ECONOMETRICS II (ECO 2401)

Victor Aguirregabiria (Winter 2018)

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

3.4. Anderson-Hsiao (JE, 1982)

3.5. Arellano-Bond (REStud, 1991)

3.6. Blundell-Bond (JE, 1998)

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.

$$\text{Data} = (\mathbf{x}_{it} : i = 1, 2, ..., N; \quad t \in \{1, 2, ..., T\})$$

$$\begin{array}{rcl} \mathbf{x}_{it} & : & K \times 1 \text{ vector of observed variables} \\ \\ i & : & \text{subindex for individuals} \\ t & : & \text{subindex for time} \\ \\ N & = & \# \text{ of individuals} \\ T_i & = & \# \text{ of periods of time the individual } i \text{ is observed} \end{array}$$

• 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.

$$
\begin{array}{ll}
i = 1 & \qquad\qquad \vdash - - \quad - - - - - - \dashv \\
i = 2 & \vdash - - - - \dashv \\
i = 3 & \qquad\qquad\qquad\qquad\qquad\qquad \vdash - - \quad - - - - - - \dashv \\
i = 4 & \qquad\qquad \vdash - - \quad - - - - - - - \quad - - - - - - \dashv
\end{array}
$$

• 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:

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

• ε_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 $cov(l_i, \varepsilon_i) > 0$ and $cov(k_i, \varepsilon_i) > 0$.

• OLS estimation will provide inconsistent estimates of β_1 and β_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:

- α_i represents time-invariant unobserved inputs, e.g., quality of land;

- γ_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 α_i and γ_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 Δu_{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 Δu_{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:

- ΔL_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;

- γ_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: $\varepsilon_t \equiv \gamma_t + u_t$. While u_t is not correlated with the regressors, the error γ_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:

$$y_{it} \;=\; f\left(\mathbf{x}_{it},\ \varepsilon_{it},\ \beta\right)$$

where:

> $f(.)$ is a known function;
>
> β is a vector of unknown parameters.
>
> $\mathbf{x}_{it}$ is a vector of observable explanatory variables;
>
> ε_{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 α's and γ's as parameters to estimate.

• We can include as explanatory variables $(N - 1)$ individual dummies associated to α's and $(T - 1)$ time dummies associated to γ's.

• The number of parameters α's increases at the same rate as N as $N \to \infty$. We cannot estimate these parameters consistently.

• Inconsistency of our estimator of α's may contaminate / bias our estimation of the parameters of interest β.

INCIDENTAL PARAMETERS PROBLEM [more generally]

• Given a model with vector of parameters θ and a random sample with size n, we say that the model has an *incidental parameters problem* if:

$$\text{As } n \rightarrow \infty \text{, we have that } \textbf{dim}(\theta) \rightarrow \infty$$

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

• Suppose that $\theta = (\alpha, \beta)$, such that as $n \rightarrow \infty$,

$$\textbf{dim}(\alpha) \rightarrow \infty, \text{ but } \textbf{dim}(\beta) = K \text{ for any value of } n$$

Can we obtain a consistent estimator of β despite our estimation of α 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: $$y_{it} = h(\mathbf{x}_{it},\ \beta) + \varepsilon_{it}$$

Non additive separable unobservable: e.g., $$y_{it} = \max\{\mathbf{x}_{it}'\,\beta + \varepsilon_{it}\ ,\ 0\}$$

• (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 $\{\mathbf{x}_{it}\}$ and the transitory shock $\{u_{it}\}$.

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

$$\mathbb{E}\left(u_{it}\ \mathbf{x}_{it+s}\right) = 0 \qquad \text{for any } s \in \{..., -2, -1, 0, +1, +2, ...\}$$

• **Dynamic models** include predetermined or lagged endogenous variables. For instance, y_{it-1} is included in $\mathbf{x}_{it}$. In these models, $\mathbf{x}_{it+1}$ includes y_{it}, and therefore

$$\mathbb{E}\left(u_{it}\ \mathbf{x}_{it+1}\right) \neq 0$$

- 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:

$$y_{it} = \mathbf{x}_{it}' \beta + \varepsilon_{it}$$

where: $\mathbf{x}_{it}$ is a $K \times 1$ of regressors. ε_{it} is the error term.

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

$$\varepsilon_{it} = \alpha_i + \gamma_t + u_{it}$$

γ_t is a common aggregate effect;

α_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 $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, ..., \gamma_T\}$ using time dummies.

- Let $D_t^{(s)}$ be the time-dummy for period s such that:

$$D_{it}^{(s)} = 1 \quad \text{if } t = s; \qquad \text{and} \quad D_{it}^{(s)} = 0 \quad \text{if } t \neq s$$

- Then,

$$\gamma_t = \sum_{s=1}^{T} D_{it}^{(s)} \gamma_s \quad = \quad [D_{it}^{(1)}, ..., D_{it}^{(T)}] \, \boldsymbol{\gamma} \quad = \quad \mathbf{D}_{it}' \, \boldsymbol{\gamma}$$

- Therefore, the model:

$$y_{it} \quad = \quad [\mathbf{x}_{it} \; , \mathbf{D}_{it}]' \begin{bmatrix} \beta \\ \boldsymbol{\gamma} \end{bmatrix} + \alpha_i + u_{it}$$

Time dummies (cont.)

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

$$y_{it} = \mathbf{x}_{it}' \beta + \alpha_i + u_{it}$$

But it should be understood that $\mathbf{x}_{it}$ includes the time-dummy variables, and β includes the time-effect parameters $\boldsymbol{\gamma}$.

• This notation applies to large N and T small. In this case, $\boldsymbol{\gamma}$ and β are not subject to the incidental parameters problem [in cotrast to $\boldsymbol{\alpha}$].

• When T is large and N is small we can include individual dummies in $\mathbf{x}_{it}$. In that case, $\boldsymbol{\alpha}$ and β are not subject to the incidental parameters problem [in cotrast to $\boldsymbol{\gamma}$].

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.

$$\begin{array}{rcl} y_{i1} & = & \mathbf{x}'_{i1}\,\beta + \alpha_i + u_{i1} \\ y_{i2} & = & \mathbf{x}'_{i2}\,\beta + \alpha_i + u_{i2} \\ \vdots & & \vdots \\ y_{iT} & = & \mathbf{x}'_{iT}\,\beta + \alpha_i + u_{iT} \end{array}$$

• We can represent this system in vector form as:

$$\mathbf{Y}_i \;=\; \mathbf{X}_i\,\beta + \mathbf{1}\,\alpha_i + \mathbf{U}_i$$

where:

$\mathbf{Y}_i$ and $\mathbf{U}_i$ are $T \times 1$ vectors; and $\mathbf{1}$ is a $T \times 1$ vector of $1's$;

$\mathbf{X}_i$ 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:

$$\mathbf{Y}_i \;=\; \mathbf{X}_i\,\beta + \mathbf{d}_i\,\alpha_i + \mathbf{U}_i$$

where:

$$\underset{T\times 1}{\mathbf{Y}_i} = \begin{bmatrix} d_{i1}\ y_{i1} \\ d_{i2}\ y_{i2} \\ \vdots \\ d_{iT}\ y_{iT} \end{bmatrix}; \quad \underset{T\times K}{\mathbf{X}_i} = \begin{bmatrix} d_{i1}\ \mathbf{x}'_{i1} \\ d_{i2}\ \mathbf{x}'_{i2} \\ \vdots \\ d_{iT}\ \mathbf{x}'_{1T} \end{bmatrix}; \quad \underset{T\times 1}{\mathbf{d}_i} = \begin{bmatrix} d_{i1} \\ d_{i2} \\ \vdots \\ d_{iT} \end{bmatrix}$$

The expressions of the different estimators we will see below apply to this definition of $\mathbf{Y}_i$, $\mathbf{X}_i$, and $\mathbf{d}_i$.

- **Assumption:** **Strict exogeneity (Static Model)**

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

- **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 $y_{i,t-1} \subset \mathbf{x}_{it}$. The model establishes that $y_{i,t-1}$ depends on $u_{i,t-1}$. Therefore, it is clear that:

$$\mathbb{E}[\mathbf{x}_{it}\ u_{i,t-1}] \neq 0$$

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.

$$\mathbb{E}\left(\mathbf{x}_{it}\ \alpha_i\right) \neq 0$$

• 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 $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ and α_i.

