2. IV ESTIMATION AND TWO STAGE LEAST SQUARES

[1] Motivation

(1) Consider a regression model:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \varepsilon_t = x_{t\bullet}'\beta + \varepsilon_t .$$

- In what cases is the OLS estimator of β ($\hat{\beta}$) consistent?

  $\rightarrow E(x_{t\bullet}'\varepsilon_t) = 0$.

- In what cases is this condition violated?

(2) Examples

1) Measurement errors in regressors

- A simple example:
  - True model: $y_t = \beta x_t^* + \varepsilon_t$.
  - But we only observe $x_t = x_t^* + v_t$ (v_t: measurement error).

    [x_t may be a proxy variable for x_t^*.]
  - If we use x_t for x_t^*, $y_t = x_t\beta + [\varepsilon_t - \beta v_t]$ (model we estimate).
  - x_t and $(\varepsilon_t - \beta v_t)$ correlated.
  - Assume that the ε_t are i.i.d. with $N(0,\sigma^2)$; v_t i.i.d. $N(0,\sigma_v^2)$; and ε_t and v_t are stochastically independent.

- $p\lim_{T\to\infty} \hat{\beta} = \dfrac{\beta}{1+\sigma_v^2/a}$,

  where a = plim $T^{-1}\Sigma_t(x_t^*)^2$. (Greene)

- $p\lim_{T\to\infty} |\hat{\beta}| < |\beta|$, if $\sigma_v^2 > 0$.
- $p\lim_{\to\infty} \hat{\beta} = \beta$, only if $\sigma_v^2 = 0$.

2) Omitted Variables

- True model: $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \varepsilon_t$.
- The model you estimate: $y_t = \beta_1 + \beta_2 x_{t2} + v_t$,

  where $v_t = \beta_3 x_{t2} + \varepsilon_t$.

- The OLS estimator of β_2 from this misspecfied model is generally inconsistent. It is consistent only if x_{t2} and x_{t3} are uncorrelated.

3) Simultaneous Equations models

(a) $c_t = \beta_1 + \beta_2 y_t + \varepsilon_t$;

(b) $c_t + i_t = y_t$

→ $y_t = \beta_1 + \beta_2 y_t + \varepsilon_t + i_t$.

→ $y_t = [\beta_1/(1-\beta_2)] + i_t[1/(1-\beta_2)] + \varepsilon_t/(1-\beta_2)$.

→ y_t is correlated with ε_t in (a).

→ OLS inconsistent.

[2] The Method of Instrumental Variables (IV)

(1) Intuition:

- Consider a simple regression model $y_t = \beta x_t + \varepsilon_t$.
- Suppose there is a variable z_t such that (i) $E(z_t x_t) \neq 0$ and (ii) $E(z_t \varepsilon_t) = 0$.
- Assume that $(z_t, x_t, \varepsilon_t)'$ is iid.
- Let $\hat{\beta}_{IV} = \frac{\Sigma_t z_t y_t}{\Sigma z_t x_t}$:

$$\hat{\beta}_{IV} = \frac{\frac{1}{T}\Sigma_t z_t y_t}{\frac{1}{T}\Sigma_t z_t x_t} = \frac{\frac{1}{T}\Sigma_t z_t (x_t \beta + \varepsilon_t)}{\frac{1}{T}\Sigma_t z_t x_t} = \beta + \frac{\frac{1}{T}\Sigma_t z_t \varepsilon_t}{\frac{1}{T}\Sigma_t z_t x_t}$$

$$\to_p \beta + \frac{0}{E(z_t x_t)} = \beta.$$

(2) Formal Assumptions for instrumental variables:

Digression to Weak Ideal Conditions (WIC):

(WIC.1) The conditional mean of y_t (dependent variable) given $\varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}$, $x_{1\bullet}, x_{2\bullet}, ... , x_{t\bullet}$ (vectors of explanatory variables) is linear in $x_{t\bullet}$:

$$y_t = E(y_t \mid x_{1\bullet}, ... x_{t\bullet}, \varepsilon_1, ..., \varepsilon_{t-1}) + \varepsilon_t = x_{t\bullet}'\beta + \varepsilon_t .$$

Comment:

- $\text{cov}(\varepsilon_t,\varepsilon_s)$ (t > s) $= E(\varepsilon_t\varepsilon_s) = E_{\varepsilon_s}[E(\varepsilon_t\varepsilon_s | \varepsilon_s)] = E_{\varepsilon_s}[\varepsilon_s E(\varepsilon_t | \varepsilon_s)]$

  $= E_{\varepsilon_s}(0) = 0$ (LIE: Law of Iterative Expectation).

