# Temporally Dynamic Panel-Data Models
## (Wawro again)

- How to model persistence?

1. Lags of the dependent variable (LDV) are included as regressors
   - ➢ Account for partial adjustment of behavior over time (e.g., to reach a long-run equilibrium).
   - ➢ Account for particular factors, including exogenous shocks, that have continual effects over time (coefficient on LDV indicate whether these factors have greater impact over time or whether their impact decays and the rate at which it decays).
   - ➢ Eliminate serial correlation in the disturbance term.
   - ➢ Parsimonious way of accounting for the persistent effects of explanatory variables w/o including their lags.
2. Individual-specific effects that do not vary over time
3. *Dynamic panel models* employ both of these approaches: dynamics plus individual-level heterogeneity.

# The Model:

- Consider the following representative regression model for dynamic panel data:

$$y_{i,t} = \gamma y_{i,t-1} + \beta x_{i,t} + \alpha_i + u_{i,t} \tag{6.1}$$

  where $i$ denotes the cross-sectional units ($i = 1, \ldots, N$), $t$ denotes the time period ($t = 1, \ldots, T$), $x_{i,t}$ is an exogenous explanatory variable, $\gamma$ and $\beta$ are parameters to be estimated, $\alpha_i$ is an individual-specific effect, and $u_{i,t}$ is a random disturbance term.

- $\alpha_i$ can be either fixed or random effects, since estimators have been derived for both cases.

- Assume

$$E[u_{i,t}|y_{i,t-1}, \ldots, y_{i,1}, x_{i,t}, x_{i,t-1}, \ldots, x_{i,1}] = 0. \tag{6.2}$$

- For now, also assume that the $u_{i,t}$ are serially uncorrelated and homoskedastic.

# OLS Biased if Inadequate Model Unit-Specific

- If we have not adequately accounted for individual-specific effects, then OLS is inappropriate; eq. 6.1 becomes

$$y_{i,t} = \gamma y_{i,t-1} + \beta x_{i,t} + u^*_{i,t} \tag{6.3}$$

where $u^*_{i,t} = \alpha_i + u_{i,t}$.

- To see why this is problematic, consider what happens if we lag eq. 6.1 one period:

$$y_{i,t-1} = \gamma y_{i,t-2} + \beta x_{i,t-1} + \alpha_i + u_{i,t-1} \tag{6.4}$$

- *By construction*, $y_{i,t-1}$ is correlated with $\alpha_i$. $\therefore$, $y_{i,t-1}$ is correlated with $u^*_{i,t}$.
- For the OLS estimates of $\gamma$ and $\beta$ to be unbiased,

$$E[u^*_{i,t}|y_{i,t-1}, x_{i,t}] = 0, \tag{6.5}$$

- Furthermore, the performance of OLS does not improve as sample size $\uparrow$, b/c the fundamental requirement for consistency is violated.

**I called this inverse Hurwicz/Nickell before. It's just another example of OVB.**

# LSDV/FE LDV is Small-($T$-)Sample Biased

- LSDV transformation to remove the individual effects produces biased and inconsistent estimates because correlation remains between the transformed lagged dependent variable and the transformed disturbance:

$$y_{i,t-1} - \bar{y}_{i,t-1}, \text{ where } \bar{y}_{i,t-1} = \sum_{t=2}^{T} y_{i,t-1}/(T-1) \tag{6.6}$$

$$u_{i,t} - \bar{u}_{i,t-1}, \text{ where } \bar{u}_{i,t-1} = \sum_{t=2}^{T} u_{i,t-1}/(T-1) \tag{6.7}$$

- Maybe okay as $T$ gets big—Hurwicz/Nickell bias.

**In fact, bias is "of order 1/*T*", which means proportionate bias is 1/*T*.**
**(So 5% for T=20, e.g., but 25% for *T*=4.)**
**(So big-to-huge issue for Panel, moderate-to-small issue for TSCS.)**

**Panel-analyst's response to conundrum:**

Model unit-effects, then redress Hurwicz Bias.

## 6.3 The Anderson-Hsiao Estimator

- Anderson and Hsiao ('81 *JASA*; '82 *J. Econometrics*) pointed out that first differencing eq. 6.10 eliminates the problem of correlation between the lagged endogenous variable and the individual-specific effect.