• $CDF(\alpha_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ is completely unrestricted. In this sense, the FE model is **nonparametric** with respect the distribution $CDF(\alpha_i \mid \mathbf{x}_i)$.

• 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 $CDF(\alpha_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$.

• The stronger restriction is that α_i is independent of $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$ and $iid(0, \sigma_\alpha^2)$. Some textbooks define RE in this restrictive way.

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

$$\alpha_i = \lambda_0 + \mathbf{x}_{i1}' \, \lambda_1 + ... + \mathbf{x}_{iT}' \, \lambda_T + e_i$$

where e_i is independent of $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$. Based on this assumption, we estimate the parameters β and $\lambda's$. It is a **parametric approach** because it depends on a parametric assumption on the distribution of $\{x_{i1}, x_{i2}, ..., x_{iT}\}$ and α_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 $\mathbb{E}[\mathbf{x_{it}}\ \alpha_\mathbf{i}] = \mathbf{0}$

• 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:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\ \mathbf{x}'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\ y_{it}\right) \\
&= \left(\sum_{i=1}^{N}\ \mathbf{X}'_i\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\ \mathbf{X}'_i\mathbf{Y}_i\right)
\end{aligned}
$$

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 $\mathbb{E}[\mathbf{x_{it}}\ \alpha_\mathbf{i}] = \mathbf{0}$ (cont.)

• The pooled OLS estimator is not efficient because the error term ε_{it} is serially correlated:

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

• The asymptotically efficient estimator is the GLS.

• To obtain the efficient GLS estimator of β it is convenient to write the model in matrix form:

$$\mathbf{Y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$$

GLS ESTIMATOR WHEN $\mathbb{E}[\mathbf{x_{it}}\ \alpha_\mathbf{i}] = 0$ (cont.)

where:

$$\underset{NT \times 1}{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} \quad ; \quad \underset{NT \times K}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \quad ; \quad \underset{NT \times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}$$

• Let $\mathbf{V}_\varepsilon$ be the variance matrix of $\boldsymbol{\varepsilon}$. We know that the GLS estimator can be defined as the OLS estimator of the following transformed model:

$$\left(\mathbf{V}_\varepsilon^{-1/2}\ \mathbf{Y}\right) = \left(\mathbf{V}_\varepsilon^{-1/2}\ \mathbf{X}\right)\ \beta + \left(\mathbf{V}_\varepsilon^{-1/2}\ \boldsymbol{\varepsilon}\right)$$

We now derive the expressions of $\mathbf{V}_\varepsilon$, $\mathbf{V}_\varepsilon^{-1/2}$, and of the transformed variables $\mathbf{V}_\varepsilon^{-1/2}\ \mathbf{Y}$ and $\mathbf{V}_\varepsilon^{-1/2}\ \mathbf{X}$.

GLS ESTIMATOR WHEN $\mathbb{E}[\mathbf{x_{it}}\ \alpha_\mathbf{i}] = 0$ (cont.)

- We have

$$\mathbf{V}_\varepsilon = \mathbb{E}[\boldsymbol{\varepsilon}\ \boldsymbol{\varepsilon}'] = \mathbb{E}\begin{bmatrix} \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2' & \dots & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_N' \\ & \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2' & \dots & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_N' \\ & & \ddots & \vdots \\ & & & \boldsymbol{\varepsilon}_N\boldsymbol{\varepsilon}_N' \end{bmatrix}$$

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

$$\mathbf{V}_\varepsilon = \begin{bmatrix} \Omega & 0 & \dots & 0 \\ 0 & \Omega & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \Omega \end{bmatrix} = \mathbf{I}_N \otimes \mathbf{\Omega}$$

where Ω is the $T \times T$ matrix $\mathbb{E}[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i']$.

• **GLS and FGLS when $\mathbf{u}$ is not i.i.d.** $\mathbf{V}_\varepsilon = \mathbf{I}_N \otimes \mathbf{\Omega}$ where for any individual i the variance-covariance matrix

$$\mathbf{\Omega} = \mathbb{E}\begin{bmatrix} \varepsilon_{i1}^2 & \varepsilon_{i1}\varepsilon_{i2} & \cdots & \varepsilon_{i1}\varepsilon_{iT} \\ & \varepsilon_{i2}^2 & \cdots & \varepsilon_{i2}\varepsilon_{iT} \\ & & \ddots & \vdots \\ & & & \varepsilon_{iT}^2 \end{bmatrix}$$

is an unrestricted.

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

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

• This FGLS estimator can be obtained without any restriction on the structure of the variance matrix $\mathbf{\Omega}$.

• Let $\widehat{\boldsymbol{\varepsilon}}_i = (\widehat{\varepsilon}_{i1}, \widehat{\varepsilon}_{i2}, ..., \widehat{\varepsilon}_{iT})'$ be the vector of OLS residuals for individual i. Then, a consistent estimator of $\boldsymbol{\Omega}$ is:

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\varepsilon}}_i \; \widehat{\boldsymbol{\varepsilon}}_i'$$

The FGLS estimator uses this estimator of Ω.

ESTIMATION WHEN $\mathbb{E}[\mathbf{x_{it}}\ \alpha_\mathbf{i}] \neq \mathbf{0}$

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

- 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:

$$\Delta y_{it} = \Delta \mathbf{x}_{it}' \beta + \Delta u_{it}$$

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

• Strict exogeneity of x_{it} implies that $\mathbb{E}[\Delta \mathbf{x}_{it} \; \Delta u_{it}] = 0$:

$$\begin{array}{ccccc} \mathbb{E}[\Delta \mathbf{x}_{it} \Delta u_{it}] = & \mathbb{E}\left[\mathbf{x}_{it} u_{it}\right] - & \mathbb{E}\left[\mathbf{x}_{it} u_{i,t-1}\right] - & \mathbb{E}\left[\mathbf{x}_{i,t-1} u_{it}\right] + & \mathbb{E}\left[\mathbf{x}_{i,t-1} u_{i,t-1}\right] \\ = & 0+ & 0+ & 0+ & 0 \end{array}$$

• Therefore, OLS in FD is a consistent estimator.

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it} \; \Delta \mathbf{x}_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it} \; \Delta y_{it} \right)$$

- 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 β 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:

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

• The WG estimator is just the OLS estimator of the WG transformation.

$$\hat{\beta}_{WG} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)(y_{it} - \bar{y}_i) \right)$$

Consistency of WG estimator in Static-Linear PD

• This estimator is consistent if there is not perfect collinearity in $(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)$ and $\mathbb{E}\left((\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\ (u_{it} - \bar{u}_i)\right) = 0$.

• Note that:

$$\begin{array}{ccccc} \mathbb{E}\left((\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\ (u_{it} - \bar{u}_i)\right) = & \mathbb{E}\left[\mathbf{x}_{it}u_{it}\right] - & \mathbb{E}\left[\mathbf{x}_{it}\bar{u}_i\right] & -\mathbb{E}\left[\overline{\mathbf{x}}_i u_{it}\right] & +\mathbb{E}\left[\overline{\mathbf{x}}_i\bar{u}_i\right] \\ = & 0+ & 0+ & 0+ & 0 \end{array}$$

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

• However, the estimator of α_i is not consistent for fixed T:

$$\begin{aligned} p \lim_{N \longrightarrow \infty} \widehat{\alpha_i} &= \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}'_{it}\left[p \lim_{N \longrightarrow \infty} \widehat{\beta}\right]\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}'_{it}\beta\right) \\ &= \alpha_i + \bar{u}_i \neq \alpha_i \end{aligned}$$

(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 β.

• We can write the model in vector form as:

$$\mathbf{Y} = \mathbf{X}\,\beta + \mathbf{D}\,\alpha + \mathbf{U} = (\mathbf{X} \;\vdots\; \mathbf{D}) \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + \mathbf{U}$$

• $\mathbf{D}$ is a $NT \times N$ matrix of dummy variables, one for each individual.

• The $i - th$ column of matrix $\mathbf{D}$ contains the observations of the dummy variable for individual i, i.e., 1 if the observation belongs to i and 0 otherwise.

