ApEc 8213: Econometric Analysis III -- Lecture #12

Instrumental Variables and Dynamic Models for Panel Data
Hansen, Chapter 17, Sections 17.28 – 17.42

Panel data models were introduced in Apec 8212. Today's lecture covers two topics that were not covered in Apec 8212: instrumental variable estimation for panel data and dynamic panel models (models with lagged Y variables). The focus is on fixed effects panel data models.

I. Instrumental Variables for Panel Data (17.28 – 17.31)

Recall the fixed effects model for panel data:

$$Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it} \qquad (17.68)$$

where $i = 1, 2, \ldots N$ (and n = total number of observations)

For the fixed effects model, **we can difference out $\boldsymbol{u_i}$**, and so the only **concern** is that some of the $\boldsymbol{X_{it}}$ **variables** could be **correlated with $\boldsymbol{\varepsilon_{it}}$**. If such correlation occurs, standard fixed effects estimation will be **biased and inconsistent**.

Let Z_{it} **be** a **vector of $\boldsymbol{\ell}$ instrumental variables**, with $\ell \geq k$ (k is the number of X variables). Z_{it} includes both X_{it} variables that are not correlated with ε_{it} and "excluded" instruments that are not in equation (17.68). Let $\boldsymbol{Z_i}$ be the $T_i \times \ell$ matrix of instruments for individual i and let $\boldsymbol{Z}$ be the $n\times\ell$ matrix that "stacks" all the $\boldsymbol{Z}_i$ matrices (n is the total number of observations). Stacking Y_{it}, X_{it} and ε_{it} in the same way gives:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{\varepsilon} \quad (17.69)$$

where u is the $N\times 1$ column vector of fixed effects, and:

$$\boldsymbol{D} = \text{diag}\{\mathbf{I}_{T_1}, \ldots \mathbf{I}_{T_N}\} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 0 \\ 0 & 0 & & 1 \\ 0 & 0 & & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

OLS estimation of (17.69) gives the fixed effects estimate of β. Note that the $\boldsymbol{D}$ dummy variables are exogenous.

Turn now to **2SLS (IV) estimation of $\boldsymbol{\beta}$**. The first stage is to regress all X variables on the instruments Z. Since Z

includes all exogenous X variables, it includes the $\boldsymbol{D}$ variables. Thus 2SLS/IV estimation of (17.68) can be done by applying in 2SLS/IV to equation (17.69), where $\boldsymbol{Z}$ and $\boldsymbol{D}$ are the instruments for $\boldsymbol{X}$ and $\boldsymbol{D}$.

Estimating (17.69) is not very practical, so **within estimation should be used**. This is done for all variables by subtracting the individual specific means. For all variables in equation (17.69) this is done by using $\boldsymbol{M_D} = \boldsymbol{I}_n - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$: $\dot{\boldsymbol{Y}} = \boldsymbol{M_D Y}$, $\dot{\boldsymbol{X}} = \boldsymbol{M_D X}$, and $\dot{\boldsymbol{Z}} = \boldsymbol{M_D Z}$.

Question: What is $\dot{\boldsymbol{D}} = \boldsymbol{M_D D}$?

We can then write the 2SLS/IV estimate of β as:

$$\hat{\beta}_{2sls} = (\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{X}})^{-1}(\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Y}})$$

You can easily add time fixed effects for "two-way" fixed effects by adding dummy variables for each time period.

Identification with Instrumental Variables

There are **two requirements that IV panel data models must satisfy** to be able to identify (estimate) β. They are most easily expressed in terms of the above transformed variables. They are:

$$\mathrm{E}[\dot{\boldsymbol{Z}}_i'\dot{\boldsymbol{Z}}_i] > 0 \text{ (is a positive definite matrix)} \quad (17.70)$$

$$\text{rank}(\text{E}[\dot{\boldsymbol{Z}}_i'\dot{\boldsymbol{X}}_i]) = k \qquad (17.71)$$

Condition (17.70) is the requirement that, after "within differencing", each instrument Z has variation that is not a linear combination of the variation in the other instruments. Condition (17.71) is the requirement that the (demeaned) instruments have predictive power for all the $\dot{X}$ variables.