- $E(x_{s\bullet}\varepsilon_t) = 0_{k\times 1}$ for all $t \geq s$.

(WIC.2) $E(\varepsilon_t | x_{1\bullet}, x_{2\bullet}, ..., x_{t\bullet}, \varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}) = 0$.

Comment:

- In fact, (WIC.1) implies (WIC.2). But we write (WIC.2) as an independent assumption for convenience.
- By LIE, $E(\varepsilon_t) = 0$ for all t.

(WIC.3) The series $\{x_{t\bullet}\}$ are covariance-stationary and ergodic.

- We no longer require the random sample assumption (SIC.3) and the nonstochastic regressor assumption (SIC.4).

(WIC.4) $X'X$ is positive definite and $\text{plim}_{T\to\infty}T^{-1}X'X = \text{plim}_{T\to\infty}T^{-1}\Sigma_t x_{t\bullet}x_{t\bullet}'$ ($\equiv Q_o$) is finite.

Comment:

- In fact, $Q_o = \lim_{T\to\infty}T^{-1}\Sigma_t E(x_{t\bullet}x_{t\bullet}')$ [By GWLLN].
- Rules out perfect multicollinearity among regressors.

(WIC.5) $\text{var}(\varepsilon_t|x_{1\bullet},x_{2\bullet}, ..., x_{t\bullet}, \varepsilon_1, ... , \varepsilon_{t-1}) = \sigma^2$ for all $x_{t\bullet}$. (Homoskedasticity Assumption).

Comment:

By the law of iterative expectation, $\text{var}(\varepsilon_t) = \sigma^2$ for all t.

(WIC.6) $x_{t1} = 1$, for all t = 1, ... , T.

(WIC.7) The error terms ε_t are normally distributed.

End of Digression

- Assumptions for Instrumental Variables:

1) $y_t = x_{t\bullet}'\beta + \varepsilon_t = x_{t1}\beta_1 + ... + x_{tk}\beta_k + \varepsilon_t$,

or, $y = X\beta + \varepsilon$.

2) Exists a set of instrumental variables $z_{t\bullet} = (z_{t1}, ... , z_{tq})'$ such that:

2.1) $E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_1, z_{t\bullet}, ..., z_{1\bullet}) = 0$;

[$E(\varepsilon_t) = 0$, $\text{cov}(\varepsilon_t,\varepsilon_s) = 0$, and $E(z_{t\bullet}\varepsilon_t) = 0$;

$\text{plim}\, T^{-1}Z'\varepsilon = \text{plim}\, T^{-1} \Sigma_t z_{t\bullet}\varepsilon_t = 0$.]

2.2) The $(z_{t\bullet}',x_{t\bullet}')'$ are stationary and ergodic;

[$\text{plim}_{T\to\infty}\frac{1}{T}Z'Z = \text{plim}_{T\to\infty}\frac{1}{T}\Sigma_t z_{t\bullet}z_{t\bullet}' \equiv Q_z$ is finite and positive definite; $\text{plim}_{T\to\infty}\frac{1}{T}\Sigma_t z_{t\bullet}x_{t\bullet}'$ is finite and rank = k.]

2.3) $\text{var}(\varepsilon_t|\varepsilon_{t-1},...,\varepsilon_1,z_{t\bullet},...,z_{1\bullet}) = \sigma^2$ (Homoske. Assum.);

[$\text{var}(\varepsilon_t) = \sigma^2$.]

2.4) $\text{plim}_{T\to\infty}\frac{1}{T}Z'X = \text{plim}_{T\to\infty}\frac{1}{T}\Sigma_t z_{t\bullet}x_{t\bullet}'$ is finite, nonzero, with rank = k.

Comment on 2):

$z_{t\bullet}$ can contain exogenous regressors in $x_{t\bullet}$.

- Suppose $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \varepsilon_t$, where x_{t2} is endogenous and other regressors are exogenous. Suppose that some extra variables $h_{t1}, h_{t2}, ... , h_{tr}$ are correlated with x_{t2} but not correlated with ε_t. Then, $z_{t\bullet}$ can contain $(1,x_{t3},...,x_{tk},h_{t1},...,h_{tr})'$.

Comment on 2.1):

The most important part of this assumption is that $z_{t\bullet}$ and ε_t should be uncorrelated; that is, $E(z_{t\bullet}\varepsilon_t) = 0$.