• α 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:

$$
\begin{aligned}
\begin{pmatrix} \widehat{\beta}_{LSDV} \\ \widehat{\alpha}_{LSDV} \end{pmatrix} &= \left[(\mathbf{X} \ \vdots \ \mathbf{D})'(\mathbf{X} \ \vdots \ \mathbf{D})\right]^{-1} \left[(\mathbf{X} \ \vdots \ \mathbf{D})'\mathbf{Y}\right] \\
&= \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{D}'\mathbf{Y} \end{bmatrix}
\end{aligned}
$$

• 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 $\mathbf{D}'\mathbf{D}$ and $\mathbf{D}'\mathbf{X}$.

• By **Frish-Waugh Theorem,** the OLS estimator of β can be written:

$$\widehat{\beta}_{LSDV} = \left[\mathbf{X}'\,\mathbf{M_D}\,\mathbf{X}\right]^{-1}\left[\mathbf{X}'\,\mathbf{M_D}\,\mathbf{Y}\right]$$

where $\mathbf{M_D}$ is the idempotent matrix:

$$\begin{aligned}
\mathbf{M_D} &= \mathbf{I}_{NT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \\
&= (\mathbf{I}_N \otimes \mathbf{I}_T) - \frac{1}{T}(\mathbf{I}_N \otimes \mathbf{1}\mathbf{1}') \\
&= \mathbf{I}_N \otimes \left(\mathbf{I}_T - \frac{1}{T}\mathbf{1}\mathbf{1}'\right)
\end{aligned}$$

LSDV estimator (cont.)

• When we pre-multiply $\mathbf{Y}$ and $\mathbf{X}$ by $\mathbf{M_D}$, these variables are tramsformed in deviations with respect to individual means: $\mathbf{M_D}\,\mathbf{Y} = \mathbf{Y}^*$ and $\mathbf{M_D}\,\mathbf{X} = \mathbf{X}^*$:

$$y^*_{it} = y_{it} - \bar{y}_i \qquad ; \qquad \mathbf{x}^*_{it} = \mathbf{x}_{it} - \overline{\mathbf{x}}_i$$

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

• Since $\mathbf{M_D}$ is an idempotent matrix, we can write:

$$\widehat{\beta}_{LSDV} = \left[\mathbf{X}^{*\prime}\,\mathbf{X}^*\right]^{-1}\left[\mathbf{X}^{*\prime}\,\mathbf{Y}^*\right]$$

• Therefore, the **least squares dummy variables estimator** of β is **EQUIVALENT** to the OLS estimator of the transformed model:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)'\beta + \varepsilon^*_{it}$$

LSDV estimator (cont.)

• 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 α_i is:

$$\widehat{\alpha_{i,LSDV}} = T^{-1} \sum_{t=1}^{T} \left(y_{it} - \mathbf{x}_{it}' \, \hat{\beta} \right)$$

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 α grow at the same rate as N, and we have only T observations for each α_i. Therefore, $\widehat{\alpha_{LSDV}}$ is NOT consistent.

• In general, the inconsistency of $\widehat{\alpha_{LSDV}}$ affects the estimation of β 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 β despite our estimator of α 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 α_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 α_i. Then, how is it possible that we get a consistent estimator of β?

• The answer is that the estimator of α_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:

$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}_{it}\ \beta + \tilde{\varepsilon}_{it}$$

$\widetilde{y}_{it}$, $\widetilde{\mathbf{x}}_{it}$, $\tilde{\varepsilon}_{it}$ are transformations of the original variables y_{it}, $\mathbf{x}_{it}$, ε_{it}.

- This representation includes as particular cases:

  - Model in levels: $\widetilde{y}_{it} = y_{it}$

  - Model in FD: $\widetilde{y}_{it} = \Delta y_{it}$

  - Within Groups: $\widetilde{y}_{it} = y_{it} - \bar{y}_i$

  - Balestra-Nerlove: $\widetilde{y}_{it} = y_{it} - \theta \bar{y}_i$

ROBUST STANDARD ERRORS (cont.)

• In vector form:

$$\underset{(T\times 1)}{\widetilde{\mathbf{Y}}_i} = \underset{(T\times K)}{\widetilde{\mathbf{X}}_i} \underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{\widetilde{\varepsilon}_i}$$

• The OLS estimator in this transformed model is:

$$\widehat{\beta} = \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \, \widetilde{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i \, \widetilde{\mathbf{Y}}_i\right)$$

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:

$$\mathbf{V}(\widehat{\beta}) = \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \, \widetilde{\mathbf{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \, \mathbb{E}\left[\widetilde{\varepsilon}_i \, \widetilde{\varepsilon}_i' \mid \widetilde{\mathbf{X}}_i\right] \, \widetilde{\mathbf{X}}_i\right) \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \, \widetilde{\mathbf{X}}_i\right)^{-1}$$

ROBUST STANDARD ERRORS (cont.)

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

$$\widehat{\mathbf{V}}(\widehat{\beta}) = \left(\sum_{i=1}^{N} \widetilde{\boldsymbol{X}}_i' \, \widetilde{\boldsymbol{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{\boldsymbol{X}}_i' \, \widehat{\boldsymbol{\varepsilon}}_i \, \widehat{\boldsymbol{\varepsilon}}_i' \, \widetilde{\boldsymbol{X}}_i\right) \left(\sum_{i=1}^{N} \widetilde{\boldsymbol{X}}_i' \, \widetilde{\boldsymbol{X}}_i\right)^{-1}$$

where $\widehat{\boldsymbol{\varepsilon}}_i = \widetilde{\boldsymbol{Y}}_i - \widetilde{\boldsymbol{X}}_i\widehat{\beta}$.

• Note that $cov(u_{it}, u_{is})$ and $var(u_{it})$ 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 β. Using cross-sectional data, estimates of β are very small. We are concerned with $cov(W_{it}, \alpha_i) \neq 0$.

And Example (Labor Supply) [Cont.]

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

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

• Main results:

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

- Very important to use robust s.e.;

- Substantial loss of precision in FE estimation

- Estimates

Labor Supply Dependent variable: log(Annual Hours of Work)				
Parameter	**Pool OLS**	**Within**	**FD**	**FGLS**
β	0.083	0.168	0.109	0.120
Robust s.e.	(0.030)	(0.085)	(0.084)	(0.052)
Default s.e.	(0.009)	(0.020)	(0.021)	(0.014)
σ_α		0.181		
σ_u		0.232		
θ		0.624		0.586
Number Obs.	5,320	5,320	4,788	5,320

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 α_i and $\mathbf{x}$.

• However, there are two relevant issues with FE estimators:

(a) Variance of FE cab 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 $\mathbf{x}_{it}$ in the sum of Within Groups and Between Groups variance:

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\overline{x})^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}([x_{it}-\overline{x}_i]+[\overline{x}_i-\overline{x}])^2 \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(x_{it}-\overline{x}_i)^2+\sum_{i=1}^{N}T_i(\overline{x}_i-\overline{x})^2 \\
SS_{TOTAL} \qquad & \qquad SS_{WG}+SS_{BG}
\end{aligned}
$$

• Tyypically, 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. Typcally, this represents a small fraction of SS_{TOTAL}.

• Therefore, Var(WG) $>>>$ Var(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 α_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 α_i, $cov(x_{it}^*, \alpha_i) = \gamma > 0$.

• 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:

$$Bias(OLS) = \frac{cov(x,\varepsilon)}{var(x_{it})} = \frac{\gamma - \beta \; \sigma_e^2}{SS_{TOTAL}/NT}$$

- 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:

$$Bias(WG) = \frac{cov((x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i))}{var((x_{it} - \bar{x}_i))} = \frac{-\beta \; \sigma_e^2}{SS_{WG}/NT}$$

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., $SS_{TOTAL} >> SS_{WG}$), 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 $\mathbb{E}(\mathbf{x}_{it}\,\alpha_i) = 0$. We do not want to get rid of the BG variation of the regressots if this variation is not correlated with α_i. **[Hausman test]**

(b) An estimator that allows for $\mathbb{E}(\mathbf{x}_{it}\,\alpha_i) \neq 0$ 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

- $H_0 : \mathbb{E}\left(\mathbf{x}_{it}\ \alpha_i\right) = 0$ and $H_1 : \mathbb{E}\left(\mathbf{x}_{it}\ \alpha_i\right) \neq 0$

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

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

- Under $[H_1 : \mathbb{E}\left(\mathbf{x}_{it}\ \alpha_i\right) \neq 0] \Rightarrow \begin{cases} \hat{\beta}_{FGLS} \text{ is inconsistent} \\ \hat{\beta}_{WG} \text{ is still consistent} \end{cases}$

HAUSMAN TEST OF ENDOGENOUS INDIVIDUAL EFFECTS

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

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

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

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

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

• Therefore,

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

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

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

• Our main concern is the correlation between α_i and the regressors $\mathbf{x}$.

