FD is consistent.
Arellano-Bond test 2nd order correlation in $\Delta u_{it}$

- Let $r_{2t} \equiv \mathbb{E}(\Delta u_{it} \Delta u_{it-2})$ be the auto-covariance of order 2 at period $t$ of $\{\Delta u_{it}\}$. Its sample counterpart is:

$$\hat{r}_{2t} = \frac{1}{N} \sum_{i=1}^{N} \Delta \hat{u}_{it} \Delta \hat{u}_{it-2}$$

- We can obtain $\hat{r}_{2t}$ for any $t \in \{5, 6, ..., T\}$. Note: To test for second order serial correlation in $\Delta u_{it}$ we need a panel with at least $T = 5$ periods.

- Let $r_2 \equiv \sum_{t=5}^{T} r_{2t}$, and let $\hat{r}_2$ be its sample counterpart. Arellano & Bond (1991) derive the expression for the the asymptotic variance $Var(\hat{r}_2)$.

- Then, under $H_0: r_2 = 0$.

$$\hat{m}_2 \equiv \frac{\hat{r}_2}{se(\hat{r}_2)} \sim N(0, 1)$$
Arellano-Bond test j-th order correlation in $\triangle u_{it}$

- Let $r_{jt}$ be the auto-covariance of order $j$ at period $t$ of $\{\triangle u_{it}\}$: i.e., $r_{jt} \equiv E\left(\triangle u_{it} \triangle u_{it-j}\right)$. And its sample counterpart:

$$\hat{r}_{jt} = \frac{1}{N} \sum_{i=1}^{N} \triangle \hat{u}_{it} \triangle \hat{u}_{it-j}$$

- We can obtain $\hat{r}_{jt}$ for any $t \in \{3 + j, 6, ..., T\}$.

- Let $r_{j} \equiv \sum_{t=3+j}^{T} r_{tj}$, and let $\hat{r}_{j}$ be its sample counterpart.

- Then, under $H_0$: $r_{j} = 0$.

$$\hat{m}_{j} \equiv \frac{\hat{r}_{j}}{se(\hat{r}_{j})} \sim N(0, 1)$$
What if we reject null hypothesis \( r^2 = 0 \)?

- If \( u_{it} \) follows a \( MA(1) \): \( u_{it} = a_{it} - \theta a_{it-1} \),

- Then we have that \( y_{i,t-3} \) is a valid instrument, i.e., \( \mathbb{E}(y_{i,t-3} \Delta u_{it}) = 0 \).

- Of course, these instruments can be weak, i.e., \( corr(y_{i,t-3} \Delta y_{i,t-1}) \) could be close to zero.
What if we reject null hypothesis $r^2 = 0$? [Cont]

- If $u_{it}$ follows an $AR(1)$ process, $u_{it} = \rho u_{it-1} + a_{it}$,

- We can transform the model $y_{it} = \beta y_{it-1} + x'_{it} \delta + \alpha_i + u_{it}$, to get:

$$
\Delta y_{it} = (\rho + \beta) \Delta y_{it-1} + (-\rho) \Delta y_{it-2} + \Delta x'_{it} \delta + \Delta x'_{it-1}(-\rho \beta) + \Delta a_{it}
$$

- Note that $\mathbb{E}(\Delta y_{it-2} \Delta a_{it}) = 0$, and for $\Delta y_{it-1}$ we can use $y_{it-2}$ and $y_{it-3}$ as instruments.

- Separate identification of $\rho$ and $\beta$ could be difficult due to collinearity and weak instruments.
Test of Individual Effects in DPD models

- It is a **differential Hansen-Sargan test** of overidentifying restrictions based on two sets of comment conditions: (1) the Arellano-Bond moment conditions in first differences; (2) the OLS moment conditions in the equation in levels, which are valid only if there is not individual heterogeneity.