Comment on 2.4):

- q must be greater than or equal to k.
- Suppose $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \varepsilon_t$, where x_{t2} is endogenous and other regressors are exogenous. Then, there must be (a least one) some extra variable (say, $h_{t1}, ... , h_{tr}$) that are correlated with x_{t2} conditional on $(1,x_{t3},x_{t4},...,x_{tk})'$ and uncorrelated with ε_t.

(3) IV estimator (q = k):

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

Theorem:

$$p\lim_{T\to\infty} \hat{\beta} = \beta ;$$

$$\sqrt{T}\left(\hat{\beta}_{IV} - \beta\right) \to_d N\left(0_{k\times 1}, \sigma^2 T (Z'X)^{-1} Z'Z (X'Z)^{-1}\right);$$

$$\text{plim}_{T\to\infty} s_{IV}^2 = \sigma^2 \text{, where } s_{IV}^2 = \frac{(y - X\hat{\beta}_{IV})'(y - X\hat{\beta}_{IV})}{T-k}.$$

Implication:

- The IV estimator is consistent.
- $\hat{\beta}_{IV} \approx N\left(\beta, \hat{\Omega}\right)$, where $\hat{\Omega} = s_{IV}^2 (Z'X)^{-1} Z'Z (X'Z)^{-1}$.
- We can use t- or Wald statistics to test hypotheses regarding β.

  [e.g., t = $\frac{(R\hat{\beta}_{IV} - r)}{\sqrt{R\hat{\Omega}R'}}$, W_T = $(R\hat{\beta}_{IV} - r)'(R\hat{\Omega}R')^{-1}(R\hat{\beta}_{IV} - r)$.]

Digression to t and Wald test

(D.1) Testing a single restriction on β:

- H_o: $R\beta - r = 0$, where R is $1\times k$ and r is a scalar.

Example: $y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3 + \varepsilon_t$.

- We would like to test H_o: $\beta_3 = 0$.
  - Define R = [0 0 1] and r = 0.
  - Then, $R\beta - r = 0 \rightarrow \beta_3 = 0$.
- H_o: $\beta_2 - \beta_3 = 0$ (or $\beta_2 = \beta_3$).
  - Define R = [0 1 -1] and r = 0.
  - $R\beta - r = 0 \rightarrow \beta_2 - \beta_3 = 0$
- H_o: $2\beta_2 + 3\beta_3 = 3$.
  - R = [0 2 3] and r = 3.
  - $R\beta - r = 0 \rightarrow H_o$.

(D.2) Testing several restrictions

Assume that R is $m\times k$ and r is $m\times 1$ vector, and H_o: $R\beta = r$.

Example:

- A model is given: $y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3 + \varepsilon_t$.
- Wish to test for H_o: $\beta_1 = 0$ and $\beta_2 + \beta_3 = 1$.
- Define:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix};\ r = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Then, $H_o \rightarrow R\beta = r$.

(D.3) Testing Nonlinear restrictions:

General form of hypotheses:

- Let $w(\theta) = [w_1(\theta), w_2(\theta), \ldots , w_m(\theta)]'$, where $w_j(\theta) = w_j(\theta_1, \theta_2, \ldots , \theta_p)$ = a function of $\theta_1, \ldots , \theta_p$.
- H_o: The true θ (θ_o) satisfies the m restrictions, $w(\theta) = 0_{m\times 1}$ ($m \leq p$).

Examples:

1) θ: a scalar

$H_o: \theta_o = 2 \rightarrow H_o: \theta_o - 2 = 0 \rightarrow H_o: w(\theta_o) = 0$, where $w(\theta) = \theta - 2$.

2) $\theta = (\theta_1, \theta_2, \theta_3)'$.

$H_o: \theta_{0,1}^2 = \theta_{o,2} + 2$ and $\theta_{o,3} = \theta_{o,1} + \theta_{o,2}$.

$\rightarrow H_o: \theta_{o,1}^2 - \theta_{o,2} - 2 = 0$ and $\theta_{o,3} - \theta_{o,1} - \theta_{o,2} = 0$.

$$\rightarrow H_o:\ w(\theta) = \begin{pmatrix} w_1(\theta) \\ w_2(\theta) \end{pmatrix} = \begin{pmatrix} \theta_1^2 - \theta_2 - 2 \\ \theta_3 - \theta_1 - \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

3) linear restrictions

$\theta = [\theta_1, \theta_2, \theta_3]'$.