• Suppose that we make the following assumption about the joint distribution of α_i and $\mathbf{x}$:

$$\alpha_i = \mathbf{x}_{i1}' \, \lambda_1 + ... + \mathbf{x}_{iT}' \, \lambda_T + e_i$$

where e_i is independent of $\mathbf{x}_i \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$.

$$e_i \perp \mathbf{x}_i$$

CHAMBERLAIN'S CORRELATED RANDOM EFFECTS ESTIMATOR

- Based on this assumption, we have:

$$\begin{aligned}
y_{it} &= \mathbf{x}'_{i1}\ \lambda_1 + ... + \mathbf{x}'_{i1}\,[\lambda_t + \beta] + ...\mathbf{x}'_{iT}\ \lambda_T + e_i + u_{it} \\
&= [\mathbf{x}'_{i1}\ , \mathbf{x}'_{i2}\ , ..., \mathbf{x}'_{it}, ..., \mathbf{x}'_{iT}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_t + \beta \\ \vdots \\ \lambda_T \end{bmatrix} + e_i + u_{it} \\
&= \mathbf{x}'_i\ \Pi_t + u^*_{it}
\end{aligned}$$

where $\mathbf{x}'_i \equiv (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, ..., \mathbf{x}'_{iT})$ is $1 \times KT$ vector of regressors, Π_t is a $KT \times 1$ vector of parameters, 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):

$$\begin{aligned} y_{i1} &= \mathbf{x}_i' \, \Pi_1 + u_{i1}^* \\ y_{i2} &= \mathbf{x}_i' \, \Pi_2 + u_{i2}^* \\ \vdots & \quad \vdots \\ y_{iT} &= \mathbf{x}_i' \, \Pi_T + u_{iT}^* \end{aligned}$$

• 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 β, not Π's.

CHAMBERLAIN'S CORRELATED RANDOM EFFECTS ESTIMATOR

• Note that we can represent the relationship between Π's and λ's and β as a linear regression-like equation:

$$\mathbf{\Pi} \equiv \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_T \end{bmatrix} = \begin{bmatrix} \mathbf{I}_K & \mathbf{I}_K & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{I}_K & \mathbf{0} & \mathbf{I}_K & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_K \end{bmatrix} \begin{bmatrix} \beta \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_T \end{bmatrix}$$

or in a compact form:

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

where $\mathbf{A}$ 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 $\widehat{\mathbf{\Pi}}$, we can obtain a consistency estimator of λ's and β as:

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} = \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \, \widehat{\mathbf{\Pi}}$$

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

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} = \left(\mathbf{A}' \, \widehat{\mathbf{V}}(\widehat{\mathbf{\Pi}})^{-1} \, \mathbf{A} \right)^{-1} \left(\mathbf{A}' \, \widehat{\mathbf{V}}(\widehat{\mathbf{\Pi}})^{-1} \, \widehat{\mathbf{\Pi}} \right)$$

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:

$$\alpha_i = \lambda_1^x \, x_{i1} + ... + \lambda_T^x \, x_{iT} + \lambda^z \, z_i + e_i$$

• 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 $\left\|\widehat{\beta}_{WG} - \widehat{\beta}_{FGLS}\right\|$ has very little power.

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

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

• 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 $\boldsymbol{\lambda} = \mathbf{0}$.

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:

$$log(crime_{it}) = \alpha_0 + \alpha_1 \ dum78_t + \beta \ avgpsolv_{it} + \alpha_i + u_{it}$$

$crime_{it}$ 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 β).

- **OLS in levels**

OLS in levels		
Dependent variable: log(crime)		
Standard errors robust of heterocedasticity		
Regressor	**Estimate**	**Std. error**
constant	4.182	0.193
dummy 78	-0.056	0.088
avgpsolv	-0.036	0.004
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 $\mathbb{E}(avgpsolv_{it}{}^{*}\ \alpha_i) = 0$.

• 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 $\mathbb{E}(avgpsolv_{it}{}^{*}\ \alpha_i) < 0$ and the OLS estimator of β will be downward biased.

• Therefore, the OLS may be a asymp. biased estimator of the causal effect of $avgpsolv$ on $crime$.

- **Within-groups estimator**

Within-Groups		
Dependent variable: log(crime)		
Standard errors robust of heterocedasticity		
Regressor	**Estimate**	**Std. error**
constant	3.316	0.231
dummy 78	0.099	0.062
avgpsolv	-0.017	0.005
Number obs.	106	
R-square (within)	0.408	
R-square (all)	0.436	
$\hat{\sigma}_\alpha$	0.471	
$\hat{\sigma}_u$	0.244	

- The deterrence effect is still significantly different to zero, but much smaller in magnitude. Though we should make a formal test, it seems that α_i is correlated with the regressors. There may be unobserved district characteristics (i.e., α_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 α_i) where crime is low and most crimes are easy to solve, and "bad districts" (i.e., high α_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 β, i.e., an estimate of the deterrence effect that is partly spurious.

FGLS estimator

FGLS Estimates		
Dependent variable: log(crime)		
Regressor	**Estimate**	**Std. error**
constant	3.799	0.186
dummy 78	0.013	0.013
avgpsolv	-0.027	0.004
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 β is between the OLS and the WG. Still, it seems that α_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 α_i.

• The Hausman statistic is equal to 14.85. Under the null hypothesis that $\mathbb{E}(\alpha_i X_i) = 0$ 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 β do you find more reasonable?

• There is clear evidence that the unobserved district effect α_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):

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

$n = \ln(\mathbb{E}mployment)$, $w = \ln(Wage)$, and $y = \ln(Output)$; γ_t represents aggregate shocks, and α_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 β_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 wit 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 β_1 has important economic implications. Short-run elasticity $=$ β_2 : Long-run elasticity $= \dfrac{\beta_2}{1-\beta_1}$.

• **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.** α_i represents firm heterogeneity in the steady-state optimal level of employment (due to productivity, other labor costs). α_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 β_1.

- Two possible explanations for the OLS estimate of β_1.

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

(B) **Persistent unobserved heterogeneity.** ε_{it} represents firm heterogeneity in the profitability of investment in R&D. If ε_{it} is persistent over time, then firms with higher values of ε_{it} tend to have also higher past levels of R&D: $cov(RD_{it-1}, \varepsilon_{it}) > 0$. OLS estimation will provide an upward biased estimate of β_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 β in this model.

• These expressions will help us to understand the problems and merits of different estimators.

• Solving backwards, we have that

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

• If $\beta \neq 0$:

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

(2) The regressor $y_{i,t-1}$ is correlated with $u_{i,t-1}$, $u_{i,t-2}, ...$, 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 $Cov(\Delta y_{i,t-1}, \Delta u_{it}) = 0$.

- However,

$$\begin{aligned}
& Cov(\triangle y_{i,t-1}, \triangle u_{it}) = \\
= \ & \underbrace{\mathbb{E}(y_{t-1}u_t)}_{0} - \underbrace{\mathbb{E}(y_{t-1}u_{t-1})}_{\sigma_u^2} - \underbrace{\mathbb{E}(y_{t-2}u_t)}_{0} + \underbrace{\mathbb{E}(y_{t-2}u_{t-1})}_{0} \\
= \ & 0 \qquad - \sigma_u^2 \qquad 0 \qquad 0 \\
= \ & -\sigma_u^2 \neq 0
\end{aligned}$$