Asymptotic Distribution of Fixed Effects 2SLS Estimator

We are now **ready to show that $\widehat{\boldsymbol{\beta}}_{\mathbf{2sls}}$**, the fixed effects 2SLS estimate of β in equation (17.68) **is consistent**, and to work out its asymptotic distribution. Hansen does this for the **case of a "balanced" panel**, and then discusses the extension to the "unbalanced" case. Here are the assumptions (for the balanced case), followed by the theorem.

Assumption 17.4

1. $Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$ for $i = 1, 2, \dots N$ and $t = 1, 2, \dots T$.
2. The variables ($\boldsymbol{\varepsilon}_i$, $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$) are i.i.d for $i = 1, 2, \dots N$.
3. $\text{E}[Z_{is}\varepsilon_{it}] = 0$ for all $s = 1, 2, \dots T$. (strict exogeneity)
4. $\boldsymbol{Q}_{ZZ} = \text{E}[\dot{\boldsymbol{Z}}_i'\dot{\boldsymbol{Z}}_i] > 0$.

5. rank($\boldsymbol{Q}_{ZX}$) = k, where $\boldsymbol{Q}_{ZX} = \mathrm{E}[\dot{\boldsymbol{Z}}_i'\dot{\boldsymbol{X}}_i]$.

6. $\mathrm{E}[\varepsilon_{it}^4] < \infty$.

7. $\mathrm{E}[\|X_{it}\|^2] = \mathrm{E}[\sum_{j=1}^{k} X_{ijt}^2] < \infty$.

8. $\mathrm{E}[\|Z_{it}\|^4] = \mathrm{E}[(\sum_{j=1}^{k} Z_{ijt}^2)^2] < \infty$.

Theorem 17.4. Under Assumption 17.4, as $N \to \infty$,

$$\sqrt{N}(\hat{\beta}_{2\text{sls}} - \beta) \underset{d}{\to} \mathrm{N}(0, \boldsymbol{V}_\beta)$$

where

$$\boldsymbol{V}_\beta = (\boldsymbol{Q}_{ZX}'\boldsymbol{\Omega}_{ZZ}^{-1}\boldsymbol{Q}_{ZX})^{-1}(\boldsymbol{Q}_{ZX}'\boldsymbol{\Omega}_{ZZ}^{-1}\boldsymbol{\Omega}_{Z\varepsilon}\boldsymbol{\Omega}_{ZZ}^{-1}\boldsymbol{Q}_{ZX})(\boldsymbol{Q}_{ZX}'\boldsymbol{\Omega}_{ZZ}^{-1}\boldsymbol{Q}_{ZX})^{-1}$$

$$\boldsymbol{\Omega}_{ZZ} = \mathrm{E}[\dot{\boldsymbol{Z}}_i'\dot{\boldsymbol{Z}}_i], \quad \boldsymbol{\Omega}_{Z\varepsilon} = \mathrm{E}[\dot{\boldsymbol{Z}}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\dot{\boldsymbol{Z}}_i]$$

$\boldsymbol{V}_\beta$, which **allows for general heteroscedasticity within observations** for unit/person i (i.e. for a given unit/person ε can be correlated over time), can be estimated as:

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}} = (\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{X}})^{-1}(\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}})(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}(\textstyle\sum_{i=1}^{N}\dot{\boldsymbol{Z}}_i'\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\dot{\boldsymbol{Z}}_i)$$
$$\times(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{X}})(\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}(\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Z}})^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{X}})^{-1}$$

where $\hat{\boldsymbol{\varepsilon}}_i$ is the 2SLS/IV fixed effects residuals.

Hansen also presents a version of $\boldsymbol{V}_\beta$ under the (unlikely) assumption that ε_{it} is homoscedastic and uncorrelated over time (bottom of p.647), but this assumption is doubtful, so I do not recommend it.

