5. GENERALIZED METHODS OF MOMENTS (GMM)

[1] THE PRINCIPLE OF GMM

(1) Notation:

- θ : p×1 vector of possible parameters.
- $g_t(\theta) \equiv g(\theta, w_t)$: q×1 vector of functions of θ and data (w_t).

    [$g_t(\theta)$ is called moment function.]
- $g_T(\theta) = \frac{1}{T}\Sigma_t g_t(\theta)$: sample mean of $g_t(\theta)$.
- $G_T(\theta) = \frac{\partial g_T(\theta)}{\partial \theta'}$: q×p matrix.
- Identification condition: $q \geq p$

(2) Assumptions:

- $E[g_t(\theta)] = 0$ iff θ = the true value of θ (θ_o).

    [Called orthogonality (moment) conditions.]
- $E[G_T(\theta_o)]$ is full column.
- $\frac{1}{\sqrt{T}}\Sigma_t g_t(\theta_o) = \sqrt{T} g_T(\theta_o) \rightarrow_d N(0,V)$.

(3) About V:

- Let $\Gamma_i = E[g_t(\theta_o)g_{t-i}(\theta_o)']$.
- Hansen (1982, Econ) shows, under stationarity conditions:

$$V = \Gamma_0 + \Sigma_{i=1}^{\infty}(\Gamma_i + \Gamma_i').$$

- Estimation of V when the w_t are autocorrelated over t:
  - See Newey and West (1987, Econometrica), Andrews (1991, Econometrica), or Andrews and Monahan (1992, Econometrica).
- Estimation of V when the w_t are independent over t:

$$\hat{V} = \frac{1}{T}\Sigma_t g_t(\hat{\theta})g_t(\hat{\theta})',$$

where $\hat{\theta}$ is any consistent estimator.

(4) Example:

- Consider a regression model:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t, \text{ t = 1, ..., T.}$$

- Assume $E(x_{t\bullet}\varepsilon_t) = 0$.
- $g_t(\beta) = x_{t\bullet}(y_t - x_{t\bullet}'\beta)$: $E[g_t(\beta)] = E[x_{t\bullet}(y_t - x_{t\bullet}'\beta)] = E[x_{t\bullet}\varepsilon_t] = 0$.
- $g_T(\beta) = \frac{1}{T}\Sigma_t x_{t\bullet}(y_t - x_{t\bullet}'\beta) = T^{-1}X'(y-X\beta)$.
- For cross-section data,

$$\hat{V} = \frac{1}{T}\Sigma_{t=1}^{T} g_t(\hat{\theta})g_t(\hat{\theta})' = \frac{1}{T}\Sigma_{t=1}^{T} e_t^2 x_{t\bullet}x_{t\bullet}',$$

where $\hat{\beta}$ is the OLS estimator and e_t = OLS residual.

(5) Definition of GMM estimator (Hansen, 1982, Econometrica):

Let A be a q×q weighting matrix which is pd.

A GMM estimator, $\hat{\theta}_A$, minimizes:

$$Tg_T(\theta)' A^{-1} g_T(\theta).$$

[2] PROPERTIES OF GMM ESTIMATORS

(1) Facts:

- Consistent and asymptotically normal.
- The covariance matrix of $\hat{\theta}_A$ is estimated by:

$$\frac{1}{T}\left(\hat{G}'A^{-1}\hat{G}\right)^{-1}\hat{G}'A^{-1}\hat{V}A^{-1}\hat{G}\left(\hat{G}'A^{-1}\hat{G}\right)^{-1},$$

where $\hat{G} = G_T(\hat{\theta}_A)$.

(2) The optimal choice of A

- Hansen (1982) shows that $\hat{V}$ is the optimal (efficient) choice of A.
- The optimal GMM estimator, $\tilde{\theta}$, minimizes:

$$Tg_T(\theta)'\hat{V}^{-1}g_T(\theta).$$

- $Cov(\tilde{\theta}) = \left(T\tilde{G}'\hat{V}^{-1}\tilde{G}\right)^{-1}$, where $\tilde{G} = G_T(\tilde{\theta})$.

(3) Example:

- Consider a regression model:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t, \text{ t = 1, ..., T.}$$

- Assume $E(x_{t\bullet}\varepsilon_t) = 0$.
- $g_t(\beta) = x_{t\bullet}(y_t - x_{t\bullet}'\beta)$: $E[g_t(\beta)] = E[x_{t\bullet}(y_t - x_{t\bullet}'\beta)] = E[x_{t\bullet}\varepsilon_t] = 0$.
- $g_T(\beta) = \frac{1}{T}\Sigma_t x_{t\bullet}(y_t - x_{t\bullet}'\beta) = T^{-1}X'(y - X\beta)$.

- For cross-section data,

$$\hat{V} = \frac{1}{T}\Sigma_{t=1}^{T} g_t(\hat{\theta}) g_t(\hat{\theta})' = \frac{1}{T}\Sigma_{t=1}^{T} e_t^2 x_{t\bullet} x_{t\bullet}',$$

where $\hat{\beta}$ is the OLS estimator and e_t = OLS residual.

- For any pd matrix A, minimize

$$\frac{1}{T}(y - X\beta)' X A^{-1} X'(y - X\beta).$$

- FOC:

$- 2X'XA^{-1}X'(y-X\beta) = 0$

$\rightarrow X'(y-X\beta) = 0.$

$\rightarrow \hat{\beta} = (X'X)^{-1}X'y$.

- The OLS estimator is the GMM estimator based on the moment condition $E(x_{t\bullet}\varepsilon_t) = 0$.
- Notice that the GMM estimator (OLS) does not depend on A. In fact, if p = q (cases of exact identification), GMM estimators do not depend on A. That is, the GMM estimator is unique in cases of exact identification.

[3] THREE WAYS TO OBTAIN OPTIMAL GMM

- Need an initial estimate:

  min. $Tg_T(\theta)'A^{-1}g_T(\theta)$, setting A = I.

(1) Two-step GMM

STEP 1: Compute $\hat{V}$ using $\hat{\theta}_A$.

STEP 2: Min. $Tg_T(\theta)'\hat{V}^{-1}g_T(\theta)$, and get $\tilde{\theta}$.

(2) Iterative GMM

STEP 3: Compute $\tilde{V}$ using $\tilde{\theta}$.

STEP 4: Min. $Tg_T(\theta)'\tilde{V}^{-1}g_T(\theta)$, and get $\bar{\bar{\theta}}$.

STEP 5: Do while estimates do not change.

(3) Continuous-updating GMM

(Hansen, Heaton and Yaron, 1996, JBES)

- Note that $\hat{V}$ is a function of θ: $\hat{V}(\theta)$.
- Min. $Tg_T(\theta)'[\hat{V}(\theta)]^{-1}g_T(\theta)$.

(4) Facts:

- All three GMMs are asymptotically idenitcal.
- For finite sample, C-U GMM seems to perform better than two other GMM.
- No unaminous results for two-step and iterative GMM.

[4] ESTIMATION OF COVARIANCE MATRICES.

- See Hamilton (Ch. 10)

Question: How can we estimate $V = \Gamma_0 + \Sigma_{i=1}^{\infty}\left(\Gamma_i + \Gamma_i'\right)$.

Answer: Use nonparametric methods.

- Define $S_0 = \frac{1}{T}\Sigma_t g_t(\hat{\theta})g_t(\hat{\theta})'$;

$$S_i = \frac{1}{T}\Sigma_{t=i+1}^{T} g_t(\hat{\theta})g_{t-i}(\hat{\theta})'.$$

$$\hat{V} = S_0 + \Sigma_{i=1}^{b} k\left(\frac{i}{b+1}\right)\left(S_i + S_i'\right),$$

  where b is called bandwidth and k(•) is a kernel function.

- Andrews (1991, ECON) provides a list of kernels that warrant psd $\hat{V}$ matrices.

Question: How can we choose b?

- If we believe that $\Gamma_i = 0$ for $i > \tau$ [it happens if $g_t(\theta_o)$ follows MA(τ)], choose $b = \tau$ or more.
- If we don't know? Choose b such that $b \to \infty$ and $b/T^{1/2} \to 0$, as $T \to \infty$. [Andrews (1991, ECON)].
- The theoretically optimal b (say b^*), which minimizes some MSE measure, depends on the choice of kernel.
- See Newey and West (1994, RESTUD) for the data-dependent optimal choice of b.

(1) Newey-West Method (1987, ECON)

- Use Bartlett's kernel, k(z) = 1 - z.
- $b^* \propto T^{1/3}$ (Andrews (1991, ECON), Newey and West (1994, RESTUD)).