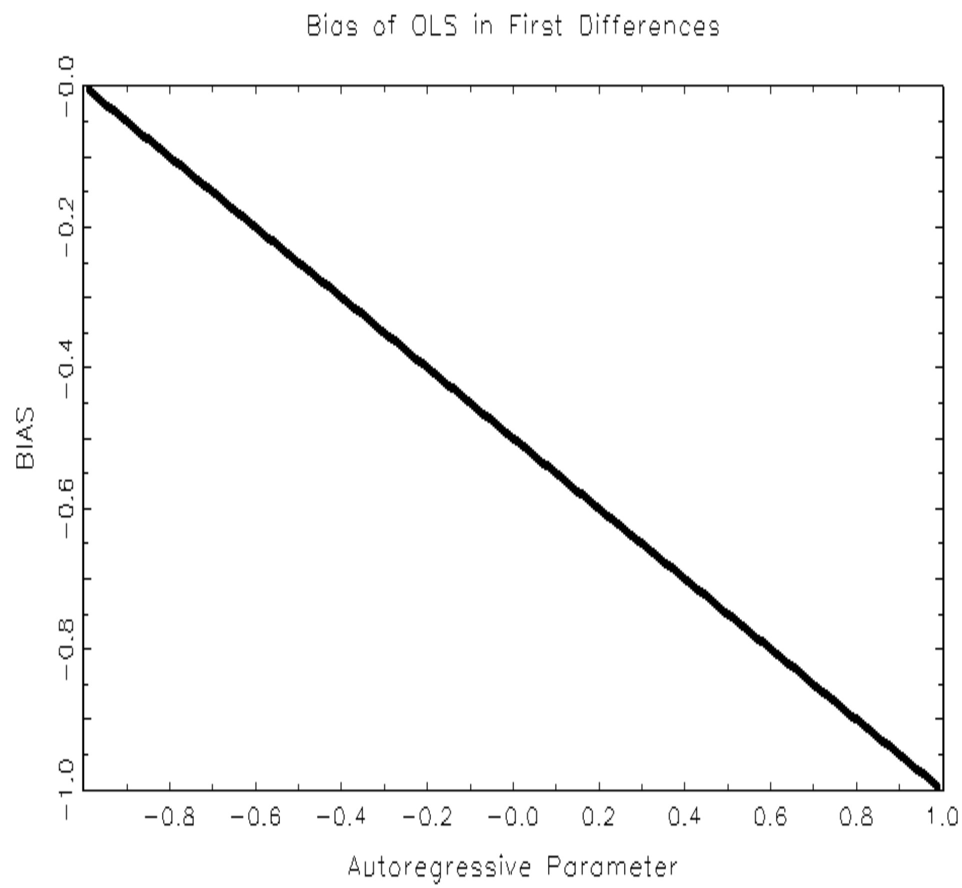
• **Exercise:** Show that $Var(\triangle y_{i,t-1}) = \dfrac{2\sigma_u^2}{1+\beta}$.

• Therefore

$$\begin{aligned} p\lim_{N\to\infty} \widehat{\beta}_{OLS-FD} &= \beta + \frac{Cov(\triangle y_{i,t-1}, \triangle u_{it})}{Var(\triangle y_{i,t-1})} \\ &= \frac{\beta - 1}{2} \end{aligned}$$

• 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 β 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 $T = 500$.



Bias of OLS in First Differences
BIAS
-0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.6
-0.4
-0.2
-0.0
0.2
0.4
0.6
0.8
1.0
Autoregressive Parameter

- **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!]

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

- 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 $\widehat{\beta}_{OLS-FD} \rightarrow_p \dfrac{\beta - 1}{2}$

• This provides a very simple approach to correct for the bias and obtain a consistent estimator of β.

• 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 ρ 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 $Cov((y_{i,t-1} - \bar{y}_i), (u_{it} - u_i)) = 0$. However, this covariance is

$$\begin{aligned} & Cov((y_{i,t-1} - \bar{y}_i), (u_{it} - u_i)) \\ = \;& \underbrace{\mathbb{E}(y_{i,t-1}, u_{it})}_{0} - \underbrace{\mathbb{E}(y_{i,t-1}, \bar{u}_i)}_{\neq 0} - \underbrace{\mathbb{E}(\bar{y}_i, u_{it})}_{\neq 0} + \underbrace{\mathbb{E}(\bar{y}_i, \bar{u}_i)}_{\neq 0} \\ < \;& 0 \end{aligned}$$

• Nickell (Econometrica, 1981) derived the expression for the asymptotic bias of the WG estimator in an AR(1) panel data model (for fixed T):

$$bias\left(\hat{\beta}_{WG}\right) = \frac{-\left(1-\beta^2\right)\ h(T,\beta)}{(T-1) - 2\beta\ h(T,\beta)}$$

where:

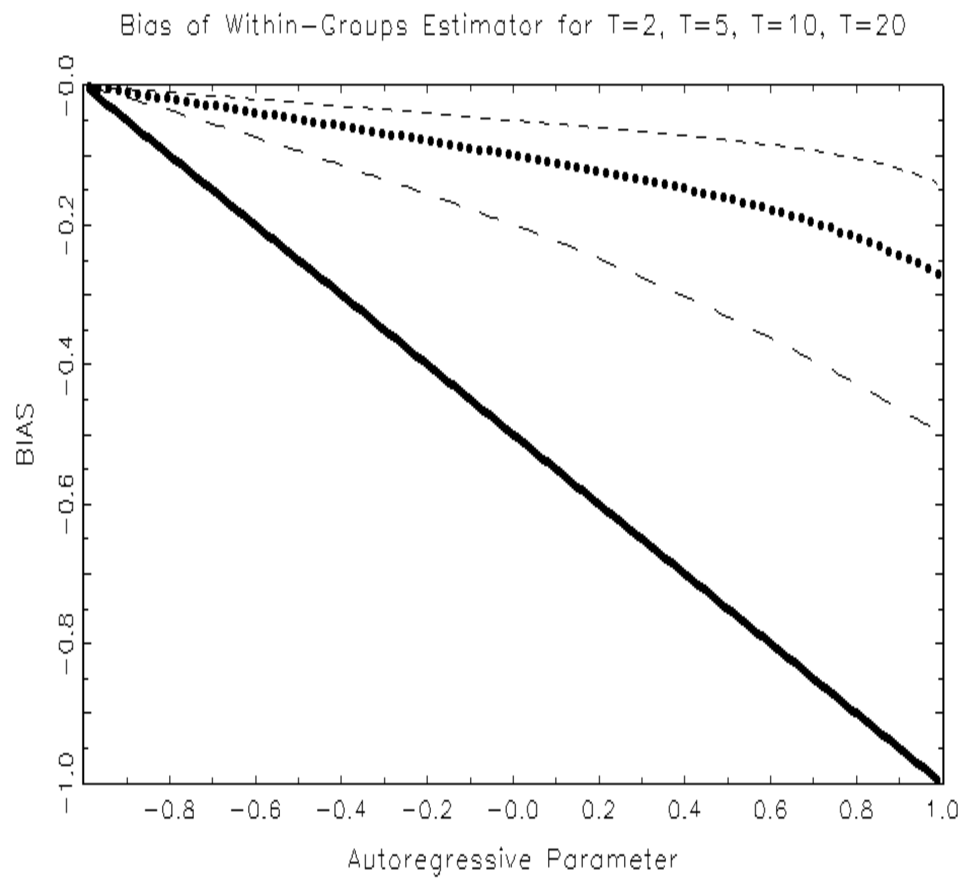
$$h(T,\beta) = \frac{1}{1-\beta}\left(1 - \frac{1-\beta^T}{T\,(1-\beta)}\right)$$

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

(2) It increases in absolute when β goes to one.

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

(4) If β is not small, the bias can be important even for T greater than 20.



Bias of Within-Groups Estimator for T=2, T=5, T=10, T=20
BIAS
-0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.6
-0.4
-0.2
-0.0
0.2
0.4
0.6
0.8
1.0
Autoregressive Parameter

• **Remark: WG is consistent as $\mathbf{T} \rightarrow \infty$**

• 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 β.

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

- **Remark: Bias reduction**

- Again, we have obtained an expression: $plim\hat{\beta}_{WG} = p(\beta, T)$, where the function $p(.)$ is known, and it strictly monotonic in β, 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) $\triangle u_{it}$ is NOT correlated with $y_{i,t-2}$, $y_{i,t-3}$, ...