The proof for the unbalanced panel adds the assumption that missing data are not correlated with ε_{it}. See p.647.

Linear GMM

An alternative to 2SLS/IV is GMM estimation which, in theory, is more efficient. In the "just-identified" case (ℓ = k), 2SLS selects β to "solve" the equation $\dot{\boldsymbol{Z}}'(\dot{\boldsymbol{Y}} - \dot{\boldsymbol{X}}\beta) = 0$. The underlying moment condition is $\mathrm{E}[\dot{\boldsymbol{Z}}_i'(\dot{\boldsymbol{Y}}_i - \dot{\boldsymbol{X}}_i\beta)] = 0$. The moment condition should also hold for the over-identified case. GMM can be applied to this case. Let $\widehat{\boldsymbol{W}}$ be the estimator of $\boldsymbol{W} = \mathrm{E}[\dot{\boldsymbol{Z}}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\dot{\boldsymbol{Z}}_i]$. That is:

$$\widehat{\boldsymbol{W}} = (1/N)\sum_{i=1}^{N} \dot{\boldsymbol{Z}}_i'\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\dot{\boldsymbol{Z}}_i$$

The GMM fixed effects estimator is then:

$$\hat{\beta}_{\mathrm{gmm}} = (\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}\widehat{\boldsymbol{W}}^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{X}})^{-1}(\dot{\boldsymbol{X}}'\dot{\boldsymbol{Z}}\widehat{\boldsymbol{W}}^{-1}\dot{\boldsymbol{Z}}'\dot{\boldsymbol{Y}})$$

Hansen explains how to estimate this using Stata 16 on p.648 (but Stata 18 or 19 may have an option for estimating $\hat{\beta}_{\mathrm{gmm}}$).

II. Panel Data with Time-Invariant Regressors (17.32-17.33)

Standard fixed effects estimation removes any individual **variables that do not change over time**, so **their coefficients cannot be estimated**. **Random effects** estimation **does allow for such variables**, **but** at the "cost" of **assuming** that the **individual effects are not correlated with the *X* variables**.

In fact, **there are ways to keep variables that do not vary over time by using instrumental variable methods** (and making more assumptions). Hansen gives an example in Section 17.32, but this is just a special case of the Hausman-Taylor method, so let's go to that more general method.

Hausman-Taylor Model

Generalize equation (17.68) to **allow for variables that vary across people but not over time**, denoted by Z_i, and to **distinguish between variables** that are **correlated with $\boldsymbol{u_i}$ and** variables that are **not correlated with $\boldsymbol{u_i}$**:

$$Y_{it} = X_{1it}'\beta_1 + X_{2it}'\beta_2 + Z_{1i}'\gamma_1 + Z_{2i}'\gamma_2 + u_i + \varepsilon_{it}$$

There are k_1 X_{1it} variables, k_2 X_{2it} variables, ℓ_1 Z_{1i} variables, and ℓ_2 Z_{2i} variables. Write this model in matrix form:

$$\boldsymbol{Y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{Z}_1\gamma_1 + \boldsymbol{Z}_2\gamma_2 + \boldsymbol{u} + \boldsymbol{\varepsilon} \quad (17.80)$$

Denote matrices of the same dimensions as $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ but with individual-specific means as $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$, so that the "within" transformations of those matrices are $\dot{\boldsymbol{X}}_1 = \boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1$ and $\dot{\boldsymbol{X}}_2 = \boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2$.

In this model, **all variables are assumed to be uncorrelated with ε_{it}**, and the following **additional assumptions** are made:

$$\mathrm{E}[X_{1it}u_i] = 0$$

$$\mathrm{E}[Z_{1i}u_i] = 0$$

However, **X_{2it} and Z_{2i} may still be correlated with u_i**.