(2) Gallant Method (1987, Nonlinear Statistic Models)

- Use Pazen kernel,

  k(z) = $1-6z^2+ 6z^3$ for $0 \le z \le 1/2$;

  = $2(1-z)^3$, $½ \le z \le 1$;

  = 0, otherwise.
- $b^* \propto T^{2/5}$.

(3) Andrews (1991, ECON)

- Use a quadratic kernel:

$$k(z) = \frac{3}{(6\pi z/5)^2}\left(\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right).$$

- $b^* \propto T^{2/5}$.
- $\hat{V} = \frac{T}{T-p}\left(S_0 + \Sigma_{i=1}^{T-1} k\left(\frac{i}{b+1}\right)(S_i + S_i')\right).$

(4) Others:

- Newey-West (1994, RESTUD)
- Andrews-Monahan (1992, ECON)
- West (1997, Journal of Econometrics)

[5] TESTING HYPOTHESES

- $\hat{\theta}$: a GMM estimator; and $\hat{\Omega} = Cov(\hat{\theta})$.

- Wald test for H_o: $w(\theta_o) = 0_{m\times 1}$: $W_T = w(\hat{\theta})'\left(W(\hat{\theta})\hat{\Omega}W(\hat{\theta})'\right)^{-1} w(\hat{\theta})$.

- For other tests, see Newey and West (1987, IER):
  - LR-type test:
    - Let $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ where θ_1 is $p_1\times 1$ vector
    - H_o: $\theta_{1,o} = 0_{p_1\times 1}$.
    - Do optimal GMM using the moment condition $E\left(g_t(0_{p_1\times 1}, \theta_2)\right) = 0_{q\times 1}$.
    - Denote the estimator of θ_2 from this GMM by $\breve{\theta}_2$.
    - Let $\breve{\theta} = \begin{pmatrix} 0_{p_1\times 1} \\ \breve{\theta}_2 \end{pmatrix}$.
    - Then, under H_o, $T\tilde{g}'\hat{V}^{-1}\tilde{g} - T\breve{g}'\hat{V}^{-1}\breve{g} \to_d \chi^2(p_1)$.

[6] SPECIFICATION TESTS

(1) Testing H_o: $E[g_t(\theta)] = 0$.

- Suppose that q (# of moment conditions) > p (# of parameters).

$$J_T \equiv T g_T(\tilde{\theta})' \hat{V}^{-1} g_T(\tilde{\theta}) \to_d \chi^2(df = q - p)$$

- $J_T = 0$ if q = p (exact identification). So, it is not possible to test the hypothesis. [If q = p, $g_T(\tilde{\theta}) = 0$.]

- Test based on any consistent estimator, $\hat{\theta}$.

$$MJ_T = T\left[\hat{g}'\hat{V}^{-1}\hat{g} - \hat{g}'\hat{V}^{-1}\hat{G}\left(\hat{G}'\hat{V}^{-1}\hat{G}\right)^{-1}\hat{G}'\hat{V}^{-1}\hat{g} \right]$$

= SSE from a GLS regression of $\hat{g}$ on $\hat{G}$,

where $\hat{g} = g_T(\hat{\theta})$ and $\hat{G} = G_T(\hat{\theta})$.

- MJ_T is asymptotically identical to J_T. [Ahn (1995).]
- Virtually almost all GMM specification tests are of the form MJ_T.

(2) Testing subsets of moment conditions

- Partition:

$$g_t(\theta) = \begin{pmatrix} b_t(\theta) \\ c_t(\theta) \end{pmatrix},$$

where $b_t(\theta)$ and $c_t(\theta)$ are $q_b \times 1$ and $q_c \times 1$ vectors, respectively, and $q_b + q_c = q$. Assume that $q_b \geq p$, so that θ_o can be estimated by $b_t(\theta)$ only.

- Definitions:
  - $b_T = \frac{1}{T}\Sigma_t b_t(\theta)$; $c_T(\theta) = \frac{1}{T}\Sigma_t c_t(\theta)$.
  - $B_T(\theta) = \frac{\partial b_T(\theta)}{\partial \theta'}$; $C_T = \frac{\partial c_T(\theta)}{\partial \theta'}$.
  - Corresponding to $b_T(\theta)$ and $c_T(\theta)$, we also divide $\hat{V}$ into

$$\begin{pmatrix} \hat{V}_{bb} & \hat{V}_{bc} \\ \hat{V}_{cb} & \hat{V}_{cc} \end{pmatrix}.$$

- H_o: $E[g_t(\theta)] = 0$ against $H_A^*: E(b_t(\theta_o)) = 0 \text{ and } E(c_t(\theta_o)) \neq 0$.
- Eichenbaum, Hansen and Singleton [JPE, 1988]
  - $\widehat{\theta}$ minimizes $Tb_T(\theta)'\left(\hat{V}_{bb}\right)^{-1} b_T(\theta)$.
  - The EHS statistic is akin to likelihood ratio test:

$$D_T = T\tilde{g}_T'\hat{V}^{-1}\tilde{g}_T - T\widehat{b}_T'\left(\hat{V}_{bb}\right)^{-1}\widehat{b}_T,$$

where $\widehat{b}_T = b_T(\widehat{\theta})$.

  - This statistic is asymptotically $\chi^2(q_c)$.

- Comments:
  - More powerful than the Hansen test if it is truly that $E[b_t(\theta)] = 0$.
  - Newey (1985, Journal of Econometrics) considers a Wald-type test. His test is asymptotically identical to EHS. For the class of tests that are identical to EHS, refer to Ahn (1995).

[7] EXAMPLES FOR GMM

[7.1] Consider a regression model:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t, \text{ t = 1, ..., T.}$$

- Assume $E(x_{t\bullet}\varepsilon_t) = 0$.
- $g_t(\beta) = x_{t\bullet}(y_t - x_{t\bullet}'\beta)$: $E[g_t(\beta)] = E[x_{t\bullet}(y_t - x_{t\bullet}'\beta)] = E[x_{t\bullet}\varepsilon_t] = 0$.
- $g_T(\beta) = \frac{1}{T}\Sigma_t x_{t\bullet}(y_t - x_{t\bullet}'\beta) = T^{-1}X'(y - X\beta)$.
- For cross-section data,

$$\hat{V} = \frac{1}{T}\Sigma_{t=1}^{T} g_t(\hat{\theta})g_t(\hat{\theta})' = \frac{1}{T}\Sigma_{t=1}^{T} e_t^2 x_{t\bullet}x_{t\bullet}' \equiv \frac{1}{T}\hat{\Delta},$$

  where $\hat{\beta}$ is the OLS estimator and e_t = OLS residual.

  - GMM estimator = $\hat{\beta} = (X'X)^{-1}X'y$ (OLS).

- $Cov(\hat{\beta})$ is given:

$$\frac{1}{T}\left(G_T(\hat{\beta})'\hat{V}^{-1}G_T(\hat{\beta})\right)^{-1} = \left[(X'X)\hat{\Delta}^{-1}(X'X)^{-1}\right]^{-1} = (X'X)^{-1}\hat{\Delta}(X'X)^{-1}.$$

[7.2] Consumption Decision Model

[Hansen and Singleton (1982, Econometrica)]

- Utility function of a representative consumer: $u(c_t) = [1/(1-\alpha)]c_t^{1-\alpha}$.
- The consumer maximizes expected utility given information set available at time t (Ω_t) :

$$E\left(\Sigma_{j=0}^{\infty} \beta^j \frac{c_{t+j}^{1-\alpha}}{1-\alpha} | \Omega_t \right),$$

subject to the intertemporal budget constraint:

$$c_t + A_t p_t = A_{t-1}(p_t + d_t).$$

Here, A_t: the amount of asset at time t;

p_t: price of asset at time t;

d_t: dividend at time t;

$c_t + A_t p_t$: expenditure at time t

$A_{t-1}(p_t+d_t)$: total value of asset at time t.

- FOC:

$$E\left(v_t(\alpha,\beta) | \Omega_t\right) = 0,$$

where $v_t(\alpha,\beta) = \beta[c_t/c_{t-1}]^{\alpha} r_t - 1$; $r_t = (p_t+d_t)/p_{t-1} = 1 +$ rate of return.

- Any variable in Ω_t should be uncorrelated with $v_t(\alpha,\beta)$.
- $\Omega_t = \{1, c_{t-1}/c_{t-2}, c_{t-2}/c_{t-3}, \ldots, r_{t-1}, r_{t-2}, \ldots\}$. Let w_t be a vector of some variables in Ω_t. Then,

$$E[w_t v_t(\alpha,\beta)] = 0.$$

[7.3] Models with endogenous regressors and non-Gaussian errors

1. Estimation of a single equation model:

$$y_t = z_{t\bullet}'\delta + \varepsilon_t,\ t = 1, \ldots, T,$$

where δ is p×1.