- First differencing eq. 6.1 gives

$$y_{i,t} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + \beta(x_{i,t} - x_{i,t-1}) + u_{i,t} - u_{i,t-1} \quad (6.8)$$

which can be rewritten as

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \beta \Delta x_{i,t-1} + \Delta u_{i,t} \quad (6.9)$$

where $\Delta$ is the difference operator such that the notation $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$.

- Still correlation between RHS variables and the disturbance term because $y_{i,t-1}$ in $\Delta y_{i,t-1}$ is by construction correlated with $u_{i,t-1}$ in $\Delta u_{i,t}$.

# Instruments from TSCS Structure Dataset…

- Still correlation between RHS variables and the disturbance term because $y_{i,t-1}$ in $\Delta y_{i,t-1}$ is by construction correlated with $u_{i,t-1}$ in $\Delta u_{i,t}$.
- Use IV w/ set of instruments conveniently supplied by the panel structure of the data.
- $y_{i,t-2} - y_{i,t-3}$ and $y_{i,t-2}$ are correlated with $y_{i,t-1} - y_{i,t-2}$ but not $u_{i,t} - u_{i,t-1}$ (assuming eq. 6.2 holds and there is no serial correlation).
- The same is true for $x_{i,t-2} - x_{i,t-3}$ and $x_{i,t-2}$.
- Suppose eq. 6.1 includes $y_{i,t-1}$ as the only explanatory variable:

$$y_{i,t} = \gamma y_{i,t-1} + \alpha_i + u_{i,t}. \tag{6.10}$$

# Anderson-Hsiao Insight & Problems

- Anderson and Hsiao ('81 *JASA*) showed that

$$\gamma_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t} \Delta y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} \qquad (6.11)$$

  and

$$\gamma_{IV} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t} y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \Delta y_{i,t-1} y_{i,t-2}} \qquad (6.12)$$

  are consistent estimators of $\gamma$.

- Anderson-Hsiao (A-H) estimators have some problems though.

- Arellano ('89 *Econ. Letters*) shows that the estimator given in eq. 6.11 has a singularity point as well as large variances over a range of values for $\gamma$.

- Arellano and Bover ('95 *J. Econometrics*, p. 46) concluded from a Monte Carlo study that a variant of this first-difference estimator is "useless" when $N = 100$, $T = 3$ and the coefficient on the lagged endogenous variable is .8.

- Others have shown that it is inefficient b/c it neglects important information in the data.

# "Improvements": GMM Estimators - i.e., (Asymptotic-)Efficiency Enhancers (if true)

- The subsequent improvements on A-H have built on their innovation of using IVs made available by the panel structure of the data. These studies have adopted the Generalized Method of Moments (GMM) framework to derive estimators that surmount the problems of A-H.
- GMM estimators: key intuition is that once the individual-specific effects are removed, the panel structure of the data provides a large number of IVs in the form of lagged endogenous and exogenous vari-
  - More generally, GMM : 2SLS : : 2SLS : IV – namely, uses more info & *id* restricts
  - Familiar "exclusion restrictions" regard 1st moment; extra GMM ones regard 2nd (&, in principle, higher) moments.
  - In implementation, can show GMM is matrix-wtd 2SLS. So could also say GMM : 2SLS : :GLS : OLS

# "Review" GMM Estimation (1)

- You've done GMM before—OLS and maximum likelihood can be derived as GMM estimators—just like GLS & 2SLS.
- Main idea: from a set of basic assumptions about a DGP, we can establish population moment conditions and then use sample analogs of these moment conditions to compute parameter estimates.
- Pop. moment conditions typically involve expectations of functions of the disturbance term and explanatory variables, while the sample analogs of the population moment conditions typically take the form of sample means.

MoM: use sample analog as estimator population parameter.

So, e.g.: $\bar{x}$ is a MoM estimator for $\mu_x$.

G-MoM: use higher moment conditions too.