Let $\boldsymbol{X}$ denote $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Z}_1, \boldsymbol{Z}_2)$ and β denote $(\beta_1, \beta_2, \gamma_1, \gamma_2)$, which implies that $\boldsymbol{Y} - \boldsymbol{X}\beta = \boldsymbol{u} + \boldsymbol{\varepsilon}$. The assumptions imply the following (population) moment conditions:

$$\mathrm{E}[\dot{\boldsymbol{X}}_1'(\boldsymbol{Y} - \boldsymbol{X}\beta)] = 0$$
$$\mathrm{E}[\dot{\boldsymbol{X}}_2'(\boldsymbol{Y} - \boldsymbol{X}\beta)] = 0$$
$$\mathrm{E}[\overline{\boldsymbol{X}}_1'(\boldsymbol{Y} - \boldsymbol{X}\beta)] = 0$$
$$\mathrm{E}[\boldsymbol{Z}_1'(\boldsymbol{Y} - \boldsymbol{X}\beta)] = 0$$

Question: Since we allow X_{2it} to be correlated with u_i, how is it that $\mathrm{E}[\dot{\boldsymbol{X}}_2'(\boldsymbol{Y} - \boldsymbol{X}\beta)] = 0$?

Thus we have $2k_1 + k_2 + \ell_1$ moment conditions and $k_1 + k_2 + \ell_1 + \ell_2$ coefficients to estimate. **At minimum, identification** of all parameters **requires $\boldsymbol{k_1 \geq \ell_2}$**: there are at least as many exogenous time-varying varying regressors (X_1 variables) as there are endogenous time-invariant regressors (Z_2 variables).

There is also a rank condition that is sufficient for identification. See Wooldridge (2010, p.362); this condition is analogous to Assumptions 17.4.4 and 17.4.5 in Hansen.

Equation (17.80) can be estimated by 2SLS/IV using the instruments $\boldsymbol{Z} = (\dot{\boldsymbol{X}}_1, \dot{\boldsymbol{X}}_2, \overline{\boldsymbol{X}}_1, \boldsymbol{Z}_1)$. The exogenous variables are $\boldsymbol{X}_1$ and $\boldsymbol{Z}_1$, and the endogenous variables are $\boldsymbol{X}_2$ and $\boldsymbol{Z}_2$.

The intuition is the following. We can always use fixed effects estimation to estimate β_1 and β_2. We can then estimate the following regression:

$$Y_{it} - X_{1it}'\hat{\beta}_1 - X_{2it}'\hat{\beta}_2 = Z_{1i}'\gamma_1 + Z_{2i}'\gamma_2 + u_i + \varepsilon_{it}$$

if we have enough X_1 variables to serve as instruments for the Z_2 variables.

Question: Why not use $\dot{X}_2$ variables and instruments for Z_2?

Always use a clustered and robust covariance matrix, with the individual as the "cluster".

When the model is overidentified ($k_1 > \ell_2$), you can use GMM estimation, which in principle is more efficient.

Hansen shows (p.651) how to estimate this model when the errors are homoscedastic, but this assumption is not credible, so this method is not recommended. However, it is possible to use this method and then calculate a cluster-robust covariance matrix, but it is not clear to me that this is preferred to the GMM method.

In **Sections 17.34 and 17.35**, Hansen briefly discusses how to use **jackknife and bootstrap methods** to calculate the variance-covariance matrices for panel data models. These can be useful if you are worried that the number of individuals (N, not n) is "small", so that the asymptotic approximations of variance-covariance matrices are not accurate.

III. Dynamic Panel Data Models (17.36 – 17.37)

For many economic and social phenomena the value of Y at time t could depend, in a causal sense, on the value of Y in previous time periods. For such panel data models, you should regress Y_{it} on both X_{it} and lagged values of Y_{it}. Such models are **dynamic panel data models**.

The general dynamic panel data model is:

$$Y_{it} = \alpha_1 Y_{i,t\text{-}1} + \alpha_2 Y_{i,t\text{-}2} + \ldots + \alpha_p Y_{i,t\text{-}p} + X_{it}'\beta + u_i + \varepsilon_{it} \quad (17.81)$$

As before, X_{it} is a $k\times1$ vector of regressors. Usually, **ε_{it} is assumed to be serially uncorrelated**, with a mean of 0.