- The test statistics is:

\[
D = N \left[ \hat{m}'_{ALL} \hat{\Omega}^{-1}_{ALL} \hat{m}_{ALL} \right] - N \left[ \hat{m}'_{AB} \hat{\Omega}^{-1}_{AB} \hat{m}_{AB} \right]
\]

Under the null, \(D \sim_d \chi^2_K\)

- Though, typically we will reject the null, this idea can be useful to design a test for individual-specific time trends, i.e., heterogeneity in the equation in first differences.
Weak Instruments Problem in Arellano-Bond Estimator

- Even if we use all the moment conditions of the model and use the optimal GMM estimator, at the end of the day we are instrumenting $\Delta y_{it-1}$ with lagged values of $\{y_{it-j} : j \geq 2\}$.

- Therefore, if $\text{corr}(\Delta y_{it-1}, y_{i,t-j})$ is small, the optimal GMM estimator will still give us imprecise estimates of the parameters, i.e., weak instruments problem. Under weak instruments, the GMM estimator can be seriously biased in small samples.

- This is an important problem in some empirical applications of dynamic PD models.
• The solution to this potential problem should necessarily come from incorporating additional structure/assumptions into the model.

• Several approaches have been proposed to deal with this problem:

  1. Exploit stationarity assumptions: Arellano and Bover (JoE, 1995) and Blundell and Bond (JoE, 1998).

  2. Assumptions on the joint distribution of $\alpha_i$ and $x_i$: Chamberlain’s Correlated Random Effects.

3.6. ARELLANO-BOVER (JoE 1995) - BLUNDELL-BOND (JoE 1998)

• Consider the PD model

\[ y_{it} = \beta y_{i,t-1} + x_{it} \delta + \gamma_t + \alpha_i + u_{it} \]

• Suppose that \(|\beta| < 1\), and that at some period \(t^* \leq 0\) in the past (that can be individual specific), the process \(\{y_{it}\}\) visited his individual-specific mean,

\[ y_{it^*} = \mu_{y,i} \equiv \mathbb{E}(y_{it} | i) \]

• Stationarity implies that:

\[ \mu_{y,i} = \beta \mu_{y,i} + \mu_{x,i} \delta + \mathbb{E}(\gamma_t) + \alpha_i \]

and

\[ \mu_{y,i} = \frac{\alpha_i + \mu_{x,i} \delta + \mathbb{E}(\gamma_t)}{1 - \beta} \]
• Under this stationarity assumption, we have that, one period after \( t^* \):

\[
y_{i,t^*+1} = \beta \mu_{y,i} + x_{i,t^*+1} \delta + \gamma_{t^*+1} + \alpha_i + u_{i,t^*+1}
\]

\[
= \mu_{y,i} - (1 - \beta) \mu_{y,i} + x_{i,t^*+1} \delta + \gamma_{t^*+1} + \alpha_i + u_{i,t^*+1}
\]

\[
= \mu_{y,i} - (1 - \beta) \left[ \frac{\alpha_i + \mu_{x,i} \delta + \mathbb{E}(\gamma_t)}{1 - \beta} \right] + x_{i,t^*+1} \delta + \gamma_{t^*+1} + \alpha_i + u_{i,t^*+1}
\]

\[
= \mu_{y,i} + \left[ x_{i,t^*+1} - \mu_{x,i} \right] \delta + \left[ \gamma_{t^*+1} - \mathbb{E}(\gamma_t) \right] + u_{i,t^*+1}
\]

• And for any \( t > t^* \), we have that:

\[
y_{it} = \mu_{y,i} + \sum_{j=0}^{t-t^*} \beta^j \left\{ \left[ x_{i,t-j} - \mu_{x,i} \right] \delta + \left[ \gamma_{t-j} - \mathbb{E}(\gamma_t) \right] + u_{i,t-j} \right\}
\]
• Given

\[
y_{it} = \mu_{y,i} + \sum_{j=0}^{t-t^*} \beta^j \left\{ [x_{i,t-j} - \mu_{x,i}] \delta + [\gamma_{t-j} - \mathbb{E}(\gamma_t)] + u_{i,t-j} \right\}
\]

• We have that:

\[
\Delta y_{it} = \sum_{j=0}^{t-t^*} \beta^j \left\{ \Delta x_{i,t-j} \delta + \Delta \gamma_{t-j} + \Delta u_{i,t-j} \right\}
\]

• \(\Delta y_{it}\) does NOT dependent of \(\alpha_i\) !!!
• This property implies that there are valid instruments for the equation in levels.

\[ y_{it} = \beta y_{i,t-1} + x_{it} \delta + \gamma_t + \alpha_i + u_{it} \]

• If \( u_{it} \) is not serially correlated, then \( \Delta y_{i,t-1}, \Delta y_{i,t-2}, \ldots \) are not correlated with \( (\alpha_i + u_{it}) \).