So, G-MoM : MoM : : GLS : OLS

# "Review" GMM Estimation (2)

- Consider the cross-sectional regression

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i \tag{6.13}$$

where we adopt the key identifying assumption

$$E\left[\mathbf{x}_i' u_i\right] = \mathbf{0} \tag{6.14}$$

(here $\mathbf{x}_i$ is a $1 \times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters to be estimated, and $u_i$ is the disturbance).

- This basic assumption defines a set of moment conditions and is a weaker variant of the assumption in eq. 6.5 discussed above (violated in a dynamic panel setting).

- Substituting in for $u_i$, we can rewrite eq. 6.14 as

$$E\left[\mathbf{x}_i'(y_i - \mathbf{x}_i \boldsymbol{\beta})\right] = \mathbf{0} \tag{6.15}$$

to get the moment conditions in terms of observables and parameters.

OLS is also a MoM, in other words...

- Pop. moments are estimated *consistently* with sample moments, so the next step is to write down the sample analog of eq. 6.15:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) = \mathbf{0} \tag{6.16}$$

where $\hat{\boldsymbol{\beta}}$ is our estimator.

- Multiplying this out and solving for $\hat{\boldsymbol{\beta}}$ gives

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{x}_i'y_i\right), \tag{6.17}$$

which is identical to the equation for the OLS estimator of $\boldsymbol{\beta}$.

- We can rewrite eq. 6.17 as $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ by stacking the $\mathbf{x}_i$ and $y_i$ for observations $i = 1, \ldots, N$ into an $N \times K$ matrix $\mathbf{X}$ and $N \times 1$ vector $\mathbf{y}$, respectively.

# Instrumental Variables in GMM Framework

- Suppose eq. 6.14 does not hold, for example, because $x_{i,k}$ in $\mathbf{x}_i$ is correlated with $u_i$.

- Suppose also that there are some variables $\mathbf{z}_i$ available for which

$$E\left[\mathbf{z}_i' u_i\right] = \mathbf{0} \tag{6.18}$$

  does hold and that the elements of $\mathbf{z}_i$ are partially correlated with $x_{i,k}$.

- Then $\mathbf{z}_i$ can serve as instrumental variables.
- The pop. moment conditions for the GMM estimator of $\boldsymbol{\beta}$ are

$$E\left[\mathbf{z}_i'(y_i - \mathbf{x}_i\boldsymbol{\beta})\right] = \mathbf{0} \tag{6.19}$$

  which have the sample analog

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i' \left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) = \mathbf{0}. \tag{6.20}$$

- If the number of columns in $\mathbf{z}_i$ (i.e., the number of moment conditions) > the number of parameters to be estimated (which is typically the case), then our equation is overidentified and there is not a closed form solution as with eq. 6.16 (which was just identified).

# Just identified = IV; Overidentified = 2SLS

## *Capitalizing on* the Overidentifying Information

- To get around this problem we choose $\boldsymbol{\beta}$ so that it minimizes the quadratic

$$\left(\sum_{i=1}^{N} \mathbf{z}_i' \left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right)' \mathbf{W} \left(\sum_{i=1}^{N} \mathbf{z}_i' \left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right), \qquad (6.21)$$

where $\mathbf{W}$ is a positive semi-definite weighting matrix.

- The solution to this minimization problem does have a closed form, and with a little manipulation, we obtain

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{X}\right)^{-1} \left(\mathbf{X}'\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{y}\right). \qquad (6.22)$$

- Note the similarities between this GMM estimator and expressions for 2SLS estimators (note the $\mathbf{Z}$s) and GLS estimators (note the $\mathbf{W}$s).

- It can be shown that the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is

$$\boldsymbol{\Omega} = \left(E\left[\mathbf{X}_i'\mathbf{Z}_i\right]\mathbf{W}E\left[\mathbf{Z}_i'\mathbf{X}_i\right]\right)^{-1} E\left[\mathbf{X}_i'\mathbf{Z}_i\right]\mathbf{W}\mathbf{V}\mathbf{W}E\left[\mathbf{Z}_i'\mathbf{X}_i\right]\left(E\left[\mathbf{X}_i'\mathbf{Z}_i\right]\mathbf{W}E\left[\mathbf{Z}_i'\mathbf{X}_i\right]\right)^{-1} \qquad (6.23)$$

where

$$\mathbf{V} = \mathrm{Var}\left[\mathbf{Z}_i'u_i\right] = E\left[\mathbf{Z}_i'u_iu_i'\mathbf{Z}_i\right].$$

# Optimizing the Weights in GMM

- The efficiency of the GMM estimator depends crucially on the choice of $\mathbf{W}$. In order to obtain an efficient estimator, we should choose $\mathbf{W}$ so that it makes $\mathbf{\Omega}$ as small as possible.