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

• 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 $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} y_{i,t-2}\,\Delta y_{i,t-1} \;\to_p\; \mathbb{E}\left(y_{t-2}\,\Delta y_{t-1}\right)$$

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

By Mann-Wald Theorem, as $N \to \infty$,

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

$$\text{with } V_{AH} \;=\; \frac{\sum_{t=3}^{T} Var\left(y_{t-2}\,\Delta u_t\right)}{\left[\sum_{t=3}^{T}\mathbb{E}\left(y_{t-2}\,\Delta y_{t-1}\right)\right]^2}$$

• **Asymptotic Variance Anderson-Hsiao Estimator**

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

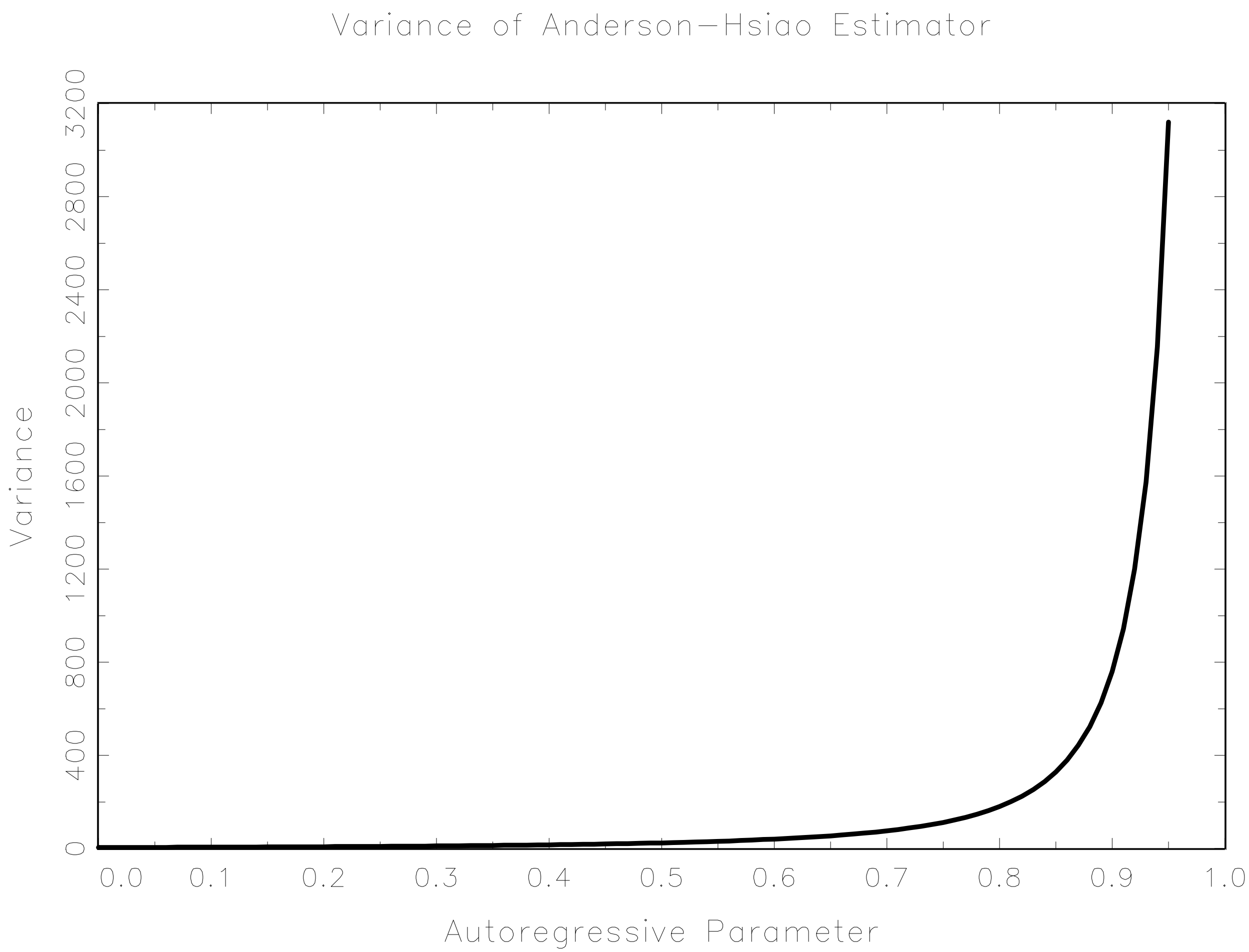
• Given these expressions, we have that the variance of the limiting distribution is:

$$
\begin{aligned}
V_{AH} &= \frac{(T-2)\ 2\sigma_u^2 \left(\dfrac{\sigma_\alpha^2}{(1-\beta)^2} + \dfrac{\sigma_u^2}{1-\beta^2}\right)}{\left[(T-2)\ \sigma_u\ \ (1+\beta)^{-1}\right]^2} \\
&= \frac{2}{(T-2)}\left(\frac{1+\beta}{1-\beta}\right)\left(1+\left(\frac{1+\beta}{1-\beta}\right)\frac{\sigma_\alpha^2}{\sigma_u^2}\right)
\end{aligned}
$$

• **Asymptotic Variance AH Estimator [Cont]**

• **Comment 1:** The variance increases more than exponentially with β. It can be pretty large if β is close to 1.

• For $\beta = 1$, the variance is infinite, i.e., the instrument $y_{i,t-2}$ is not correlated with $\Delta y_{i,t-1}$.



Variance of Anderson–Hsiao Estimator
Variance
0
400
800
1200
1600
2000
2400
2800
3200
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Autoregressive Parameter

• **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:

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

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

$$95\% \ CI \text{ for } \beta = \ \ [-0.511 \ \ , \ \ +1.711]$$

That basically does not provide any information about the true value of β.

• **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}, ...) 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 Δy when β 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-differnces:

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

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 Δy_{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:

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

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.

$$\begin{cases} \text{Equation for } t=3: & \Delta y_{i3} = \beta\ \Delta y_{i2} + \Delta u_{i3} \\ \text{Equation for } t=4: & \Delta y_{i4} = \beta\ \Delta y_{i3} + \Delta u_{i4} \\ \vdots & \vdots \\ \text{Equation for } t=T: & \Delta y_{iT} = \beta\ \Delta y_{iT-1} + \Delta u_{iT} \end{cases}$$

- If $\{u_{it}\}$ is iid, we have the following moment conditions:

$$\text{For } t = 3: \quad \mathbb{E}[y_{i1} \triangle u_{i3}] = 0$$

$$\text{For } t = 4: \quad \begin{array}{l} \mathbb{E}[y_{i1} \triangle u_{i4}] = 0 \\ \mathbb{E}[y_{i2} \triangle u_{i4}] = 0 \end{array}$$

$$\text{For } t = 5: \quad \begin{array}{l} \mathbb{E}[y_{i1} \triangle u_{i5}] = 0 \\ \mathbb{E}[y_{i2} \triangle u_{i5}] = 0 \\ \mathbb{E}[y_{i3} \triangle u_{i5}] = 0 \end{array}$$

- The total number of moment conditions is:

$$q \equiv 1 + 2 + ... + (T - 2) = \frac{(T - 2)(T - 1)}{2}$$

• We can represent these q moment conditions in vector form as:

$$\mathbb{E}\begin{bmatrix} y_{i1} \triangle u_{i3} \\ y_{i1} \triangle u_{i4} \\ y_{i2} \triangle u_{i4} \\ \ldots \end{bmatrix} = \mathbb{E}\left[\mathbf{Z}_i\left(\triangle\mathbf{Y}_i - \beta \ \triangle \mathbf{Y}_{i(-1)}\right)\right]$$

where $\triangle\mathbf{Y}_i$ and $\triangle\mathbf{Y}_{i(-1)}$ are $(T-2)\times 1$ vectors

$$\triangle\mathbf{Y}_i = (\triangle y_{i3}, \ \triangle y_{i4}, \ ..., \ \triangle y_{iT})'$$

and $\mathbf{Z}_i$ is the $q \times (T-2)$ matrix of instruments.

- And $\mathbf{Z}_i$ is the $q \times (T-2)$ matrix of instruments:

$$\mathbf{Z}_i = \begin{bmatrix} y_{i1} & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & y_{i1} \\ \\ 0 & y_{i2} & 0 & \dots & 0 \\ 0 & 0 & y_{i2} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & y_{i2} \\ \\ \vdots & \vdots & \vdots & & \vdots \\ \\ 0 & 0 & 0 & & y_{iT-2} \end{bmatrix}$$

• To understand the structure of the $q \times (T-2)$ matrix of instruments $\mathbf{Z}_i$, note that:

- Each row of $\mathbf{Z}_i$ 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 $\left[\begin{array}{ccccc} y_{i1} & 0 & 0 & \ldots & 0 \end{array}\right]$ represents y_{i1} as instrument in the first equation, $t=3$.