(1) Assumptions:

- $z_{t\bullet}$ could be correlated with ε_t.
- There exists a q×1 (q ≥ p) vector of instrumental variables, $x_{t\bullet}$, such that $E(x_{t\bullet}\varepsilon_t) = 0$.

(2) Estimation Methods:

(2.1) 2SLS:

$$\hat{\delta} = [Z'P(X)Z]^{-1}Z'P(X)y, \text{ where } P(X) = X(X'X)^{-1}X'.$$

- Consistent.
- If the ε_t are iid over t and independent of $x_{t\bullet}$,

$$Cov(\hat{\delta}) = s_\varepsilon^2(Z'P(X)Z)^{-1} = s_\varepsilon^2[Z'X(X'X)^{-1}X'Z]^{-1},$$

  where $Z'X = \Sigma_{t=1}^T z_{t\bullet}x_{t\bullet}'$ and $s_\varepsilon^2 = \frac{1}{T}\Sigma_t\left(y_t - x_{t\bullet}'\hat{\delta}\right)^2$.

- A GMM based on $E[x_{t\bullet}(y_t - z_{t\bullet}'\delta)] = 0$ with $A = s_\varepsilon^2(X'X/T)$.
- If the ε_t are heteroskedastic, this estimated covariance matrix is incorrect, even though the 2SLS estimator is still consistent.

(2.2) Optimal GMM ($\tilde{\delta}$).

- Moment function:

  $g_t(\delta) = x_{t\bullet}(y_t - z_{t\bullet}'\delta) \rightarrow E[g_t(\delta)] = E[x_{t\bullet}(y_t - z_{t\bullet}'\delta)] = E(x_{t\bullet}\varepsilon_t) = 0.$

- Sample mean of moment function:

  $g_T(\delta) = T^{-1}\Sigma_t x_{t\bullet}(y_t - z_{t\bullet}'\delta) = T^{-1}X'(y-Z\delta).$

  - Gradient of $g_T(\delta)$:

    $G_T(\delta) = -T^{-1}\Sigma_t x_{t\bullet}z_{t\bullet}' = -T^{-1}X'Z$;

  - $Cov\left(\sqrt{T}g_T(\delta)\right) = V$.

    - If the ε_t are serially uncorrelated,

      $$\hat{V} = \frac{1}{T}\Sigma_{t=1}^T g_t(\hat{\delta})g_t(\hat{\delta})' = \frac{1}{T}\Sigma_{t=1}^T \left(y_t - z_{t\bullet}'\hat{\delta}\right)^2 x_{t\bullet}x_{t\bullet}'$$

      [See White (1982, ECON).]

      If not, use Newey-West, Andrews, etc.

  - $\tilde{\delta}$ minimizes

    $$Tg_T(\delta)\hat{V}^{-1}g_T(\delta) = (y - Z\delta)'X\left[T\hat{V}\right]^{-1}X'(y - Z\delta).$$

  - $\tilde{\delta} = \left[Z'X\left(T\hat{V}\right)^{-1}X'Z\right]^{-1}Z'X(T\hat{V})^{-1}X'y$ [called 2SIV estimator].

  - $Cov(\tilde{\delta}) = \left[Z'X\left(T\hat{V}\right)^{-1}X'Z\right]^{-1}$.

  - If ε_t are i.i.d., then 2SIV $\Rightarrow$ 2SLS.

    If ε_t are not i.i.d., 2SIV is more efficient than 2SLS.

  - 2SLS is a GMM estimator using $A = s_\varepsilon^2(X'X/T)$, which is suboptimal unless the ε_t are i.i.d.

(3) Specification Test

- If q > p, $J_T = \left(y - Z\tilde{\delta}\right)' X \left[T\hat{V}\right]^{-1} X'\left(y - Z\tilde{\delta}\right) \Rightarrow \chi^2$(df = q-p).
- Testing exogeneity of a variable.
  - Wish to test whether a variable in $z_{t\bullet}$, say h_t, is endogenous or not. Use the EHS test method.

STEP 1: Estimate $y_t = z_{t\bullet}'\delta + \varepsilon_t$ by 2SIV with IV = $x_{t\bullet}$, and get J_T [Hansen statistic].

STEP 2: Estimate the model by 2SIV with IV = $(x_{t\bullet}', h_t)'$, and get J_T^* [Hansen statistic].

STEP 3: $D_T = J_T^* - J_T \Rightarrow \chi^2(1)$.

2. Estimation of Multiple Equations Model

- Two equations models:

$$y_{t1} = z_{t1,\bullet}'\delta_1 + \varepsilon_{t1};$$
$$y_{t2} = z_{t2,\bullet}'\delta_2 + \varepsilon_{t2}.$$

Here, the $z_{tj,\bullet}$ (j = 1, 2) are $p_j \times 1$ vectors of regressors. Here, $z_{t1,\bullet}$ and $z_{t2,\bullet}$ could be correlated with ε_{t1} and ε_{t2}.

- Assumptions:
  - $x_{tj,\cdot}$ is $q_j \times 1$ (j = 1,2, $q_j \geq p_j$) such that $E(x_{tj,\cdot}\,\varepsilon_{tj}) = 0$.
  - $Cov(\varepsilon_{t1},\varepsilon_{t2}) \neq 0$ [If = 0, use 2SIV.]
  - Allow ε_{t1} and ε_{t2} to be heteroskedastic: $var(\varepsilon_{t1})$, $cov(\varepsilon_{t1},\varepsilon_{t2})$ and $var(\varepsilon_{t2})$ vary across different t.

- Estimation

  1) 2SIV: Consistent, but not efficient unless $cov(\varepsilon_{t1},\varepsilon_{t2}) = 0$, or $q_1 = p_1$ and $q_2 = p_2$.

  2) 3SIV: Optimal GMM estimator: $\delta_*^* = \begin{pmatrix} \delta_1^* \\ \delta_2^* \end{pmatrix}$.

  [Chamberlain (1984, Handbook of Econometrics, Ch. 22).]

- 3SIV

  - Define $\delta_* = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$; $y_1 = \begin{pmatrix} y_{t1} \\ y_{t2} \\ : \\ y_{T1} \end{pmatrix}$

    and y_2, X_1, X_2, Z_1 and Z_2 are similarly defined.

  - Define:

$$y_* = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}; Z_* = \begin{pmatrix} Z_1 & 0_{T\times p_2} \\ 0_{T\times p_1} & Z_2 \end{pmatrix}; X_* = \begin{pmatrix} X_1 & 0_{T\times q_2} \\ 0_{T\times q_1} & X_2 \end{pmatrix}.$$

- Moment function:

$$g_t(\delta_*) = \begin{pmatrix} x_{t1,\bullet}(y_{t1} - z_{t1,\bullet}{}'\delta_1) \\ x_{t2,\bullet}(y_{t2} - z_{t2,\bullet}{}'\delta_2) \end{pmatrix}.$$

[Apparently, $E[g_t(\delta_*)] = 0$.]

- Sample mean of the moment function:

$$g_T(\delta_*) = \frac{1}{T}\Sigma_{t=1}^{T} \begin{pmatrix} x_{t1,\bullet}\left(y_{t1} - z_{t1,\bullet}{}'\delta_1\right) \\ x_{t2,\bullet}\left(y_{t2} - z_{t2,\bullet}{}'\delta_2\right) \end{pmatrix}$$

$$= \frac{1}{T}\begin{pmatrix} X_1{}'(y_1 - Z_1\delta_1) \\ X_2{}'(y_2 - Z_2\delta_2) \end{pmatrix} = \frac{1}{T}X_*{}'(y_* - Z_*\delta_*).$$

- $G_T = -\frac{1}{T}X_*{}'Z_* = -\frac{1}{T}\begin{pmatrix} X_1{}'Z_1 \\ X_2{}'Z_2 \end{pmatrix}$.

- $\hat{V}$ is the estimate of V using any consistent estimator of δ_*.

- 3SIV minimizes:

$$\left(y_* - Z_*\delta_*\right)'\left(T\hat{V}\right)^{-1}\left(y_* - Z_*\delta_*\right).$$

- From the FOC,

$$\delta_*^* = \begin{pmatrix} \delta_1^* \\ \delta_2^* \end{pmatrix} = \left[Z_*{}'X_*\left(T\hat{V}\right)^{-1}X_*{}'Z_*\right]^{-1}Z_*{}'X_*\left(T\hat{V}\right)^{-1}X_*{}'y_*.$$

- $Cov\left(\delta_*^*\right) = \left[Z_*{}'X_*\left(T\hat{V}\right)^{-1}X_*{}'Z_*\right]^{-1}$.