$H_o: \theta_{o,1} = \theta_{o,2} + 2$ and $\theta_{o,3} = \theta_{o,1} + \theta_{o,2}$

$$\rightarrow H_o:\ w(\theta) = \begin{pmatrix} w_1(\theta) \\ w_2(\theta) \end{pmatrix} = \begin{pmatrix} \theta_1 - \theta_2 - 2 \\ \theta_3 - \theta_1 - \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$\rightarrow \text{H}_o\text{: } w(\theta) = \begin{pmatrix} 1 & -1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \end{pmatrix} = R\theta - r \,.$$

Remark:

If all restrictions are linear in θ, H_o takes the following form:

H_o: $R\theta_o - r = 0_{m\times 1}$,

where R and r are known m×p and m×1 matrices, respectively.

Definition:

$$W(\theta) \equiv \frac{\partial w(\theta)}{\partial \theta'} = \begin{pmatrix} \frac{\partial w_1(\theta)}{\partial \theta_1} & \frac{\partial w_1(\theta)}{\partial \theta_2} & \cdots & \frac{\partial w_1(\theta)}{\partial \theta_p} \\ \frac{\partial w_2(\theta)}{\partial \theta_1} & \frac{\partial w_2(\theta)}{\partial \theta_2} & \cdots & \frac{\partial w_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial w_m(\theta)}{\partial \theta_1} & \frac{\partial w_m(\theta)}{\partial \theta_2} & \cdots & \frac{\partial w_m(\theta)}{\partial \theta_p} \end{pmatrix}_{m\times p} .$$

Example: (Nonlinear restrictions)

Let $\theta = [\theta_1,\theta_2,\theta_3]'$.

H_o: $\theta_{o,1}^2 - \theta_{o,2} = 0$ and $\theta_{o,1} - \theta_{o,2} - \theta_{o,3}^2 = 0$.

$$\rightarrow w(\theta) = \begin{pmatrix} \theta_1^2 - \theta_2 \\ \theta_1 - \theta_2 - \theta_3^2 \end{pmatrix};\; W(\theta) = \begin{pmatrix} 2\theta_1 & -1 & 0 \\ 1 & -1 & -2\theta_3 \end{pmatrix}.$$

Example: (Linear restrictions)

$\theta = [\theta_1,\theta_2,\theta_3]'$.

H_o: $\theta_{o,1} = 0$ and $\theta_{o,2} + \theta_{o,3} = 1$.

$$\rightarrow w(\theta) = \begin{pmatrix} \theta_1 \\ \theta_2 + \theta_3 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\rightarrow w(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is of form $w(\theta) = R\theta - r$.

Theorem:

Assume that $\sqrt{T}(\hat{\theta} - \theta_o) \rightarrow_d N\left(0_{k\times 1}, p\lim_{T\to\infty} \hat{\Omega}\right)$. Then, under H_o: $w(\theta_o) = 0_{m\times 1}$,

$$\sqrt{T}\left(w(\hat{\theta}) - w(\theta_o)\right) \rightarrow_d N\left(0_{m\times 1}, W(\theta_o)\Omega W(\theta_o)'\right).$$

<Proof>

Taylor's expansion around β_o:

$$w(\hat{\theta}) = w(\theta_o) + W(\bar{\theta})(\hat{\theta} - \theta_o),$$

where $\bar{\theta}$ is between $\hat{\theta}$ and θ_o. Since $\hat{\theta}$ is consistent, so is $\bar{\theta}$. Thus,

$$\sqrt{T}\left(w(\hat{\theta}) - w(\theta_o)\right) \approx W(\theta_o)\sqrt{T}(\hat{\theta} - \theta_o)$$

$$\rightarrow_d N\left(0_{m\times 1}, p\lim_{T\to\infty} W(\theta_o) T\hat{\Omega} W(\theta_o)'\right).$$

Implication:

$$\left(w(\hat{\theta}) - w(\theta_o)\right) = w(\hat{\theta}) \approx N\left(0_{m\times 1}, W(\hat{\theta})\hat{\Omega}W(\hat{\theta})'\right).$$

Comment:

If m = 1. $t = \dfrac{w(\hat{\theta})}{\sqrt{W(\hat{\theta})\hat{\Omega}^{-1}W(\hat{\theta})'}} \rightarrow_d N(0,1)$.

Theorem:

Under WIC and H_o: $w(\theta_o) = 0$,

$$W_T = w(\hat{\theta})'\left[W(\hat{\theta})\hat{\Omega}W(\hat{\theta})'\right]^{-1} w(\hat{\theta}) \rightarrow_d \chi^2(m).$$

<Proof>

Under H_o: $w(\theta_o) = 0$,

$$w(\hat{\beta}) \approx N\left(0_{p\times 1}, W(\hat{\theta})\hat{\Omega}W(\hat{\theta})'\right).$$

For a normal random vector $h_{m\times 1} \sim N(0_{m\times 1},\Psi_{m\times m})$, $h'\Psi^{-1}h \sim \chi^2(m)$. Thus, we obtain the desired result.