- Row $\left[\begin{array}{ccccc} 0 & 0 & y_{i1} & \ldots & 0 \end{array}\right]$ represents y_{i1} as instrument in the third equation, $t=5$.

- Row $\left[\begin{array}{ccccc} 0 & y_{i2} & 0 & \ldots & 0 \end{array}\right]$ 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 $\mathbb{E}\left[\mathbf{Z}_i\left(\triangle\mathbf{Y}_i - \beta \ \triangle \mathbf{Y}_{i(-1)}\right)\right]$

$$m_N(\beta) \ = \ N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\left(\triangle\mathbf{Y}_i - \beta \ \triangle \mathbf{Y}_{i(-1)}\right)$$

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

• The **Arellano-Bond GMM estimator** is:

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

where $\hat{\Omega}$ is a consistent estimator of the optimal weighting matrix, $\Omega = \mathbb{E}\left(\mathbf{Z}_i \ \Delta\mathbf{U}_i \ \Delta\mathbf{U}_i' \ \mathbf{Z}_i\right)$, with $\Delta\mathbf{U}_i = (\Delta u_{i3}, \Delta u_{i4}, ..., \Delta u_{iT})'$.

- Deriving the first order conditions and solving for $\hat{\beta}_{AB}$ we get:

$$\hat{\beta}_{AB} = \frac{\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{\Delta Y}_{i(-1)}\right)' \hat{\Omega}^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{\Delta Y}_{i}\right)}{\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{\Delta Y}_{i(-1)}\right)' \hat{\Omega}^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{\Delta Y}_{i(-1)}\right)}$$

- 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 $\Omega = \sigma_u^2 \ \mathbb{E}\left(\mathbf{Z}_i \ \mathbf{H} \ \mathbf{Z}_i'\right)$ 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.

$$\mathbf{H} = \begin{bmatrix} 2 & -1 & ... & 0 & 0 \\ -1 & 2 & ... & 0 & 0 \\ ... & ... & & ... & ... \\ 0 & 0 & ... & 2 & -1 \\ 0 & 0 & ... & -1 & 2 \end{bmatrix}$$

- Then, the **first stage GMM estimator is:**

$$\hat{\beta}_{AB}^{(1)} = \frac{\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{Y}_{i(-1)}\right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{H} \, \mathbf{Z}_i'\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{Y}_i\right)}{\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{Y}_{i(-1)}\right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{H} \, \mathbf{Z}_i'\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{Y}_{i(-1)}\right)}$$

• In the **second stage GMM estimator,** given $\hat{\beta}_{AB}^{(1)}$ we can get the residuals $\Delta \hat{u}_{it} = \Delta y_{it} - \hat{\beta}_{AB}^{(1)} \Delta y_{i,t-1}$. And we can use these residuals to obtain a consistent estimator of Ω that is robust to heterocedasticity.

$$\hat{\Omega} = N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\widehat{\mathbf{U}}_i \; \Delta\widehat{\mathbf{U}}_i' \; \mathbf{Z}_i'$$

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

• Finally, the two-stage GMM estimator is:

$$\hat{\beta}_{AB}^{(2)} = \frac{\left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\mathbf{Y}_{i(-1)}\right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\widehat{\mathbf{U}}_i \; \Delta\widehat{\mathbf{U}}_i' \; \mathbf{Z}_i'\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\mathbf{Y}_i\right)}{\left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\mathbf{Y}_{i(-1)}\right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\widehat{\mathbf{U}}_i \; \Delta\widehat{\mathbf{U}}_i' \; \mathbf{Z}_i'\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \; \Delta\mathbf{Y}_{i(-1)}\right)}$$

• 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:

$$\mathbf{Y}_i = \mathbf{X}_i \, \beta + \alpha_i + \mathbf{U}_i$$

• Let $\mathbf{Z}_i$ be the matrix of instruments construncted in the way described above, such that:

$$\mathbb{E}\left[\mathbf{Z}_i \left(\triangle \mathbf{Y}_i - \mathbf{X}_i \, \beta\right)\right] = 0$$

- The **two-stage GMM estimator of** β is:

$$\hat{\beta}_{AB}^{(2)} = \left[\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{X}_i \right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\widehat{\mathbf{U}}_i \, \Delta\widehat{\mathbf{U}}_i' \, \mathbf{Z}_i' \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{Y}_i \right) \right]^{-1}$$

$$\left[\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{X}_i \right)' \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\widehat{\mathbf{U}}_i \, \Delta\widehat{\mathbf{U}}_i' \, \mathbf{Z}_i' \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \, \Delta\mathbf{X}_i \right) \right]$$

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.

$$\mathbb{E}\left[\Delta w_{it}\ \Delta u_{it}\right] \neq 0,\ \mathbb{E}\left[\Delta k_{it}\ \Delta u_{it}\right] \neq 0,\ \mathbb{E}\left[\Delta q_{it}\ \Delta u_{it}\right] \neq 0$$

• 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 Δu_{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)

Regressor	**OLS-Levels**	**WG**	**AH**	**AB-1step**	**AB-2step**
$ln(emp)$ $[t-1]$	1.045	0.734	2.308	0.686	0.629
	(0.051)	(0.058)	(1.055)	(0.145)	(0.090)
$ln(wage)$ $[t]$	-0.524	-0.557	-0.810	-0.608	-0.526
	(0.172)	(0.155)	(0.283)	(0.178)	(0.054)
$ln(wage)$ $[t-1]$	0.477	0.326	1.422	0.393	0.311
	(0.169)	(0.143)	(0.851)	(0.168)	(0.094)
Sargan (d.o.f.)				65.8 (25)	31.4 (25)
m2				-0.516	-0.434

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 Δu_{it}

(a) Test of over-identifying restrictions (Hansen-Sargan):

$$H_0 : \mathbb{E}\left(\mathbf{Z}_i \; \mathbf{\Delta U}_i\right) = 0 \qquad\qquad (q \text{ restrictions})$$

• The model is based on the q restrictions $\mathbb{E}\left(\mathbf{Z}_i \; \mathbf{\Delta U}_i\right) = 0$. In principle, we would like to impose all these restrictions in the estimation of β.

$$m_N(\widehat{\beta}) = \sum_{i=1}^{N} \mathbf{Z}_i \; \left[\mathbf{\Delta Y}_i - \mathbf{\Delta X}_i \; \widehat{\beta}\right] = 0$$

• However, if $q > K$, there not exists a value of $\widehat{\beta}$ that solves this system. Therefore, we define $\widehat{\beta}$ 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, $\widehat{\beta}$ satisfies the conditions:

$$\left(\sum_{i=1}^{N} \mathbf{Z}_i \, \mathbf{\Delta X}_i\right)' \hat{\Omega}^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i \left[\mathbf{\Delta Y}_i - \mathbf{\Delta X}_i \, \widehat{\beta}\right]\right) = 0$$

- However, this is compatible with having that:

$$\left\| m_N(\widehat{\beta}) \right\| = \left\| \sum_{i=1}^{N} \mathbf{Z}_i \left[\mathbf{\Delta Y}_i - \mathbf{\Delta X}_i \, \widehat{\beta}\right] \right\| >>>>> 0$$

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

Hansen-Sargan Test (cont)

- Under $H_0 : \mathbb{E}\left(\mathbf{Z}_i\ \boldsymbol{\Delta}\mathbf{U}_i\right) = 0$, we have that $\left\| m_N(\widehat{\beta}) \right\| \rightarrow_p 0$ and

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

If $\boldsymbol{\xi}_i \equiv \mathbf{Z}_i\ \boldsymbol{\Delta}\widehat{\mathbf{U}}_i$, note that:

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

- Under the alternative hypothesis $H_1 : \mathbb{E}\left(\mathbf{Z}_i\ \boldsymbol{\Delta}\mathbf{U}_i\right) \neq 0$, we have that $\left\| m_N(\widehat{\beta}) \right\| \rightarrow_p c > 0$, and

$$HS \equiv\ m_N(\hat{\beta})'\ \hat{\Omega}^{-1}\ m_N(\hat{\beta}) \sim_d \text{non central } \chi^2_{q-K} \text{ with noncentrality } c$$

Example: Employment equation. Arellano & Bond (REStud, 91)

Statistic	AB-1step	AB-2step
Sargan (d.o.f.)	65.8 (25)	31.4 (25)

- Why 25 degrees of freedom?