For now, also **assume that X_{it} is strictly exogenous**. That is, $E[X_{is}\varepsilon_{it}] = 0$ for all t and for all $s = 1, 2, \ldots T$. **To demonstrate some concepts and methods, sometimes a simple AR(1)** panel data **model will be used**:

$$Y_{it} = \alpha Y_{i,t\text{-}1} + u_i + \varepsilon_{it} \qquad (17.82)$$

For both of the above models, u_i can be thought of as an individual-specific intercept term, so **in all applications of time series estimation the error term is ε_{it}, not u_i.**

For the model in (17.82), if $|\alpha| < 1$ then this model is stationary, and (similar to the AR(1) model in Lecture 10) backwards substitution allows (17.82) to be written as:

$$Y_{it} = \sum_{j=0}^{\infty} \alpha^j (u_i + \varepsilon_{it}) = (1 - \alpha)^{-1} u_i + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,t\text{-}j} \qquad (17.83)$$

Conditional on u_i the mean of Y_{it} is $(1 - \alpha)^{-1}u_i$, the variance of Y_{it} is $(1 - \alpha^2)^{-1}\sigma_\varepsilon^2$, and the k^{th} autocorrelation (conditional on u_i) is α^k. Thus u_i affects the mean but not the variance or autocorrelation.

Hansen points out that sometimes apparent autocorrelation (α's $\neq 0$) is simply a time trend, so detrending or differencing may reduce or eliminate autocorrelation.

The Bias of Fixed Effects Estimation

One possible approach to estimate the general model in (17.81) is using **fixed effects** (within) estimation. However, Nickel (1981) showed that when T is small this **yields inconsistent estimates**. To see the problem, apply the within estimator to the simple AR(1) model in (17.82):

$$\dot{Y}_{it} = \alpha \dot{Y}_{i,t\text{-}1} + \dot{\varepsilon}_{it}$$

Question 1: What happened to u_i?

Question 2: Why might OLS estimation (with fixed effects) of this equation by biased?

For simplicity, assume we have a balanced panel with $T =$ 3. In this case the within estimator is the same as the differenced estimator since there are, in effect, only two observations per person/unit. The above equation is then:

$$\Delta Y_{i3} = \alpha \Delta Y_{i2} + \Delta \varepsilon_{i3} \qquad (17.84)$$

The fixed effects estimate of this equation is simply OLS:

$$\hat{\alpha}_{\text{fe}} = \left(\sum_{i=1}^{N} \Delta Y_{i2}{}^2\right)^{-1}\left(\sum_{i=1}^{N} \Delta Y_{i2}\Delta Y_{i3}\right)$$

$$= \alpha + \left(\sum_{i=1}^{N} \Delta Y_{i2}{}^2\right)^{-1}\left(\sum_{i=1}^{N} \Delta Y_{i2}\Delta \varepsilon_{i3}\right)$$

The bias depends on $E[\Delta Y_{i2}\Delta\varepsilon_{i3}]$, which is:

$$E[\Delta Y_{i2}\Delta\varepsilon_{i3}] = E[(Y_{i2} - Y_{i1})(\varepsilon_{i3} - \varepsilon_{i2})]$$

$$= E[Y_{i2}\varepsilon_{i3}] - E[Y_{i1}\varepsilon_{i3}] - E[Y_{i2}\varepsilon_{i2}] + E[Y_{i1}\varepsilon_{i2}]$$

$$= 0 - 0 - \sigma_\varepsilon^2 + 0$$

$$= -\sigma_\varepsilon^2$$

You should be able to show that $E[(\Delta Y_{i2})^2] = 2\sigma_\varepsilon^2/(1 + \alpha)$, which implies that:

$$\underset{N \to \infty}{plim}\ (\hat{\alpha}_{\text{fe}} - \alpha) = \frac{E[\Delta Y_{i2}\Delta\varepsilon_{i3}]}{E[(\Delta Y_{i2})^2]} = -(1 + \alpha)/2 \qquad (17.85)$$