- Specification test

$$J_T = \left(y_* - Z_*\delta_*\right)' X_*\left(T\hat{V}\right)^{-1} X_*'\left(y_* - Z_*\delta_*\right) \Rightarrow \chi^2(df = q_1 + q_2 - p_1 - p_2),$$

under H_o: $E(g_t(\delta_*)) = 0$.

[8] ISSUES ON GMM

(1) References:

Journal of Business & Economic Statistics, Volumn 14, 1996: Special section on small-sample properties of GMM.

(2) Practical Problems in GMM applications:

1) Hansen tests and t-tests reject null hypotheses too often even if T is reasonably large.

- Even severe for time-series data. The Newey-West or Andrews methods are not quite promising. Sometimes, better statistical inferences are obtained from estimation ignoring aucorrelation.
- Use bootstrap methods.

2) GMM estimates for nonlinear models are quite sensitive to intial starting values. (You have to try many different starting values!!!)

3) How many moment conditions should be used?

- Often too many moment conditions are available.
- Using all the moment conditions is impractical. Furthermore, GMM utilizing too many moment conditions could be seriously biased when T is small.
- Koenker and Machado's (1999, JEC) condition:
  - $(\# \text{ of MC})^3/T \to 0$ as $T \to \infty$ (maybe too strong).
  - Otherwise, GMM could be biased (not necessarily).
- It is not necessarily better to use smaller numbers of moment conditions. Andersen and Sørensen (1996) find that GMM using too few moment conditions is often as bad as GMM using too many moment conditions. There is a trade-off between informational gain and finite-sample bias caused by using more instruments.

4) What are the optimal moment conditions to use?

- In principle, we can find optimal moment conditions.
- Chamberlain (1987)
  - $f(y_t, z_t, x_t, \theta) = \varepsilon_t$, where y_t and z_t are endogenous variables and x_t is an exogenous variable (i.e., $E(x_t\varepsilon_t) = 0$), and ε_t is iid with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$.

    → EX: $f(y_t,x_t,\theta) = y_t - z_t'\theta$.
  - Let $m_t = E\left(\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} \mid x_t\right)$.
  - Then, the optimal moment conditions are $E[m_t\varepsilon_t] = 0$.
  - But it is not easy to compute m_t. Sometimes, impossible.
  - Newey (1990) develops an nonparametric method to estimate m_t.
  - But GMM with the estimated optimal instrumental variables m_t may have poor finite-sample properties.

[9] Empirical Example

$$\begin{aligned}\text{LRATE} = {} & \gamma_{21}\text{LHWORK} + \beta_{11} + \beta_{21}\text{ED} + \beta_{31}\text{ED}^2 + \beta_{41}\text{EXPP} \\ & + \beta_{51}\text{EXPP}^2 + \beta_{61}\text{AGE} + \beta_{71}\text{AGE}^2 + \beta_{81}\text{OCCW} \\ & + \beta_{9,1}\text{OCCB} + \beta_{10,1}\text{UNEMPR} + \beta_{11,1}\text{REGS} \\ & + \beta_{12,1}\text{MINOR} + \beta_{13,1}\text{INDUMG} + \beta_{14,1}\text{UNION} \\ & + \beta_{15,1}\text{URB} + \varepsilon_1;\end{aligned}$$

$$\begin{aligned}\text{LHWORK} = {} & \gamma_{1,2}\text{LRATE} + \beta_{1,2} + \beta_{2,2}\text{ED} + \beta_{3,2}\text{ED}^2 \\ & + \beta_{4,2}\text{AGE} + \beta_{5,2}\text{AGE}^2 + \beta_{6,2}\text{REGS} + \beta_{7,2}\text{MINOR} \\ & + \beta_{8,2}\text{URB} + \beta_{9,2}\text{KIDS5} + \beta_{10,2}\text{LOFINC} + \varepsilon_2.\end{aligned}$$

Endogenous: LRATE, LHWORK.

Exogeous: C, ED, ED^2, <u>EXPP, $EXPP^2$</u>, AGE, AGE^2, <u>OCCW, OCCB,</u>
<u>UNEMPR</u>, REGS, MINOR, <u>INDUMG, UNION</u>, URB, **KIDS5**,
LOFINC.

GMM (or 2SLS) for an Single Equation

STEP 1: Push **Objects/New Object**.

STEP 2: Choose **Equation**. And push **OK** button. Then, you are in **Equation Specification** box.

STEP 3: Choose **GMM**.

STEP 4: Choose any options you like for weighting matrix.

STEP 5: In **Equation Specification** box, type

LRATE C LHWORK

or LRATE=C(1)+C(2)*LHWORK

or LRATE=C(1)+C(2)*(EXPP^C(3)+TENURE^C(4))

STEP 6: In **Instrument List** box, type

URB TENURE KIDS5

STEP 7: Click OK.

NOTE: Whether you include C in the instrument set or not, EVIEWS will include C as an instruments.

NOTE: The reported J-statistic is equal to the Hansen statistic divided by # of observations.

GMM (or 3SLS) for Multiple Equations

STEP 1: Push **Objects/New Object**.

STEP 2: Choose **System**. Push OK button. Then, you are in **System Specification** box.

STEP 3: Type:

LRATE=C(1) + C(2)*LHWORK + C(3)*TENURE
LHWORK = C(4) + C(5)*LRATE + C(6)*KIDS5
INST C TENURE KIDS5 URB (if instruments are the same for all equations)

or LRATE=C(1) + C(2)*LHWORK + C(3)*TENURE @ TENURE KIDS5 URB
LHWORK = C(4) + C(5)*LRATE + C(6)*KIDS5 @ TENURE KIDS5 UNEMPR

STEP 4: Click **Estimate** button.

STEP 5: Choose any option you like for estimation method and weighting matrix.

- Iterative Weights and Coeffs/Simultaneous Ψ Iterative GMM
- Interative Weighs and Coeffs/Sequential Ψ Iterative GMM
- One-Step Weighting Matrix/Iterative Coeffs Ψ Two-step GMM
- One-Step Weighting Matrix/One-Step Coeffs
  Ψ Linearlized GMM of Newey (1985, Journal of Econometrics)

STEP 6: Click OK.

NOTE: Whether you include C in the instrument set or not, EVIEWS will include C as an instruments.

NOTE: The reported J-statistic is equal to the Hansen statistic divided by # of observations.

<2SIV Estimation of the LRATE equation>

Dependent Variable: LRATE
Method: Generalized Method of Moments
Sample: 1 923
Included observations: 923
White Covariance
Convergence achieved after: 9 weight matrices, 10 total coef iterations
Instrument list: C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB
UNEMPR REGS MINOR INDUMG UNION URB KIDS5 LOFINC

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LHWORK	**0.295959**	**0.191838**	**1.542755**	**0.1232**
C	0.850843	0.403910	2.106514	0.0354
ED	-0.120263	0.042498	-2.829831	0.0048
ED^2	0.007225	0.001731	4.173371	0.0000
EXPP	0.027302	0.008623	3.166101	0.0016
EXPP^2	-0.000499	0.000256	-1.950323	0.0514
AGE	-0.004613	0.015854	-0.290992	0.7711
AGE^2	3.36E-06	0.000182	0.018461	0.9853
OCCW	0.172470	0.040560	4.252188	0.0000
OCCB	-0.008701	0.047702	-0.182399	0.8553
UNEMPR	-0.189409	0.552475	-0.342838	0.7318
REGS	-0.031034	0.028483	-1.089545	0.2762
MINOR	-0.107482	0.031502	-3.411928	0.0007
INDUMG	0.148893	0.033342	4.465580	0.0000
UNION	0.149004	0.033544	4.442079	0.0000
URB	0.187054	0.028301	6.609456	0.0000

R-squared	0.135695	Mean dependent var	1.662759
Adjusted R-squared	0.121401	S.D. dependent var	0.399943
S.E. of regression	0.374881	Sum squared resid	127.4659
Durbin-Watson stat	1.754334	**J-statistic**	**0.037998**

$J_T = 0.038*923 = 35 > 3.84$ (= c from $\chi^2(1)$)

The model is rejected.