- The choice of $\mathbf{W}$ that does this is $\mathbf{W} = \mathbf{V}^{-1}$ (Hansen '82 *Ecta*).

- Substituting in $\mathbf{V}^{-1}$ for $\mathbf{W}$ in eq. 6.23 and canceling terms substantially simplifies the expression for the asymptotic variance, which becomes

$$\mathbf{\Omega} = (\mathbf{X}_i'\mathbf{Z}_i\mathbf{V}^{-1}\mathbf{Z}_i'\mathbf{X}_i)^{-1}. \tag{6.24}$$

- Next step: come up with a consistent estimate for $\mathbf{V}$.

- Do not get to observe $u_i$, so use estimated residuals produced by calculating $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}^*$, where $\hat{\boldsymbol{\beta}}^*$ is a first-stage, consistent estimator of $\boldsymbol{\beta}$.

- In the first stage, we typically use in eq. 6.22 the weighting matrix $\hat{\mathbf{W}}_1 = (\mathbf{Z}'\mathbf{Z})^{-1}$ to obtain $\hat{\boldsymbol{\beta}}^*$.

- The weighting matrix we use in the second stage, which is a consistent estimator for $\mathbf{V}^{-1}$, is

$$\hat{\mathbf{W}} = \hat{\mathbf{V}}^{-1} = \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \hat{u}_i \hat{u}_i' \mathbf{Z}_i] \right\}^{-1} . \qquad (6.25)$$

- Plugging in $\hat{\mathbf{W}}$ and $\hat{\mathbf{V}}^{-1}$ in equations (6.22) and (6.24) produces the asymptotically optimal GMM estimator.

- Note that if we assume the disturbances are homoskedastic and not serially correlated, then it would be optimal to use $(\mathbf{Z}'\mathbf{Z})^{-1}$ for $\hat{\mathbf{W}}$.

- However, using the weighting matrix given by eq. 6.25 assures that our standard errors, which we compute by taking the square root of the diagonal of

$$\hat{\mathbf{\Omega}} = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}^{-1}\mathbf{Z}'\mathbf{X})^{-1}, \tag{6.26}$$

are robust to nonspherical disturbances.

- Hansen ('82 *Ecta*) shows that GMM estimators are consistent and $\overset{a}{\sim}$ normal. Thus, if an estimator can be shown to be a GMM estimator (i.e., can be derived using the GMM framework just discussed) then the "goodness" properties of consistency and asymptotic efficiency automatically follow.

  - E.g, it follows that $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically distributed as $N(\boldsymbol{\beta}, \mathbf{\Omega})$.

- GMM estimators for DPD have same basic form as for cross-sectional models.

- Key feature: exploit the panel structure of the data to construct instruments that satisfy moment conditions like eq. 6.19.

# GMM for (Temporally) Dynamic Panel-Data

## Arrellano-Bond enhance Anderson-Hsiao

- If assume $E(u_{i,t}) = E(u_{i,t}u_{i,s}) = 0$, then the transformed residuals in eq. 6.8 have zero covariance between all $y_{i,t}$ and $x_{i,t}$ dated $t-2$ and earlier. This means we can go back through the panel from period $t-2$ to obtain appropriate instrumental variables for purging the correlation between $\Delta y_{i,t-1}$ and $\Delta u_{i,t}$. The transformed residuals satisfy a large number of moment conditions of the form

$$E[\mathbf{z}'_{i,t}\Delta u_{i,t}] = \mathbf{0}, \quad t = 2, \ldots, T, \qquad (6.27)$$

where $\mathbf{z}_{i,t} = (y_{i,t-2}, x_{i,t-2}, y_{i,t-3}, x_{i,t-3}, \ldots, y_{i,1}, x_{i,1})'$ denotes the instrument set at period $t$.