This bias is large! For $\alpha = 0$, it is -1/2. Even worse, (17.85) implies that $\text{plim}(\hat{\alpha}_{\text{fe}}) = \alpha - (1 + \alpha)/2 = \alpha/2 - 1/2$, so **for *any* $\alpha < 1$ we have $\hat{\alpha}_{\text{fe}}$ is negative!**

More generally, for the case where $T > 3$, Nickell (1981) showed that:

$$\underset{N \to \infty}{plim}\ (\hat{\alpha}_{\text{fe}} - \alpha) = \frac{1+\alpha}{\frac{2\alpha}{1-\alpha} - \frac{T-1}{1-\alpha^{T-1}}} \qquad (17.86)$$

In conclusion, **do not use standard fixed effects methods to estimate dynamic panel data models**.

IV. Estimation Methods for Dynamic Panel Data Models

So what can you use to estimate dynamic panel data models with fixed effects? This section reviews estimators that have been proposed.

Anderson-Hsiao Estimator

The idea here was to apply IV estimation after first differencing. To start, take the first difference of (17.81):

$$\Delta Y_{it} = \alpha_1 \Delta Y_{i,t\text{-}1} + \alpha_2 \Delta Y_{i,t\text{-}2} + \ldots + \alpha_p \Delta Y_{i,t\text{-}p} + \Delta X_{it}'\beta + \Delta\varepsilon_{it} \quad (17.87)$$

The problem is that $\Delta Y_{i,t\text{-}1}$ is correlated with $\Delta\varepsilon_{it}$:

$$\text{E}[\Delta Y_{i,t\text{-}1}\Delta\varepsilon_{it}] = \text{E}[(Y_{i,t\text{-}1} - Y_{i,t\text{-}2})(\varepsilon_{it} - \varepsilon_{i,t\text{-}1})] = -\sigma_\varepsilon^2$$

None of the other regressors are correlated with $\Delta\varepsilon_{it}$ (**why?**).

Anderson and Hsiao pointed out that $Y_{i,t\text{-}2}$ is a valid instrument for $\Delta Y_{i,t\text{-}1}$ because it is correlated with $\Delta Y_{i,t\text{-}1}$

but uncorrelated with $\Delta\varepsilon_{it}$. Hansen shows on p.656 that this yields a consistent estimate (for the special case where $T = 3$, $p = 1$ and there are no X_{it} variables).

This **can be implemented only if $T \geq p + 2$**. More generally, this estimation method can use $Y_{i,t\text{-}2}$, $Y_{i,t\text{-}3}$, … $Y_{i,t\text{-}p\text{-}1}$ as instruments for $\Delta Y_{i,t\text{-}1}$, $\Delta Y_{i,t\text{-}2}$, … $\Delta Y_{i,t\text{-}p}$.

Question: This estimator assumes that ε_{it} is serially uncorrelated. What problem could arise if ε_{it} is serially correlated?

Arellano-Bond Estimator

Arellano and Bond (1991) note that not only is $Y_{i,t\text{-}2}$ a valid IV for $\Delta Y_{i,t\text{-}1}$, but so are $Y_{i,t\text{-}3}$, $Y_{i,t\text{-}4}$, and other lagged values of Y_{it}. Their estimation method uses ***all*** lagged values of Y_{it} that are available, which means that the number of instruments increases as t is larger. This is best implemented using GMM estimation, as explained by Hansen on pp.657-658.

In theory, the **Arellano-Bond estimator** has a lower variance than the Anderson-Hsiao estimator, but it is **likely to suffer from a "many weak instruments" problem** if all available lagged values of Y_{it} are used as instruments. To avoid this problem, Hansen says that one should "limit the number of lags used as instruments".