• For instance, in the simple AR(1) PD model, \( y_{it} = \beta y_{i,t-1} + \alpha_i + u_{it} \), we can estimate the equation in levels by IV using \( \Delta y_{i,t-1} \) as instrument of \( y_{i,t-1} \).
This IV estimator is:

\[
\hat{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{i,t-1} y_{it}}{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{i,t-1} y_{i,t-1}}
\]

where we instrument \( y_{i,t-1} \) with \( \Delta y_{i,t-1} \) in the equation in levels.

This IV estimator does not suffer of weak instruments as \( \beta \to 1 \).

**Note:** The ABover and BB estimators are GMM estimators more efficient that this IV estimator. But it is convenient to look at the properties of this IV to understand why ABover and BB solve the weak instruments problem.
• This IV estimator (IV version of Blundell and Bond) does not suffer of a weak instruments problem when $\beta \to 1$.

• To see the difference between this IV estimator and the IV in AH, consider the auxiliary first-step estimations for the two IV estimators.

• For **AH estimator**, we instrument $\Delta y_{i,t-1}$ with $y_{i,t-2}$. In the first step of this IV (2SLS) estimator, the auxiliary regression is:

$$
\Delta y_{it} = \pi_0 + \pi y_{i,t-1} + e_{it}
$$

And the OLS estimator of $\alpha$ is such that (**Exercise**):

$$
p\lim \hat{\pi}_{OLS} = \frac{\text{cov}(\Delta y_{it}, y_{i,t-1})}{\text{var}(y_{i,t-1})} = \frac{\left(\frac{\sigma_u^2}{\sigma_\alpha^2}\right) (1 - \beta)^2}{1 + \beta + \left(\frac{\sigma_u^2}{\sigma_\alpha^2}\right) (1 - \beta)}
$$

That goes to zero as $\beta$ goes to one. Weak instruments problem!
• This IV estimator (IV version of Blundell and Bond) does not suffer of a weak instruments problem when $\beta \to 1$. [Cont.]

• For BB estimator, we instrument $y_{i,t-1}$ with $\Delta y_{i,t-1}$. In the first step of this IV (2SLS) estimator, the auxiliary regression is:

$$y_{it} = \pi_0 + \pi \Delta y_{it} + r_{it}$$

And the OLS estimator of $\pi$ is such that (Exercise):

$$p \lim \hat{\pi}_{OLS} = \frac{\text{cov}(y_{it}, \Delta y_{it})}{\text{var}(\Delta y_{it})} = \frac{1}{2 (1 + \beta)}$$

which does not go to zero as $\beta$ goes to one.
Asymptotic Variance of IV version of Blundell and Bond

- As we did in the case of the Anderson-Hsiao estimator, it is interesting to obtain the variance of this estimator in terms of the model parameters. Notice that:

\[
\sqrt{N} \left( \hat{\beta} - \beta \right) = \frac{N^{-1/2} \sum_{i=1}^{N} \Delta y_{i,t-1} (\alpha_i + u_{it})}{N^{-1} \sum_{i=1}^{N} \Delta y_{i,t-1} y_{i,t-1}}
\]

\[
\rightarrow_d N \left( 0, Var \left( \Delta y_{i,t-1} (\alpha_i + u_{it}) \right) \right) / \mathbb{E} \left( \Delta y_{i,t-1} y_{i,t-1} \right)
\]