Question: What does "Wald test" mean?

A test based on the unrestricted estimator only.

End of Digression

[Proof of the IV theorem]

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta+\varepsilon) = \beta + (Z'X)^{-1}Z'\varepsilon .$$

$$\rightarrow \hat{\beta}_{IV} = \beta + \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{T}Z'\varepsilon\right)$$

$$\rightarrow\ p\lim_{T\to\infty}\hat{\beta}_{IV} = \beta + \left(p\lim_{T\to\infty}\frac{1}{T}Z'X\right)^{-1}\left(p\lim_{T\to\infty}\frac{1}{T}Z'\varepsilon\right).$$

$\rightarrow$ By the central limit theorem and the given assumptions,

$$\frac{1}{\sqrt{T}}Z'\varepsilon \rightarrow_d N\left(0, \sigma^2 p\lim_{T\to\infty}\frac{1}{T}Z'Z\right).$$

$$\rightarrow \begin{array}{l}\left(\frac{1}{T}Z'X\right)^{-1}\frac{1}{\sqrt{T}}Z'\varepsilon \\ \qquad \rightarrow_d N\left(0_{q\times 1}, \sigma^2\left(p\lim\frac{1}{T}Z'X\right)\left(p\lim\frac{1}{T}Z'Z\right)\left(p\lim\frac{1}{T}X'Z\right)^{-1}\right)\end{array}$$

(3) Two-Stage Least Squares (2SLS) (q > k)

$$\hat{\beta}_{2SLS} = \left(X'P(Z)Z\right)^{-1}X'P(Z)y, \text{ where } P(Z) = Z(Z'Z)^{-1}Z'.$$

Note 1:

If q = k, $\hat{\beta}_{IV} = \hat{\beta}_{2SLS}$.

Note 2: (Why is 2SLS called 2SLS)

- $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t$.
- Suppose x_{t2} contains measurement errors. Then, your instrument set would look like $z_{t\bullet} = (x_{t1}, h_{t1}, h_{t2}, \ldots, h_{tr}, x_{t3}, \ldots, x_{tk})'$.
- Regress x_{t2} on $z_{t\bullet}$. Then, get fitted values $\hat{x}_{t2}$.
- Estimate β's by regressing the model:

$$y_t = \beta_1 x_{t1} + \beta_2 \hat{x}_{t2} + \beta_3 x_{t3} + \ldots + \beta_k x_{tk} + \text{error}.$$

- Use the 2SLS residuals $(y_t - \hat{\beta}_1 x_{t1} - \hat{\beta}_2 x_{t2} - \ldots - \hat{\beta}_k x_{tk})$ to estimate σ^2.

Theorem:

$$p\lim_{T\to\infty} \hat{\beta}_{2SLS} = \beta;$$

$$\sqrt{T}\left(\hat{\beta}_{2SLS} - \beta\right) \to_d N\left(0_{k\times 1}, \sigma^2 p\lim_{T\to\infty} T(X'P(Z)X)^{-1}\right).$$

$$\text{plim}_{T\to\infty} s_{2SLS}^2 = \sigma^2 \text{, where } s_{2SLS}^2 = \frac{(y - X\hat{\beta}_{2SLS})'(y - X\hat{\beta}_{2SLS})}{T-k}.$$

[3] Weak Instrumental Variables Problem

(1) What does "weak instrumental variables" mean?

- Consider a simple model: $y_t = x_t\beta + \varepsilon_t$ with z_t such that $E(z_t\varepsilon_t) = 0$.
- When $E(z_t x_t) \neq 0$, but the correlation between the two variables is low, we call z_t a "weak instrumental variable".

(2) What is the problem of the weak instrumental variables?

- The IV or 2SLS estimators are biased toward the OLS estimator, especially when the sample size (T) is small.
- The statistical inferences based on IV or 2SLS are unreliable.

(3) How to check weak instrumental variables?

- Suppose $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \varepsilon_t$,

  where x_{t2} is an endogenous variable.
- Let $z_{t\bullet} = (1, x_{t3}, x_{t4}, ..., h_{t1}, ..., h_{tr})'$.
- Estimate the following model:

$$x_{t2} = \alpha_1 + \alpha_2 x_{t3} + ... + \alpha_{k-1} x_{tk} + \gamma_1 h_{t1} + ... + \gamma_r h_{tr} + error.$$

- Test H_o: $\alpha_2 = ... = \alpha_{k-1} = \gamma_1 = ... = \gamma_r = 0$ (overall significance test).
- Test H_o: $\gamma_1 = ... = \gamma_r = 0$. If you can reject this hypothesis and the F statistic for this hypothesis is greater than 10, you may not have to worry about the weak instrumental variable. [Why 10? See Stock and Watson.]