- For notational efficiency, we can stack the time periods to write dowr a system of $T$ equations for each individual:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i \qquad (6.28$$

where

$$\mathbf{y}_i = \begin{bmatrix} \Delta y_{i,3} \\ \Delta y_{i,4} \\ \vdots \\ \Delta y_{i,T} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} \Delta y_{i,2} & \Delta x_{i,3} \\ \Delta y_{i,3} & \Delta x_{i,4} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{i,T} \end{bmatrix}, \text{ and } \mathbf{u}_i = \begin{bmatrix} \Delta u_{i,3} \\ \Delta u_{i,4} \\ \vdots \\ \Delta u_{i,T} \end{bmatrix}.$$

- The set of instruments is given by the block diagonal matrix

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i,3} & & & 0 \\ & \mathbf{z}_{i,4} & & \\ & & \ddots & \\ 0 & & & \mathbf{z}_{i,T} \end{bmatrix}.$$

- Note that this means that the number of instruments increases as move through the panel. For example, if we have the simple mo in eq. 6.10, then the instrument matrix becomes

$$\underset{(T-2)\times(T-2)(T-1)/2}{\mathbf{Z}_i^*} = \begin{bmatrix} y_{i,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i,1} & y_{i,2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,1} & y_{i,2} & \cdots & y_{i,T-2} \end{bmatrix}.$$

- Hence, if $T = 5$ then we would have

$$\mathbf{Z}_i^* = \begin{bmatrix} y_{i,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i,1} & y_{i,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i,1} & y_{i,2} & y_{i,3} \end{bmatrix}.$$

Issue in 2SLS/GMM to keep rank(**Z**) not too large relative to rank(endog)…b/c overfitting $1^{st}$-stage => fitting some of endog.

- The vector of population moment conditions is

$$E[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}. \tag{6.29}$$

- The sample analog of eq. 6.29 that we use to construct an optimal GMM estimator for $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta})$ is

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{u}_i = \mathbf{0}. \tag{6.30}$$

- Let

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \text{ and } \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}$$

(just stack the observations for all of the cross-sectional units for all

- Then we can re-express eq. 6.30 as

$$\frac{1}{N}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\theta})=\mathbf{0}.$$

- The optimal GMM estimator is then given by

$$\hat{\boldsymbol{\theta}}=(\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}^{-1}\mathbf{Z}'\mathbf{y}, \tag{6.31}$$

  where $\hat{\mathbf{V}}$ is a consistent estimate of $\mathbf{V}$, the limiting variance of the sample moments $E[\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i]$.

- If we assume conditional homoskedasticity and no autocorrelation, then the optimal choice for $\hat{\mathbf{V}}$ is $\hat{\mathbf{V}}_c=\mathbf{Z}'\mathbf{Z}$.

- But typically, want to compute 2nd stage, robust estimate. In general, the optimal choice for $\hat{\mathbf{V}}$ is

$$\hat{\mathbf{V}}_r=\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i,$$

  where $\hat{\mathbf{u}}_i$ is an estimate of the vector of residuals, $u_{i,t}$, obtained from an initial consistent estimator.

- If we assume conditional homoskedasticity and no autocorrelation, then the optimal choice for $\hat{\mathbf{V}}$ is $\hat{\mathbf{V}}_c = \mathbf{Z}'\mathbf{Z}$.

- But typically, want to compute 2nd stage, robust estimate. In general, the optimal choice for $\hat{\mathbf{V}}$ is

$$\hat{\mathbf{V}}_r = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i,$$

  where $\hat{\mathbf{u}}_i$ is an estimate of the vector of residuals, $u_{i,t}$, obtained from an initial consistent estimator.

- Arellano and Bond ('91 *Rev. Econ. Studies*) suggest using $\hat{\mathbf{V}}_c = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i$ to produce the initial consistent estimator, where

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

- By the properties of GMM estimators, with $T$ fixed and $N \to \infty$, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically distributed as $N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ (Hansen '82 *Ecta*).