Alternative Model?

$$\begin{aligned}
\text{LRATE} = {} & \gamma_{2,1}\text{LHWORK} + \beta_{1,1} + \beta_{2,1}\text{ED} + \beta_{3,1}\text{ED}^2 + \beta_{4,1}\text{EXPP} \\
& + \beta_{5,1}\text{EXPP}^2 + \beta_{6,1}\text{AGE} + \beta_{7,1}\text{AGE}^2 + \beta_{8,1}\text{OCCW} \\
& + \beta_{9,1}\text{OCCB} + \beta_{10,1}\text{UNEMPR} + \beta_{11,1}\text{REGS} \\
& + \beta_{12,1}\text{MINOR} + \beta_{13,1}\text{INDUMG} + \beta_{14,1}\text{UNION} \\
& + \beta_{15,1}\text{URB} + \beta_{16,1}\text{LOFINC} + \varepsilon_1; \\
\text{LHWORK} = {} & \gamma_{1,2}\text{LRATE} + \beta_{1,2} + \beta_{2,2}\text{ED} + \beta_{3,2}\text{ED}^2 + \beta_{4,2}\text{AGE} \\
& + \beta_{5,2}\text{AGE}^2 + \beta_{6,2}\text{REGS} + \beta_{7,2}\text{MINOR} + \beta_{8,2}\text{URB} \\
& + \beta_{9,2}\text{KIDS5} + \beta_{10,2}\text{LOFINC} + \varepsilon_2.
\end{aligned}$$

Endogenous: LRATE, LHWORK.

Exogeous: C, ED, ED^2, <u>EXPP, $EXPP^2$</u>, AGE, AGE^2, <u>OCCW, OCCB</u>, <u>UNEMPR</u>, REGS, MINOR, INDUMG, UNION, URB, **KIDS5**, LOFINC.

The LRATE equation is exactly identified. So, can't test for the specification of it.

<2SIV Estimation of the LRATE Equation with LOFINC>

Dependent Variable: LRATE
Method: Generalized Method of Moments
Sample: 1 923
Included observations: 923
White Covariance
Convergence achieved after: 1 weight matrix, 2 total coef iterations
Instrument list: C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB UNEMPR REGS MINOR INDUMG UNION URB KIDS5 LOFINC

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LHWORK	0.109040	0.150765	0.723244	0.4697
C	0.212160	0.389510	0.544684	0.5861
ED	-0.085310	0.034705	-2.458162	0.0142
ED^2	0.005798	0.001429	4.058488	0.0001
EXPP	0.026180	0.007239	3.616686	0.0003
EXPP^2	-0.000413	0.000215	-1.922814	0.0548
AGE	-0.008680	0.013081	-0.663557	0.5071
AGE^2	4.30E-05	0.000151	0.284272	0.7763
OCCW	0.143353	0.034614	4.141442	0.0000
OCCB	0.028543	0.040881	0.698211	0.4852
UNEMPR	-0.484955	0.482249	-1.005611	0.3149
REGS	-0.024463	0.024704	-0.990254	0.3223
MINOR	-0.069270	0.027469	-2.521789	0.0118
INDUMG	0.139793	0.028809	4.852432	0.0000
UNION	0.142994	0.028516	5.014581	0.0000
URB	0.145258	0.025225	5.758555	0.0000
LOFINC	0.114251	0.025136	4.545303	0.0000

R-squared	0.372912	Mean dependent var	1.662759
Adjusted R-squared	0.361837	S.D. dependent var	0.399943
S.E. of regression	0.319495	Sum squared resid	92.48170
Durbin-Watson stat	1.821699	**J-statistic**	**1.15E-29**

<2SIV Estimation of the LHWORK Equation>

Dependent Variable: LHWORK
Method: Generalized Method of Moments
Sample: 1 923
Included observations: 923
White Covariance
Convergence achieved after: 5 weight matrices, 6 total coef iterations
Instrument list: C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB
UNEMPR REGS MINOR INDUMG UNION URB KIDS5 LOFINC

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LRATE	-0.393270	0.141694	-2.775488	0.0056
C	0.974445	0.454783	2.142659	0.0324
ED	0.091067	0.048359	1.883132	0.0600
ED^2	-0.003195	0.002062	-1.549342	0.1216
AGE	0.059209	0.014035	4.218594	0.0000
AGE^2	-0.000634	0.000173	-3.660585	0.0003
REGS	-0.060450	0.042680	-1.416342	0.1570
MINOR	-0.059823	0.049757	-1.202295	0.2296
URB	0.043865	0.051322	0.854708	0.3929
KIDS5	0.126917	0.041966	3.024273	0.0026
LOFINC	0.069098	0.032668	2.115145	0.0347

R-squared	0.075309	Mean dependent var	2.909292
Adjusted R-squared	0.065170	S.D. dependent var	0.574604
S.E. of regression	0.555565	Sum squared resid	281.4911
Durbin-Watson stat	1.625299	**J-statistic**	**0.008211**

J_T = 0.008211*923 = 7.5788 < 12.59 (c from $\chi^2(6)$) at 95%.

The model is not rejected.

<Testing Exogeneity of LRATE in the LHWORK Equation>

Dependent Variable: LHWORK
Method: Generalized Method of Moments
Sample: 1 923
Included observations: 923
White Covariance
Convergence achieved after: 5 weight matrices, 6 total coef iterations
Instrument list: C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB
UNEMPR REGS MINOR INDUMG UNION URB KIDS5 LOFINC
LRATE

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LRATE	-0.303304	0.053650	-5.653329	0.0000
C	0.934103	0.446380	2.092616	0.0367
ED	0.099468	0.047055	2.113855	0.0348
ED^2	-0.003784	0.001886	-2.005599	0.0452
AGE	0.057021	0.013637	4.181360	0.0000
AGE^2	-0.000607	0.000168	-3.608604	0.0003
REGS	-0.054201	0.041482	-1.306602	0.1917
MINOR	-0.049562	0.047654	-1.040042	0.2986
URB	0.027032	0.044531	0.607028	0.5440
KIDS5	0.127720	0.042159	3.029445	0.0025
LOFINC	0.061666	0.030298	2.035310	0.0421

R-squared	0.078360	Mean dependent var	2.909292
Adjusted R-squared	0.068254	S.D. dependent var	0.574604
S.E. of regression	0.554648	Sum squared resid	280.5624
Durbin-Watson stat	1.619627	**J-statistic**	**0.008704**

$J_T^* = 0.008704*923 = 8.0337$

$D_T = J_T^* - J_T = 8.0337 - 7.5788 = 0.4549 < 3.84$ (=c from $\chi^2(1)$)

at 95%.

Can't reject H_o: LRATE is exogenous.

[10] Revisiting the Hausman and Taylor Model

(1) Model:

$y_i = H_i\delta + u_i = X_i\beta + e_T z_i\gamma + u_i$; $u_i = e_T\alpha_i + \varepsilon_i$,

where,

$X_i = [X_{1i},X_{2i}]$ and $z_i = [z_{1i},z_{2i}]$;

x_{1it} = $1\times k_1$ of time-varying regressors that are uncorrelated with α_i;

x_{2it} = $1\times k_2$ of time-varying regressors that may be correlated with α_i;

z_{1i} = $1\times g_1$ of time-invariant regressors that are uncorrelated with α_i;

z_{2i} = $1\times g_2$ of time-invariant regressors that may be correlated with α_i;

α_i = iid with $(0,\sigma_\alpha^2)$.

(2) Assumptions

- Basic Assumptions (BA):

  $E[x_{is}'(u_{it}-u_{i,t-1})] = 0$ and $E[z_i'(u_{it}-u_{i,t-1})] = 0$, for any t and s: Regressors are strictly exogenous to ε_{it}. [Note that $\Delta u_{it} \equiv u_{it}-u_{i,t-1} = \varepsilon_{it} - \varepsilon_{i,t-1} \equiv \Delta\varepsilon_{it}$.]

  $\rightarrow$ $\Delta y_{it} = \Delta x_{it}\beta + \Delta\varepsilon_{it}$

$$\rightarrow \Delta y_i \equiv \begin{pmatrix} \Delta x_{i2} \\ \Delta x_{i3} \\ \vdots \\ \Delta x_{iT} \end{pmatrix}\beta + \begin{pmatrix} \Delta \varepsilon_{i2} \\ \Delta \varepsilon_{i3} \\ \vdots \\ \Delta \varepsilon_{iT} \end{pmatrix} \equiv \Delta X_i\beta + \Delta\varepsilon_i .$$

$\rightarrow F_i = (x_{i1}, x_{i2}, \ldots, x_{iT}, z_i)$.

$$\rightarrow W_{B,i} = \begin{pmatrix} F_i & 0 & \ldots & 0 \\ 0 & F_i & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & F_i \end{pmatrix} = I_{T-1} \otimes F_i \; .$$

$\rightarrow E(W_{B,i}'\Delta\varepsilon_i) = E[W_{B,i}'(\Delta y_i - \Delta X_i\beta)] = 0$.