[4] Testing Exogeneity of Instrumental Variables

(1) Testing Exogeneity of Instrumental Variables

(Hansen, 1982, Econometrica; Hausman, 1984, Handbook)

- Model: $y_t = x_{t\bullet}'\beta + \varepsilon_t$ or $y = X\beta + \varepsilon$.

  Assume the $x_{t\bullet}$ contains 1.
- H_o: $E(z_{t\bullet}\varepsilon_t) = 0_{q\times 1}$.
- Instruments: $z_{t\bullet}$, or Z.
- Test procedure:

  Let $\hat{\varepsilon}$ be the vector of the 2SLS residuals (= $y - X\hat{\beta}_{2SLS}$).
- Regress $\hat{\varepsilon}$ on X and get R^2.
- Then, under the hypothesis $E(z_{t\bullet}\varepsilon_t) = 0_{q\times 1}$,

  $$J_T \equiv TR^2 \to_d \chi^2(q-k).$$
- This test is meaningful only if $q > k$.

(2) Testing Exogeneity of a Single Regressor Based on 2SLS

- Model: $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \varepsilon_t = x_{t\bullet}'\beta + \varepsilon_t$.
- Wish to test whether x_{t2} is exogenous or not.
- H_o: x_{t2} is exogenous.
- Let $z_{t\bullet}$ is the set of instruments under the assumption that x_{t2} is endogenous.
- Test Procedure:
  - Let $\hat{\beta}_{2SLS,1}$ be the 2SLS estimator using Z as instruments with $\Phi_{2SLS,1} = Cov(\hat{\beta}_{2SLS,1})$. Let $\hat{\beta}_{2SLS,2}$ be the 2SLS estimator using [Z,X_2] as instruments with $\Phi_{2SLS,2} = Cov(\hat{\beta}_{2SLS,1})$.
  - $H_T = \left(\hat{\beta}_{2SLS,1} - \hat{\beta}_{2SLS,2}\right)' \left(\Phi_{2SLS,1} - \Phi_{2SLS,2}\right)^+ \left(\hat{\beta}_{2SLS,1} - \hat{\beta}_{2SLS,2}\right) \rightarrow_d \chi^2(1)$.

- Alternative by Newey (1985, JEC) and Eichenbaum, Hasnsen and Singleton (1988, JPE)
  - Let $\hat{\varepsilon}_1$ be the vector of 2SLS residuals using Z as instruments.
  - Let $\hat{\varepsilon}_2$ be the vector of 2SLS residuals using [Z,X_2] as instruments.
  - $J_{T,1} = T\bullet$ (R^2 from the regression of $\hat{\varepsilon}_1$ on Z).
  - $J_{T,2} = T\bullet$ (R^2 from the regression of $\hat{\varepsilon}_2$ on [Z,X_2]).
  - $D_T \equiv J_{T,2} - J_{T,1} \rightarrow_p \chi^2(df = 1)$.
  - This test is asymptotically identical to H_T.

(3) Testing Exogeneity of Two Regressors by the Hausman Test

- Model: $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \varepsilon_t = x_{t\bullet}'\beta + \varepsilon_t$.
- Wish to test whether x_{t2} and x_{t3} are exogenous or not.
- H_o: x_{t2} and x_{t3} are exogenous.
- Let $z_{t\bullet}$ is the set of instruments under the assumption that both x_{t2} and x_{t3} are endogenous.
- Test Procedure:
  - Let $\hat{\beta}_{2SLS,1}$ be the 2SLS estimator using Z as instruments with $\Phi_{2SLS,1} = Cov(\hat{\beta}_{2SLS,1})$. Let $\hat{\beta}_{2SLS,2}$ be the 2SLS estimator using $[Z,X_2,X_3]$ as instruments with $\Phi_{2SLS,2} = Cov(\hat{\beta}_{2SLS,1})$.
  - $H_T = \left(\hat{\beta}_{2SLS,1} - \hat{\beta}_{2SLS,2}\right)' \left(\Phi_{2SLS,1} - \Phi_{2SLS,2}\right)^{+} \left(\hat{\beta}_{2SLS,1} - \hat{\beta}_{2SLS,2}\right) \rightarrow_d \chi^2(2)$.