40 momemt restrictions:

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

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

$\{n_{i3}, w_{i3}, k_{i3}, q_{i3}\}$ in eq. at $t = 5, 6$ [8]

and $\mathbb{E}(\Delta u_{it}) = 0$ $t = 3, 4, 5, 6$ [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 $\mathbf{u}_{it}$

• 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 α_i.

• However, we can get residuals for $\triangle u_{it}$. Our test of serial correlation will test the serial correlation of $\triangle u_{it}$.

Testing for serial correlation in $\mathbf{u}_{it}$ (cont)

• Note that if u_{it} is not serially correlated:

$$\begin{aligned}
\mathbb{E}\left(\triangle u_{it}\ \triangle u_{it-1}\right) &= -\sigma_u^2(t) < 0 \\
\mathbb{E}\left(\triangle u_{it}\ \triangle u_{it-j}\right) &= 0 \qquad \text{for any lag } j \geq 2
\end{aligned}$$

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

• Testing for serial correlation of order 1 in $\triangle u_{it}$ is also useful. If we find that $\mathbb{E}\left(\triangle u_{it}\ \triangle u_{it-1}\right) = 0$, then OLS-FD is consistent.

Arellano-Bond test 2nd order correlation in $\triangle \mathbf{u}_{it}$

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

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

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

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

• Then, under H_0: $r_2 = 0$.

$$\widehat{m}_2 \equiv \frac{\widehat{r}_2}{se(\widehat{r}_2)} \underset{a}{\sim} N(0,1)$$

Arellano-Bond test j-th order correlation in $\triangle \mathbf{u}_{it}$

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

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

- 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 $\widehat{r}_j$ be its sample counterpart.

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

$$\widehat{m}_j \equiv \frac{\widehat{r}_j}{se(\widehat{r}_j)} \underset{a}{\sim} N(0, 1)$$

What if we reject null hypothesis r2 = 0?

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

• 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 r2 = 0? [Cont]

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

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

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

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

- Separate identification of ρ and β 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:

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

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-Bonnd 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 $\triangle y_{it-1}$ with lagged values of $\{y_{it-j} : j \geq 2\}$.

- Therefore, if $corr(\triangle y_{it-1}, y_{i,t-j})$ 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 α_i and x_i: Chamberlain's Correlated Random Effects.

(3) LIML estimation: Alonso-Borrego and Arellano (JBES, 1999).

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,

$$y_{it^*} = \mu_{y,i} \equiv \mathbb{E}\left(y_{it} \mid i\right)$$

• 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^*:

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

- 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

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

- We have that:

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

- Δy_{it} **does NOT dependent of** α_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 $\triangle y_{i,t-1}$, $\triangle y_{i,t-2}$, ... 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 $\beta \rightarrow 1$.**

- **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 $\beta \rightarrow 1$.**

• 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 α is such that (**Exercise**):

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

That goes to zero as β goes to one. Weak instruments problem!

• **This IV estimator (IV version of Blundell and Bond) does not suffer of a weak instruments problem when $\beta \rightarrow 1$. [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 π is such that (**Exercise**):

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

which does not go to zero as β 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:

$$\begin{aligned}\sqrt{N}\left(\hat{\beta}-\beta\right) &= \frac{N^{-1/2}\sum_{i=1}^{N}\Delta y_{i,t-1}\ \left(\alpha_i+u_{it}\right)}{N^{-1}\sum_{i=1}^{N}\Delta y_{i,t-1}\ y_{i,t-1}} \\ &\to_d \frac{N\left(0, Var\left(\Delta y_{i,t-1}\left(\alpha_i+u_{it}\right)\right)\right)}{\mathbb{E}\left(\Delta y_{i,t-1}\ y_{i,t-1}\right)}\end{aligned}$$

• Therefore,

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

• Given that:

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

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

we have that:

$$\begin{aligned} Var\left(\hat{\beta}\right) &= \frac{1}{N\left(T-2\right)}\frac{2\sigma_u^2\left(\sigma_u^2+\sigma_\alpha^2\right)/\left(1+\beta\right)}{\sigma_u^4/\left(1+\beta\right)^2} \\ &= \frac{1}{N\left(T-2\right)}2\left(1+\beta\right)\left(1+\frac{\sigma_\alpha^2}{\sigma_u^2}\right) \end{aligned}$$

• In contrast with the AH and AB estimators, the variance of this estimator does not goes to infinity as β goes to one.

Example: Comparing of AH-IV and BB-IV

• Take the sample numerical example that we considered above for the AH estimator.

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

• For AH-IV estimator, we have:

$$sd\left(\hat{\beta}_{AH}\right) = 0.567 \qquad \text{and} \qquad 95\% \ CI \text{ for } \beta = \ [-0.51 \ \ , \ \ +1.71]$$

• For BB-IV estimator, we have:

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

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:

$$\text{For:} \quad \begin{array}{l} t = 3, 4, ..., T \\ s = 2, 3..., t-1 \end{array} \qquad \mathbb{E}\left(\triangle y_{is}\ \left[y_{it} - \beta\ y_{it-1}\right]\right) = 0$$

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

- The AB moment conditions are:

$$\text{For:} \quad \begin{array}{l} t = 3, 4, ..., T \\ s = 1, 2..., t-2 \end{array} \qquad \mathbb{E}\left(y_{is}\ \left[\triangle y_{it} - \beta\ \triangle y_{it-1}\right]\right) = 0$$

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:

`xtabond2` *depvar x1 x2 ... xK,*

$gmm(z1 \ ... \ zq \ lag(lmin \ \ lmax))$

$iv(v1 \ ... \ vq)$, `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:

$$y_{it} = \beta_n \, n_{it} + \beta_k \, k_{it} + \alpha_i + \gamma_t + u_{it}$$

$y \equiv \ln(Output)$; $n \equiv \ln(\mathbb{E}mployment)$; $k \equiv \ln(Capital)$. The unobservables α_i and u_{it} represent firms' productivity differences which are permanent and transitory, respectively.

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

$$\begin{array}{rcl} n_{it} & = & f_l\left(n_{i,t-1}, k_{i,t-1}, \alpha_i, u_{it}, w_{it}\right) \\ k_{it} & = & f_k\left(n_{i,t-1}, k_{i,t-1}, \alpha_i, u_{it}, w_{it}\right) \end{array}$$

• 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:

$$
\begin{aligned}
y_{it} &= \rho\, y_{it-1} + \beta_n\, n_{it} + (-\rho\beta_n) n_{it-1} + \beta_k\, k_{it} + (-\rho\beta_k) k_{it-1} \\
&\quad + (1-\rho)\alpha_i + (\gamma_t - \rho\gamma_{t-1}) + a_{it}
\end{aligned}
$$

Δy_{it-1}, Δy_{it-2}, Δn_{it-1}, Δn_{it-2}, Δk_{it-1}, and Δk_{it-2} are valid intruments in this equation [BB].

• Model in quasi-first + first differences:

$$
\begin{aligned}
\Delta y_{it} &= \rho\, \Delta y_{it-1} + \beta_n\, \Delta n_{it} + (-\rho\beta_n) \Delta n_{it-1} + \beta_k\, \Delta k_{it} + (-\rho\beta_k) \Delta k_{it-1} \\
&\quad + (\Delta\gamma_t - \rho\Delta\gamma_{t-1}) + \Delta a_{it}
\end{aligned}
$$

y_{it-2}, y_{it-3}, n_{it-2}, n_{it-3}, k_{it-2}, and k_{it-3} are valid intruments in this equation [BB].

Production Function: Blundell & Bond (ER 2000). Tables 3

509 manufacturing firms; 1982-89				
Parameter	**OLS-Levels**	**WG**	**AB-GMM**	**SYS-GMM**
β_n	0.538	0.488	0.515	0.479
	(0.025)	(0.030)	(0.099)	(0.098)
β_k	0.266	0.199	0.225	0.492
	(0.032)	(0.033)	(0.126)	(0.074)
ρ	0.964	0.512	0.448	0.565
	(0.006)	(0.022)	(0.073)	(0.078)
Sargan (p-value)	-	-	0.073	0.032
m2	-	-	-0.69	-0.35
Constant RS (p-v)	0.000	0.000	0.006	0.641

THANKS TO ALL THE STUDENTS WHO HAVE GREATELY CONTRIBUTED TO IMPROVE THESE NOTES