- Hausman and Taylor Assumption (HTA):

$$E(\overline{x}_{1i}'\overline{u}_i) = 0; E(z_{1i}'\overline{u}_i) = 0.$$

$$\rightarrow w_{HT,i} = (\overline{x}_{i1}, z_{1i}).$$

$$\rightarrow E(w_{HT,i}'\overline{u}_i) = E\left(w_{HT,i}'(\overline{y}_i - \overline{x}_i\beta - z_i\gamma)\right) = 0.$$

- Amemiya and MaCurdy Assumption (AMA):

$E(x_{1it}'\overline{u}_i) = 0; E(z_{1i}'\overline{u}_i) = 0$, for any t.

$\rightarrow w_{AM,i} = [x_{1i1}, x_{1i2}, \ldots, x_{1iT}, z_i]$.

$$\rightarrow E\left(w_{AM,i}'\overline{u}_i\right) = E\left(w_{AM,i}'(\overline{y}_i - \overline{x}_i\beta - z_i\gamma)\right) = 0.$$

CASE A: The ε_{it} are iid over i and t

1) Hausman and Taylor (1981):

- Define $G_{HT,i} = [Q_T X_i, P_T X_{1i}, e_T z_{1i}]$, $G_{HT} = [Q_V X, P_V X_1, VZ_1]$.
- Then, under BA and HT,

  $E(X_i' Q_T u_i) = E[X_i' Q_T(e_T \alpha_i + \varepsilon_i)] = E(X_i' Q_T \varepsilon_i) = 0_{k\times 1}$.

  $E(X_{1i}' P_T u_i) = 0_{k_1\times 1}$ ($\rightarrow E(\bar{x}_{1i}' \bar{u}_i) = 0_{k_1\times 1}$).

  $E(z_{1i}' e_T' u_i) = 0_{g_1\times 1}$ ($\rightarrow E(z_{1i}' \bar{u}_i) = 0_{g_1\times 1}$).

- HT estimator is an 2SLS estimator based on $E(G_{HT,i}' u_i) = 0_{(k+k_1+g_1)\times 1}$.
  - Observe that

$$E(X_i' Q_T \Sigma^{-1/2} u_i) = E(X_i' Q_T (\theta P_T + Q_T) u_i)$$
$$= E(X_i' Q_T u_i) = 0_{k\times 1}.$$

  - Similarly, you can show:

$$E(X_{1i}' P_T \Sigma^{-1/2} u_i) = 0_{k_1\times 1};$$

$$E(z_{1i}' e_T' \Sigma^{-1/2} u_i) = 0_{g_1\times 1}.$$

  - That is, $E(G_{HT}' \Sigma^{-1/2} u_i) = 0_{(k+k_1+g_1)\times 1}$.

- HT offers an estimation procedure under their assumptions.
  - Reconsider the whitened equation:

$$\Sigma^{-1/2} y_i = \Sigma^{-1/2} H_i \delta + \Sigma^{-1/2} u_i$$
$$\Omega^{-1/2} y = \Omega^{-1/2} H \delta + \Omega^{-1/2} u$$

  - HT estimate δ by 2SLS using G_{HT} as IV. [$\Sigma^{-1/2} = Q_T + \theta P_T$].

2) AM estimation:

- Define $G_{AM,i} = [Q_T X_i, e_T x_{AM,1i}, e_T z_{1i}]$, $G_{AM} = [Q_V X, VX_{AM,1}, VZ_1]$,

  where $x_{AM,1i} = [x_{1i1}, \ldots, x_{1iT}]$.

- Under the AM assumptions,

  $E(X_i' Q_T u_i) = E[X_i' Q_T (e_T \alpha_i + \varepsilon_i)] = E(X_i' Q_T \varepsilon_i) = 0.$

  $E(x_{AM,1i}' e_T' u_i) = 0.$

  $E(z_{1i}' e_T' u_i) = 0.$

  $\rightarrow E(G_{AM,i}' u_i) = 0$ and $E(G_{AM,i}' \Sigma^{-1/2} u_i) = 0.$

CASE B: The ε_{it} are not iid over time, but iid over i

Digression to 2SLS, 3SLS and GMM estimators for panel data

- Notation:
  - For T×p matrix M_i or T×1 vector m_i,

$$\rightarrow \quad M = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix}; m = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_N \end{pmatrix}.$$

    $\rightarrow$ H denotes the data matrices of NT rows.

  - With this notation, $y = H\delta + u$.
  - W_i is a T×q matrix of instruments such that $E(W_i' u_i) = 0$.

- Estimators:

  Assume that there is no cross-sectional correlations among the u_i.

  1) 2SLS: $\hat{\delta}_{2SLS}(W_i) = [H'W(W'W)^{-1}W'H]^{-1}H'W(W'W)^{-1}W'y$.

  2) 3SLS:

  $$\hat{\Sigma} = \frac{1}{N}\Sigma_i(y_i - H_i\hat{\delta}_{2SLS})(y_i - H_i\hat{\delta}_{2SLS})'; \ \hat{\Omega} = I_N \otimes \hat{\Sigma};$$

  $$\hat{\delta}_{3SLS}(W_i) = [H'W(W'\hat{\Omega}W)^{-1}W'H]^{-1}H'W(W'\hat{\Omega}W)^{-1}W'y.$$

  $$Cov\left(\hat{\delta}_{3SLS}(W_i)\right) = \left(H'W(W'\hat{\Omega}W)^{-1}W'H\right)^{-1}.$$

  3) Two-Step GMM Estimator:

  $$\hat{\delta}_{GMM}(W_i) = [H'W(N\hat{V})^{-1}W'H]^{-1}H'W(N\hat{V})^{-1}W'y,$$

  where $\hat{V} = \frac{1}{N}\Sigma_i W_i'\hat{u}_i\hat{u}_i'W_i$ and $\hat{u}_i = y_i - H_i\hat{\delta}_{2SLS}$.

  - J-stat = $(W'u_G)'(N\hat{V})^{-1}(W'u_G)$,

    where u_G = residual from GMM estimation = $y - H\hat{\delta}_{GMM}$.

- Facts:
  - If $E(u_iu_i'|W_i)$ is a constant T×T matrix (cross-sectional homoskedasticity): 3SLS and GMM estimators are asymptotically identical (when same instruments W_i are used).
  - If not, GMM estimator is strictly more efficient.

End of Digression

(1) Efficient Estimator (Arellano and Bover, 1995, JEC)

- ABo estimator

  = GMM based on (BA) and (HTA) or (AMA)

  = GMM based on:

$$E\left(W_{B,i}{}'(\Delta y_i - \Delta X_i\beta)\right) = 0 \text{ plus,}$$

$$E\left(w_{HT,i}{}'(\overline{y}_i - \overline{x}_i\beta - z_i\gamma)\right) = 0 \text{ or } E\left(w_{AM,i}{}'(\overline{y}_i - \overline{x}_i\beta - z_i\gamma)\right) = 0.$$

- HT, AM or BMS estimators are asymptotically equivalent to the ABo estimator only if the ε_{it} are iid.

(2) Modified Generalized Instrumental Variables (MGIV) Estimator (Im, Ahn, Schmidt and Wooldridge, 1999; Ahn and Schmidt, 1999)

- $W_{HT,i} = [Q_T X_i, P_T X_{1i}, e_T z_{1i}]$;
- $W_{AM,i} = [Q_T X_i, e_T x_{AM,1i}, e_T z_{1i}]$, $x_{AM,1i} = [x_{1i1}, \ldots, x_{1iT}]$ $(1\times k_1 T)$.
- Assume that the ε_i are homoskedastic across i but heteroskedastic or autocorrelated over t.

• MGIV estimator:

$$\tilde{\delta}_{MGIV,HT} \equiv \hat{\delta}_{3SLS}([Q_\Sigma Q_T X_i,\ \Sigma^{-1}P_T X_{1i}, \Sigma^{-1}e_T z_{1i}])$$

$$\tilde{\delta}_{MGIV,AM} \equiv \hat{\delta}_{3SLS}([Q_\Sigma Q_T X_i,\ \Sigma^{-1}e_T x_{AM,1i}, \Sigma^{-1}e_T z_{1i}])$$

$$\tilde{\beta}_{MGIV,KR} \equiv \hat{\beta}_{3SLS}(Q_\Sigma Q_T X_i)$$

= Kiefer's GLS for FE model (1980, JEC)

[Here, $Q_\Sigma = \Sigma^{-1} - \Sigma^{-1}e_T(e_T'\Sigma^{-1}e_T)^{-1}e_T'\Sigma^{-1}$.]