- Alternative by Newey (1985, JEC) and Eichenbaum, Hasnsen and Singleton (1988, JPE)
  - Let $\hat{\varepsilon}_1$ be the vector of 2SLS residuals using Z as instruments.
  - Let $\hat{\varepsilon}_2$ be the vector of 2SLS residuals using $[Z,X_2,X_3]$ as instruments.
  - $J_{T,1} = T\bullet$ (R^2 from the regression of $\hat{\varepsilon}_1$ on Z).
  - $J_{T,2} = T\bullet$ (R^2 from the regression of $\hat{\varepsilon}_2$ on $[Z,X_2,X_3]$).
  - $D_T \equiv J_{T,2} - J_{T,1} \rightarrow_p \chi^2(df = 2)$.
  - This test is asymptotically identical to H_T.

[EXAMPLE]

- Use **mwemp.wf1**. You can download this file from my web page.
- This is the data set of working married women in 1981 sampled from PSID. Total number of observations are 923, and 17 variables are observed.

VARIABLES	DEFINITION
LRATE	LOG OF HOURLY WAGE RATE ($)
ED	YEARS OF EDUCATION
URB	URB=1 IF RESIDENT IN SMSA
MINOR	MINOR=1 IF BLACK AND HISPANIC
AGE	YEARS OF AGE
TENURE	MONTHS UNDER THE CURRENT EMPLOYER
EXPP	NUMBER OF YEARS WORKED SINCE AGE 18
REGS	REGS=1 IF LIVES IN THE SOUTH OF U.S.
OCCW	OCCW=1 IF WHITE COLOR
OCCB	OCCB=1 IF BLUE COLOR
INDUMG	INDUMG=1 IF IN THE MANUFACTURING INDUSTRY
INDUMN	INDUMN=1 IF NOT IN MANUFACTURING SECTOR
UNION	UNION=1 IF UNION MEMBER
UNEMPR	% UNEMPLOYMENT RATE IN THE RESIDENT'S COUNTY, 1980
LOFINC	LOG OF OTHER FAMILY MEMBER'S INCOME IN 1980 ($)
HWORK	HOURS OF HOMEWORK PER WEEK
KIDS5	NUMBER OF CHILDREN . 5 YEARS OF AGE
LHWORK	ln(HWORK+1)

- Suppose we wish to estimate the following equation:

$$\begin{aligned}\text{LHWORK} \quad = \; & \gamma_{12}\text{LRATE} + \beta_{12} + \beta_{22}\text{ED} + \beta_{32}\text{ED}^2 + \beta_{42}\text{AGE} + \beta_{52}\text{AGE}^2 \\ & + \beta_{62}\text{REGS} + \beta_{72}\text{MINOR} + \beta_{8,2}\text{URB} + \beta_{9,2}\text{KIDS5} \\ & + \beta_{10,2}\text{LOFINC} + \varepsilon_2.\end{aligned}$$

Endogenous (Potentially): LRATE

Exogeous: C, ED, ED^2, ***EXPP, EXPP²***, AGE, AGE^2, ***OCCW, OCCB, UNEMPR***, REGS, MINOR, ***INDUMG, UNION***, URB, KIDS5, LOFINC.

- **First stage OLS Regression**
  - Go to **Object/New Objects...** and choose **Equation**.
  - In the equation box, type:

    LRATE C, ED, ED^2, EXPP, EXPP^2, AGE, AGE^2, OCCW, OCCB, UNEMPR, REGS, MINOR, INDUMG, UNION, URB, KIDS5, LOFINC

    Then, click on **ok**.
  - We will see:

<First-stage OLS>

Dependent Variable: LRATE
Method: Least Squares
Sample: 1 923
Included observations: 923

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.310456	0.273963	1.133203	0.2574
ED	-0.071412	0.028817	-2.478071	0.0134
ED^2	0.005213	0.001137	4.583704	0.0000
EXPP	**0.024203**	**0.005757**	**4.203963**	**0.0000**
EXPP^2	**-0.000383**	**0.000175**	**-2.187337**	**0.0290**
AGE	-0.001903	0.009134	-0.208304	0.8350
AGE^2	-2.56E-05	0.000113	-0.227435	0.8201
OCCW	**0.133112**	**0.029153**	**4.565945**	**0.0000**
OCCB	**0.028524**	**0.041328**	**0.690194**	**0.4902**
UNEMPR	**-0.550904**	**0.442207**	**-1.245806**	**0.2132**
REGS	-0.029157	0.024995	-1.166490	0.2437
MINOR	-0.073556	0.026418	-2.784329	0.0055
INDUMG	**0.134493**	**0.028844**	**4.662849**	**0.0000**
UNION	**0.145738**	**0.027660**	**5.268952**	**0.0000**
URB	0.143252	0.023811	6.016148	0.0000
KIDS5	0.011349	0.016581	0.684462	0.4939
LOFINC	0.116074	0.016043	7.235354	0.0000