Weak Instruments

Blundell and Bond (1998) pointed out that both the Anderson-Hsiao estimator and the Arellano-Bond estimator can suffer from the problem of weak instruments. Consider the case of the Anderson-Hsiao estimator for an AR(1) model. The **first-stage equation** for 2SLS/IV estimation is:

$$\Delta Y_{i,t-1} = \gamma Y_{i,t-2} + v_{it}$$

Blundell and Bond showed that:

$$\gamma = (\alpha - 1)\frac{k}{k+\sigma_u^2/\sigma_\varepsilon^2} \quad (17.97)$$

where $k = (1 - \alpha)/(1 + \alpha)$.

A weak instrument is one with γ close to 0. Equation (17.97) shows that this happens when α is close to 1 (recall that $\alpha = 1$ yields a unit root and thus a process that is not stationary) or when σ_ε^2 is small relative to σ_u^2 (most of the variation in Y_{it}, conditional on $Y_{i,t-1}$, is coming from variation in u_i, and differencing removes that variation).

Dynamic Models with "Predetermined" Regressors

The assumption that X_{it} is uncorrelated with **all** the error terms (strict exogeneity) is rather strong, and it can be weakened by assuming that the X_{it} variables are "predetermined". This is defined as follows:

Definition 17.2. The regressor X_{it} is **predetermined** for the error ε_{it} if:

$$E[X_{i,t-s}\varepsilon_{it}] = 0 \qquad (17.98)$$

for all $s \geq 0$.

This means that **"future" values of X_{it}** (i.e. $X_{i,t+1}$, $X_{i,t+2}$, etc.) **can be correlated with ε_{it}**, but current and past values of X_{it} are uncorrelated with ε_{it}.

What are the practical implications of assuming that X_{it} is predetermined instead of strictly exogenous? Recall equation (17.87):

$$\Delta Y_{it} = \alpha_1\Delta Y_{i,t-1} + \alpha_2\Delta Y_{i,t-2} + \ldots + \alpha_p\Delta Y_{i,t-p} + \Delta X_{it}'\beta + \Delta\varepsilon_{it} \quad (17.87)$$

If X_{it} is strictly exogenous then ΔX_{it} is not correlated with $\Delta\varepsilon_{it}$, but **if X_{it} is predetermined then ΔX_{it} is correlated with $\Delta\varepsilon_{it}$. [Why?]** To fix this problem, we need an instrument for ΔX_{it} that is not correlated with $\Delta\varepsilon_{it}$; **$X_{i,t-1}$ is**

a valid instrument. More generally, X_{i1}, X_{i2}, … $X_{i,t\text{-}1}$ are all valid instruments for ΔX_{it}. This is explained in more detail on pages 660-661 of Hansen.

Blundell-Bond Estimator

Recall the AR(1) model with no regressors:

$$Y_{it} = \alpha Y_{i,t\text{-}1} + u_i + \varepsilon_{it} \quad (17.82)$$

A simple OLS regression is biased and inconsistent because $Y_{i,t\text{-}1}$ is correlated with u_i. Blundell and Bond (1998) proposed to use $\Delta Y_{i,t\text{-}1}$ as an IV for $Y_{i,t\text{-}1}$. For more elaborate models that include $X_{i,t}$ variables that could be correlated with u_i, $\Delta X_{i,t}$ can be used as instruments. **If X_{it} is predetermined** than not only $\Delta X_{i,t}$ but also $\Delta X_{i,t\text{-}1}$ and $\Delta X_{i,t\text{-}2}$, etc. can be used as instruments for $X_{i,t}$, but **not** $\Delta X_{i,t+1}$, $\Delta X_{i,t+2}$, etc.

See Hansen (pp. 662-664) for more details on the Blundell-Bond estimator. They showed, using simulated data, that their estimator performed better than the Arellano-Bound estimator, especially when α is close to 1 and σ_ε^2 is small relative to σ_u^2. Note, however, that the Blundell-Bond estimator requires that X_t be stationary (actually, a slightly weaker assumption). It could also suffer from a "too many weak instruments" problem. Hanson says "it may be desirable to limit the number of lags used as instruments".