- If the u_i are cross-sectionally iid (that is, $Cov(u_i) = \Sigma$ for all i), then, the HT (AM) MGIV estimator is as efficient as the Abo-HT (Abo-AM) estimator under the HT (AM) assumptions.
- When the ε_i are cross-sectionally heteroskedastic:
  - ABo GMM is strictly more efficient asymptotically.
  - But MGIV with Hetero-corrected covariance matrices has better finite sample properties.

(3) GMM with HT or AM instruments

- ABo GMM are strictly more efficient asymptotically.
- But GMM with HT and AM instruments are better when sample is small: $E\left(G_{HT,i}'(y_i - X_i\beta - e_T z_i\gamma)\right) = 0$ or

$$E\left(G_{AM,i}'(y_i - X_i\beta - e_T z_i)\right) = 0.$$

[11] Dynamic Panel Data Models

(1) General Model

$$y_{it} = \lambda y_{i,t-1} + x_{it}\beta + u_{it} , u_{it} = \alpha_i + \varepsilon_{it} . \qquad (1)$$

Assume y_{i0} is available.

- Motivation for dynamic model:

  Short-run effect of x_{it}: β.

  Long-run effect of x_{it}: $\beta/(1-\lambda)$.

- Example:

  y_{it}^*: desired level of inventory, $y_{it}^* = x_{it}\gamma + v_{it}$.

  y_{it}: actual level of inventory, $y_{it} = (1-\lambda)y_{it}^* + \lambda y_{i,t-1}$

  $\rightarrow y_{it} = \lambda y_{i,t-1} + x_{it}[(1-\lambda)\gamma] + (1-\lambda)v_{it} = \lambda y_{i,t-1} + x_{it}\beta + u_{it}$

(2) Standard Assumptions:

(A.1) For all i, ε_{it} is uncorrelated with y_{i0} for all t.

(A.2) For all i, ε_{it} is uncorrelated with α_i for all t.

(A.3) For all i, the ε_{it} are mutually uncorrelated.

(A.4) For all i, var(ε_{it}) is the same for all t. (optional)

(3) Facts:

- Within estimator is inconsistent unless T is large.

Why?

- $$y_{it} - \overline{y}_i = (y_{i,t-1} - \overline{y}_{i,-1})\lambda + (x_{it} - \overline{x}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}_i),$$

  where $\overline{y}_i = \frac{1}{T}(y_{i,1} + y_{i,2} + ... + y_{i,T})$,

  $$\overline{y}_{i,-1} = \frac{1}{T}(y_{i,0} + y_{i,1} + ... + y_{i,T-1}),$$

  $$\overline{\varepsilon}_i = \frac{1}{T}(\varepsilon_{i1} + \varepsilon_{i2} + ... + \varepsilon_{iT}).$$

- Note that $(y_{i,t-1} - \overline{y}_{i,-1})$ and $(\varepsilon_{it} - \overline{\varepsilon}_i)$ are correlated. This correlation becomes negligible only if T is large.

(4) Notation

- $\Delta y_{it} = y_{it} - y_{i,t-1}$ (called first-difference).
- $\Delta u_{it} = u_{it} - u_{i,t-1} = \varepsilon_{it} - \varepsilon_{i,t-1} = \Delta\varepsilon_{it}$.

(5) Hsiao-Anderson estimation (1981, JASA, HA)

$$\Delta y_{it} = \Delta y_{i,t-1}\lambda + \Delta x_{it}\beta + \Delta\varepsilon_{it} = \Delta h_{it}\delta + \Delta\varepsilon_{it}; \qquad (3)$$

$$\Delta y_i = \Delta y_{i,-1}\lambda + \Delta X_i\beta + \Delta\varepsilon_i = \Delta H_i\delta + \Delta\varepsilon_i; \qquad (3')$$

$$y_{FD} = \delta y_{FD,-1} + X_{FD}\beta + u_{FD} = \delta H_{FD} + u_{FD} \qquad (3'')$$

where $\Delta y_i = \begin{pmatrix} \Delta y_{i2} \\ : \\ \Delta y_{iT} \end{pmatrix}$; $\Delta y_{i,-1} = \begin{pmatrix} \Delta y_{i,1} \\ : \\ \Delta y_{i,T-1} \end{pmatrix}$; $\Delta \varepsilon_i = \begin{pmatrix} \Delta \varepsilon_{i2} \\ : \\ \Delta \varepsilon_{iT} \end{pmatrix}$;

$$y_{FD} = \begin{pmatrix} \Delta y_1 \\ : \\ \Delta y_N \end{pmatrix};\ y_{FD,-1} = \begin{pmatrix} \Delta y_{1,-1} \\ : \\ \Delta y_{N,-1} \end{pmatrix};\ u_{FD} = \begin{pmatrix} \Delta u_1 \\ : \\ \Delta u_N \end{pmatrix}.$$

- The original model includes T equations. But (3′) has (T-1) equations for each i.
- $y_{i,t-2}$ (or $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$) is uncorrelated with ($\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$) as long as the ε_{it} are serially uncorrelated.
- $w_{HA,it} = [y_{i,t-2}, x_{it}]$ or $[y_{i,t-2}, \Delta x_{it}]$ or $[\Delta y_{i,t-2}, \Delta x_{it}]$;

$$W_{HA,i} = \begin{pmatrix} W_{HA,i2} \\ W_{HA,i3} \\ : \\ W_{HA,iT} \end{pmatrix}.$$

- In fact, $[y_{i,t-2}, \Delta x_{it}]$ is better for finite-sample properties.

  (Arellano, 1988, JEC?).
- HA estimator = 2SLS on (3) using IV = $w_{HA,it}$.

  = 2SLS on (3″) using IV = W_{HA}.

- Comment:
  - HA estimator is a (not necessarily optimal) GMM estimator based on the moment conditions $E(W_{HA,i}'\Delta\varepsilon_i) = 0$:

$$E\left(\Sigma_{t=2}^{T} y_{i,t-2}\Delta\varepsilon_{it}\right) = 0 \text{ and } E\left(\Sigma_{t=2}^{T} \Delta x_{it}'\Delta\varepsilon_{it}\right) = 0.$$

  - Is not efficient. More efficient GMM can be obtained by using the moment conditions $E\left(y_{i,t-2}\Delta\varepsilon_{it}\right) = 0$ and $E(\Delta x_{it}\Delta\varepsilon_{it}) = 0$ for all t = 2, ... , T.

(6) Arellano and Bond (1991, RESTUD, AB)

(Also, Holtz-Eakin, Newey and Rosen, 1988, ECON, HNR)

- $(y_{i,t-2}, y_{i,t-3}, \ldots, y_{i,0})$ is uncorr. with $(\varepsilon_{it}-\varepsilon_{i,t-1})$ under SA.
- Define $W_{AB,i} = \begin{pmatrix} y_{i0} & & & & & & & & & & & \Delta x_{i2} \\ & y_{i0} & y_{i1} & & & & & & & & & \Delta x_{i3} \\ & & & y_{i0} & y_{i1} & y_{i2} & & & & & & \Delta x_{i4} \\ & & & & & & \ddots & & & & & \vdots \\ & & & & & & & y_{i0} & y_{i1} & \cdots & y_{i,T-3} & \Delta x_{iT} \end{pmatrix}.$

- Moment conditions:

$$E[W_{AB,i}'\Delta\varepsilon_i] = 0. \qquad (4)$$

- AB estimator = GMM based on (4)
  - $\Delta y_i = \Delta H_i\delta + \Delta\varepsilon_i = \Delta y_{i,-1}\lambda + \Delta X_i\beta + \Delta\varepsilon_i$

Step 1: Do 2SLS with IV = $W_{AB,i}$ and get $\widehat{\Delta\varepsilon_i}$.

Step 2: $\hat{V} = \frac{1}{N}\Sigma_i W_{AB,i}{}'\widehat{\Delta\varepsilon_i}\widehat{\Delta\varepsilon_i}{}' W_{AB,i}$,

$$g_N(\delta) = \frac{1}{N}\Sigma_i W_{AB,i}{}'(\Delta y_i - \Delta H_i\delta),$$

$$G_N(\delta) = -\frac{1}{N}\Sigma_i W_{AB,i}{}'\Delta H_i.$$

Step 3: minimizing $Ng_N(\delta)'\hat{V}^{-1}g_N(\delta)$.

- AB estimator:

$$\hat{\delta}_{AB} = [H'W_{AB}(N\hat{V})^{-1}W_{AB}{}'H]^{-1}H'W_{AB}(N\hat{V})^{-1}W_{AB}{}'y_{FD},$$

with $Cov(\hat{\delta}_{AB}) = [H'W_{AB}(N\hat{V})^{-1}W_{AB}{}'H]^{-1}$.