- Therefore,

\[
Var \left( \hat{\beta} \right) = \frac{1}{N (T - 2)} \frac{Var \left( \Delta y_{i,t-1} (\alpha_i + u_{it}) \right)}{\mathbb{E} \left( \Delta y_{i,t-1} y_{i,t-1} \right)^2}
\]
Given that:

\[ \text{Var} \left( \Delta y_{i,t-1} (\alpha_i + u_{it}) \right) = 2\sigma_u^2 \left( \sigma_u^2 + \sigma_{\alpha}^2 \right) / (1 + \beta), \]

\[ \mathbb{E} \left( \Delta y_{i,t-1} y_{i,t-1} \right) = \sigma_u^2 / (1 + \beta), \]

we have that:

\[
\text{Var} \left( \hat{\beta} \right) = \frac{1}{N (T - 2)} \frac{2\sigma_u^2 \left( \sigma_u^2 + \sigma_{\alpha}^2 \right) / (1 + \beta)}{\sigma_u^4 / (1 + \beta)^2} \\
= \frac{1}{N (T - 2)} \frac{2 (1 + \beta) \left( 1 + \frac{\sigma_{\alpha}^2}{\sigma_u^2} \right)}
\]

In contrast with the AH and AB estimators, the variance of this estimator does not goes to infinity as \( \beta \) goes to one.
Example: Comparing of AH-IV and BB-IV

- Take the sample numerical example that we considered above for the AH estimator.
  
  $$\beta = 0.6 ; \quad \sigma_{\alpha}/\sigma_{u} = 6 ; \quad N (T - 2) = 3600$$

- For AH-IV estimator, we have:
  
  $$sd\left(\hat{\beta}_{AH}\right) = 0.567 \quad \text{and} \quad 95\% \ CI \ for \ \beta = \ [-0.51 , \ 1.71]$$

- For BB-IV estimator, we have:
  
  $$sd\left(\hat{\beta}_{BB-IV}\right) = 0.18 \quad \text{and} \quad 95\% \ CI \ for \ \beta = \ [0.25 , \ 0.95]$$

  that is a much narrower CI.
GMM estimation with the two sets of moment conditions (AB mc’s and BB mc’s)

System GMM

The stationarity restrictions implies the following moment conditions:

For: \( t = 3, 4, \ldots, T \)
\( s = 2, 3, \ldots, t - 1 \)
\[ \mathbb{E}(\Delta y_{is} [y_{it} - \beta y_{it-1}]) = 0 \]

These are \((T - 2)(T - 1)/2\) moment conditions.

The AB moment conditions are:

For: \( t = 3, 4, \ldots, T \)
\( s = 1, 2, \ldots, t - 2 \)
\[ \mathbb{E}(y_{is} [\Delta y_{it} - \beta \Delta y_{it-1}]) = 0 \]

These are \((T - 2)(T - 1)/2\) moment conditions.
GMM estimation with the two sets of moment conditions (AB mc’s and BB mc’s)

System GMM

- Blundell and Bond propose a GMM estimator that combines these moment conditions with the AB moment conditions. Their Monte Carlo experiments show that this estimator has smaller finite sample variance and bias than the GMM based on AB moment conditions.
The *xtabond* and *xtabond2* commands in STATA

- *xtabond2* is more complete and general than *xtabond*

- Syntax:

  \[
  \text{xtabond2 } \text{depvar } x_1 \ x_2 \ldots \ x_K, \\
  gmm(z_1 \ldots z_q \text{ lag}(l_{\text{min}} \ \ l_{\text{max}} )) \\
  iv(v_1 \ldots v_q), \text{ twostep robust nolevel}
  \]

- With the option "**nolevel**" only Arellano-Bond moment conditions are used.

- Without the option "**nolevel**" the system GMM is implemented, i.e., both Arellano-Bond and Blundell-Bond moment conditions are used.
Blundell and Bond (Econometric Reviews, 2000) "GMM ESTIMATION WITH PERSISTENT PANEL DATA: AN APPLICATION TO PRODUCTION FUNCTIONS"

- Estimation of Cobb-Douglas Production function:

\[ y_{it} = \beta_n n_{it} + \beta_k k_{it} + \alpha_i + \gamma_t + u_{it} \]

\( y \equiv \ln(Output); \ n \equiv \ln(Employment); \ k \equiv \ln(Capital) \). The unobservables \( \alpha_i \) and \( u_{it} \) represent firms’ productivity differences which are permanent and transitory, respectively.