- The asymptotic variance $\boldsymbol{\Sigma}$ is equal to

$$\left\{E(\mathbf{X}_i'\mathbf{Z}_i)E[\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i]^{-1}E(\mathbf{Z}_i'\mathbf{X}_i)\right\}^{-1}.$$

  A consistent estimator of the asymptotic variance is

$$\hat{\boldsymbol{\Sigma}} = \left(\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}_r^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}.$$

- Std. errs. for the first-difference estimates are obtained by taking the square root of the diagonal of $\hat{\boldsymbol{\Sigma}}$.

- If the disturbances are heteroskedastic, then the two-step estimator is more efficient. In practice, however, the asymptotic standard errors for the one-step estimator appear to be more reliable for making inferences in small samples (Arellano and Bond '91 *Rev. Econ. Studies*; Blundell and Bond '98 *J. Econometrics*).

**<u>TO RESTATE</u>: Properties are Asymptotic, <u>*in N*</u>**

# Twin Horns of Dilemma for Estimator:

- Instruments too weak, Instruments too many.

### 6.6.2 Finite sample considerations

- Bias/efficiency trade-off that starts to bite as $T$ increases in size (relative to $N$) $\Rightarrow$ we may not want to use all available instruments.
- More instruments become available as $T$ increases, but instruments from earlier periods in the panel become weaker the farther we progress through the panel.
- Using all of the instruments is efficient but can cause severe downward bias in GMM estimators when our sample is finite (Ziliak '97 *J. Bus. & Econ. Stats*)—overfitting.

## - Cottage Industry of Further Enhancements:

- Changes as well as levels (& v.v.) equally valid instruments.
- "Orthogonal Deviations" (actually forward obs: leads)
- "Robustness" to non-spherical V-Cov

### 6.6.3 Specification tests

- Consistency of estimators depends crucially on the assumption that the $u_{i,t}$ in eq. 6.1 are serially uncorrelated.

- If serial correlation exists, then some of our instruments will be invalid and the moment conditions used to identify parameters will not hold.

- Should test for serial correlation.
- If no serial correlation in the $u_{i,t}$ in eq. 6.1, then the first-differenced residuals should display negative 1st-order serial correlation but not 2nd-order serial correlation:

  - First differencing produces the MA(1), process $u_{i,t} - u_{i,t-1}$.
  - If our disturbances for the levels equation are $u_{i,t} - \rho u_{i,t-1}$, then differencing gives $u_{i,t} - u_{i,t-1} - \rho(u_{i,t-1} - u_{i,t-2})$
  - $y_{i,t-2}$ not valid as an instrument since it will be correlated with $u_{i,t-2}$ in the differenced disturbance term (although lagged $y$s at period $t-3$ and earlier remain valid instruments).

- Arellano and Bond ('91 *Rev. Econ. Studies*) give tests of 1st- and 2nd-order serial correlation based on the residuals from the two-step estimator of the first-differenced equation.

## AB's Omnibus Over-ID Test:

- Variant of the Sargan test (cf. Sargan '58 *Ecta*; Hansen '82 *Ecta*):

$$s = \hat{\mathbf{u}}'\mathbf{Z}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}'\hat{\mathbf{u}} \qquad (6.41)$$

  where $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N)'$, the stacked vectors of estimated first-differenced residuals for all $i$ and $T$.

- $s \overset{a}{\sim} \chi^2$ w/ df = number of columns of $\mathbf{Z}$ minus the number of explanatory variables.

- Significant $\chi^2$ value $\Rightarrow$ overidentifying restrictions are invalid.

- Intuition: if the moment conditions given by eq. 6.29 hold, then the sample moments given by eq. 6.30 when evaluated at the parameter estimates should be close to zero, and hence the value of the quadratic function in eq. 6.41 should be small.

- Rejection of the overidentifying restrictions should lead one to reconsider the specification of the model, possibly reducing the number of instruments employed or including more lags to eliminate serial correlation.

- Can use differences between Sargan test statistics to test the validity of additional moment conditions.

  ➢ The difference between the Sargan statistics $\sim \chi^2$ w/ df = number of new moment conditions that are used.

**(Help xtdpdsys)**