- If assumption (A.4) holds, i.e., ε_{it} are iid, $\hat{\delta}_{AB}$ is asymptotically identical to 2SLS with instruments $W_{AM,i}$.

- Comment:
  - AB estimator is a (not necessarily optimal) GMM estimator based on the moment conditions:

    $$E(y_{i,s}\Delta\varepsilon_{it}) = 0 \text{ and } E\left(\Sigma_{t=2}^{T}\Delta x_{it}'\Delta\varepsilon_{it}\right) = 0 \text{ for all t and s < t.}$$

  - Is not efficient. More efficient GMM can be obtained by using the moment conditions:

    $$E(y_{i,s}\Delta\varepsilon_{it}) = 0 \text{ and } E\left(\Delta x_{it}'\Delta\varepsilon_{it}\right) = 0,$$

    for all t = 2, ... , T and s < t.

    If you do, estimation also becomes simplified as we see below.

  - Let $W_{AB^*,i} = \begin{pmatrix} y_{i0} & \Delta x_{i2} & & & & & & & & & & & & & \\ & & y_{i0} & y_{i1} & y_{i2} & \Delta x_{i3} & & & & & & & & & \\ & & & & & & y_{i0} & y_{i1} & y_{i2} & \Delta x_{i4} & & & & & \\ & & & & & & & & & & \ddots & & & & \\ & & & & & & & & & & & y_{i0} & y_{i1} & \cdots & y_{i,T-2} & \Delta x_{iT} \end{pmatrix}.$

  - Then, the simultaneous equations apporach = GMM with IV = $W_{AB^*,i}$.
- This method is not recommendable for data with large T.

(7) Ahn and Schmidt (1995, JEC)

- Intuition:
  - It is a well-known fact that FIML is better than 3SLS if there are some restrictions on the covariance matrix of the errors.
  - SA implies some restrictions on the error structure. AB is akin to 3SLS. Then, we may think of the FIML-type of estimator.

- For simplicity, consider the case without x_{it}.

- AB consider only (T-1) equations:

$$\Delta y_{i2} = \lambda \Delta y_{i1} + \Delta \varepsilon_{i2} \qquad \text{(1st eq.)}$$

$$\Delta y_{i3} = \lambda \Delta y_{i2} + \Delta \varepsilon_{i3} \qquad \text{(2nd eq.)}$$

$$\vdots$$

$$\Delta y_{iT} = \lambda \Delta y_{i,T-1} + \Delta \varepsilon_{iT} \qquad \text{((T-1)'th eq.)}$$

- Include another equation:

$$y_{iT} = \lambda y_{i,T-1} + u_{i,T} \,. \qquad \text{(T'th eq.)}$$

- These T equations are equivalent to the original undifferenced equation in terms of information.

- Observe that $E[u_{iT}\Delta\varepsilon_{it}] = E[(y_{iT}-\lambda y_{i,T-1})(\Delta y_{it}-\lambda\Delta y_{i,t-1})] = 0$, t = 2, ... , T-1. These (T-2) conditions can be imposed on GMM, improving efficiency.
  - These (T-2) conditions can be alternatively expressed as

$$E[u_{i,t-2}\Delta\varepsilon_{it}] = E[(y_{i,t-2}-\lambda y_{i,t-3})(\Delta y_{it}-\lambda\Delta y_{i,t-1})] = 0, \; t = 3, 4, \ldots , T.$$

- AS estimator

$$= \text{GMM based on } E(W_{AB,i}'\Delta\varepsilon_i) = E(W_{AB,i}'(\Delta y_i-\lambda\Delta y_{i,-1})) = 0 \text{ and}$$

$$E(u_{iT}\Delta\varepsilon_{it}) = E((y_{iT}-\lambda y_{i,T-1})(\Delta y_{it}-\lambda\Delta y_{i,t-1}) = 0 \; (1 < t < T).$$

  • FIML on the T equations under SA 1-3.

- If the errors are normal: AS estimator is MLE.

  If the errors are not normal and their distributions are not known: AS estimator is semiparametrically efficient. Cannot find better estimator than AS unless you know the true distribution of the errors.

- Linearized AS estimator (Ahn and Schmidt, '97, JEC).

- Arellano and Bover (1995, JEC) stationarity assumption:
  - $E[\Delta y_{it}u_{iT}] = 0$ for t = 1, ... , T-1 (T-2 moment conditions).

    [Alternatively, $E[\Delta y_{i,t-1}u_{it}] = 0$, for t = 2, ... , 1.]

    [Call them ABo conditions.]
  - $\Delta y_{i2} = \lambda\Delta y_{i1} + \Delta\varepsilon_{i2}$: IV = y_{i0}

    $\Delta y_{i3} = \lambda\Delta y_{i2} + \Delta\varepsilon_{i3}$: IV = y_{i0}, y_{i1}

    :

    $\Delta y_{iT} = \lambda y_{i,T-1} + \Delta\varepsilon_{iT}$: IV = y_{i0}, ... , $y_{i,T-2}$.

    $y_{iT} = \lambda y_{i,T-1} + u_{i,T}$: IV = Δy_{i1}, ... , $\Delta y_{i,T-1}$.
  - Where do these conditions come from?
  - Assume that the dynamics of y_{it} began from the far remote past and that $\lambda < 1$:

$$y_{i,0} = \lambda^s y_{i,-s} + \alpha_i \frac{1-\lambda^s}{1-\lambda} + \Sigma_{j=0}^{s-1}\lambda^j \varepsilon_{i,-j}.$$

  - As s → ∞, $y_{i0} = \dfrac{\alpha_i}{1-\lambda} + v_{i0}$; $v_{i0} = \lim_{s\to\infty} \Sigma_{j=0}^{s-1}\lambda^{s-1}\varepsilon_{i,-j}$. (ST)
  - Here, $v_{i,0}$ is not correlated with α_i, ε_{i1}, ... , ε_{iT}.

- Then, (ST) and the AS conditions imply ABo conditions.
  - Why?
    - $y_{i1} = \lambda y_{i0} + \alpha_i + \varepsilon_{i1}$

      $= \lambda[\alpha_i/(1-\lambda) + v_{i,0}] + \alpha_i + \varepsilon_{i1}$

      $= \alpha_i/(1-\lambda) + (\lambda v_{i0} + \varepsilon_{i1})$.
    - $\Delta y_{i1} = (\lambda-1)v_{i0} + \varepsilon_{i1}$.
    - So, $E[\Delta y_{i1}u_{iT}] = E[\Delta y_{i1}(\alpha_i+\varepsilon_{iT})] = 0$.
    - This condition$\times\lambda$ + ($E[\Delta\varepsilon_{i2}u_{iT}] = 0$)

      $\rightarrow E(\Delta y_{i2}u_{iT}) = 0$.
    - ...

(8) Blundell and Bond (1999, JEC)

- Go back to AB and ABo:

$\Delta y_{i2} = \lambda\Delta y_{i1} + \Delta\varepsilon_{i2}$: IV = y_{i0}

$\Delta y_{i3} = \lambda\Delta y_{i2} + \Delta\varepsilon_{i3}$: IV = y_{i0}, y_{i1}

:

$\Delta y_{iT} = \lambda\Delta y_{i,T-1} + \Delta\varepsilon_{iT}$: IV = $y_{i0}, \ldots, y_{i,T-2}$.

$y_{iT} = \lambda y_{i,T-1} + x_{iT}\beta + u_{i,T}$: IV = $\Delta y_{i1}, \ldots, \Delta y_{i,T-1}$.

- For identification, IV's should be correlated with regressors. But under (ST), as δ becomes closer to one or σ_α^2 gets larger, $cov(\Delta y_{is} y_{it})$ $(t < s-1)$ converges to zero (weak instrumental variable problem). That is, AB estimator would result in an inconsistent estimator when δ is near one and σ_α^2 is quite large.
- BB find that GMM using ABo conditions become extremely useful for such cases.
- (ST) may be a quite strong assumption for practical situation. How can we test the assumption?

(9) Aggregation problems in dynamic panel data models

(Pesaran and Smith, 1995, JEC)

(*) $y_{it} = \lambda_i y_{i,t-1} + \alpha_i + \varepsilon_{it}$,

where T is large. Let $\lambda = E(\lambda_i)$.

- Estimate $y_{it} = \lambda y_{i,t-1} + \alpha_i + \varepsilon_{it}$ by any method discussed above.

  $\rightarrow$ Inconsistent for $E(\lambda_i)$.

- Consistent Estimation: Do OLS on (*) for each i. Then, use the average of the OLS estimates of λ_i as the estimator of $\lambda = E(\lambda_i)$. This method is effective only if T is large.
- Comment:
  - What is $E(\lambda_i)$?