- Input demands depend on firms’ productivity, \( \alpha_i \) and \( u_{it} \), on input prices, \( w_{it} \), but also on the level of inputs at previous period (i.e., adjustment costs).

\[
\begin{align*}
n_{it} &= f_l \left( n_{i,t-1}, k_{i,t-1}, \alpha_i, u_{it}, w_{it} \right) \\
k_{it} &= f_k \left( n_{i,t-1}, k_{i,t-1}, \alpha_i, u_{it}, w_{it} \right)
\end{align*}
\]
• It is clear that regressors are not strictly exogenous because current inputs depend on current and lagged values of the transitory shock $u$. Therefore, OLS in levels, OLS in FD and WG estimators are inconsistent.
Empirical applications of PF that attempt to control for unobserved heterogeneity and simultaneity by using Arellanor-Bond GMM estimators which take have tended to produce unsatisfactory results in this context: i.e., very low estimates of parameters (and returns to scale) and quite imprecise.

Blundell and Bond argue that these problems are related to the weak correlations that exist between the current growth rates of capital and employment and the lagged levels of these variables: weak instruments problem in the model in first differences.

They propose using the system estimator: AB + BB instruments.

They also find serial correlation in the transitory shock $u_{it}$ [m2 test clearly rejects $u_{it}$ is not serially correlated]. Therefore, they consider an AR(1) process:

$$u_{it} = \rho \, u_{i,t-1} + a_{it}$$
Model in quasi-first differences:

\[ y_{it} = \rho y_{it-1} + \beta_n n_{it} + (-\rho\beta_n)n_{it-1} + \beta_k k_{it} + (-\rho\beta_k)k_{it-1} \]

\[ + (1 - \rho)\alpha_i + (\gamma_t - \rho\gamma_{t-1}) + a_{it} \]

\( \Delta y_{it-1}, \Delta y_{it-2}, \Delta n_{it-1}, \Delta n_{it-2}, \Delta k_{it-1}, \) and \( \Delta k_{it-2} \) are valid instruments in this equation [BB].

Model in quasi-first + first differences:

\[ \Delta y_{it} = \rho \Delta y_{it-1} + \beta_n \Delta n_{it} + (-\rho\beta_n)\Delta n_{it-1} + \beta_k \Delta k_{it} + (-\rho\beta_k)\Delta k_{it-1} \]

\[ + (\Delta \gamma_t - \rho\Delta \gamma_{t-1}) + \Delta a_{it} \]

\( y_{it-2}, y_{it-3}, n_{it-2}, n_{it-3}, k_{it-2}, \) and \( k_{it-3} \) are valid instruments in this equation [BB].
Production Function: Blundell & Bond (ER 2000). Tables 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OLS-Levels</th>
<th>WG</th>
<th>AB-GMM</th>
<th>SYS-GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_n$</td>
<td>0.538</td>
<td>0.488</td>
<td>0.515</td>
<td>0.479</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.030)</td>
<td>(0.099)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>$\beta_k$</td>
<td>0.266</td>
<td>0.199</td>
<td>0.225</td>
<td>0.492</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.033)</td>
<td>(0.126)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.964</td>
<td>0.512</td>
<td>0.448</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.022)</td>
<td>(0.073)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>Sargan (p-value)</td>
<td>-</td>
<td>-</td>
<td>0.073</td>
<td>0.032</td>
</tr>
<tr>
<td>m2</td>
<td>-</td>
<td>-</td>
<td>-0.69</td>
<td>-0.35</td>
</tr>
<tr>
<td>Constant RS (p-v)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.006</td>
<td>0.641</td>
</tr>
</tbody>
</table>
THANKS TO ALL THE STUDENTS WHO HAVE GREATELY CONTRIBUTED TO IMPROVE THESE NOTES