R-squared	**0.429705**	Mean dependent var	1.662759
Adjusted R-squared	0.419633	S.D. dependent var	0.399943
S.E. of regression	0.304684	Akaike info criterion	0.479162
Sum squared resid	84.10598	Schwarz criterion	0.568078
Log likelihood	-204.1332	**F-statistic**	**42.66570**
Durbin-Watson stat	1.846204	**Prob(F-statistic)**	**0.000000**

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	**21.42607**	(7, 906)	0.0000
Chi-square	149.9825	7	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(4)	0.024203	0.005757
C(5)	-0.000383	0.000175
C(7)	-2.56E-05	0.000113
C(9)	0.028524	0.041328
C(10)	-0.550904	0.442207
C(13)	0.134493	0.028844
C(14)	0.145738	0.027660

Restrictions are linear in coefficients.

- Comment:
  - Instruments are reasonably higly correlated with LRATE.
  - We could expect that the 2SLS for the structural LHWORK equation would have good finite-sample properties.

<OLS result for the equation>

Dependent Variable: LHWORK
Method: Least Squares
Sample: 1 923
Included observations: 923

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LRATE	-0.294699	0.055614	-5.299027	0.0000
C	0.994632	0.470753	2.112854	0.0349
ED	0.105456	0.051871	2.033064	0.0423
ED^2	-0.003979	0.002076	-1.917036	0.0555
AGE	0.054478	0.014130	3.855400	0.0001
AGE^2	-0.000583	0.000178	-3.276454	0.0011
REGS	-0.059622	0.043971	-1.355931	0.1755
MINOR	-0.051056	0.046553	-1.096743	0.2730
URB	0.020359	0.043634	0.466573	0.6409
KIDS5	0.112525	0.029977	3.753647	0.0002
LOFINC	0.056628	0.029258	1.935435	0.0532

R-squared	0.078925	Mean dependent var	2.909292
Adjusted R-squared	0.068825	S.D. dependent var	0.574604
S.E. of regression	0.554478	Akaike info criterion	1.670266
Sum squared resid	280.3904	Schwarz criterion	1.727800
Log likelihood	-759.8278	F-statistic	7.814705
Durbin-Watson stat	1.622138	Prob(F-statistic)	0.000000

- 2SLS Estimation.
  - Go to **Objects/New Object...** and choose **Equation**.
  - In the **Method** box, click on the arrow. Then, choose **TSLS**.
  - Estimate the first equation by 2SLS using the following procedure.
    - In the equation box, type:

      LRATE LHWORK C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB UNEMPR REGS MINOR INDUMG UNION URB

    - In the instrument list box, type:

    C ED ED^2 EXPP EXPP^2 AGE AGE^2 OCCW OCCB UNEMPR REGS MINOR INDUMG UNION URB KIDS5 LOFINC

  - Do the same things to estimate the second equation.

<2SLS>

Dependent Variable: LHWORK
Method: Two-Stage Least Squares
Sample: 1 923
Included observations: 923
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	-0.417752	0.142145	-2.938921	0.0034
C	1.023219	0.472991	2.163296	0.0308
ED	0.095646	0.053044	1.803140	0.0717
ED^2	-0.003258	0.002218	-1.469314	0.1421
AGE	0.056944	0.014409	3.952136	0.0001
AGE^2	-0.000611	0.000181	-3.378723	0.0008
REGS	-0.067556	0.044888	-1.504993	0.1327
MINOR	-0.059135	0.047460	-1.246000	0.2131
URB	0.038790	0.047934	0.809224	0.4186
KIDS5	0.113335	0.030070	3.769026	0.0002
LOFINC	0.069324	0.032290	2.146922	0.0321

R-squared	0.073980	Mean dependent var	2.909292
Adjusted R-squared	0.063826	S.D. dependent var	0.574604
S.E. of regression	0.555964	Sum squared resid	281.8956
F-statistic	5.843729	Durbin-Watson stat	1.630243
Prob(F-statistic)	0.000000		

- Comment: Interpretation of the results for ED and ED^2
  - $\partial LHWORK/\partial ED = 0.096 + 2*(-0.003)ED = 0.096 - 0.006ED$

    $\rightarrow \partial LHWORK/\partial ED > 0$ until ED = 16; ≤ 0 if from ED ≥ 16.

- Genr TSLSE = RESID

<Exogeneity of Instruments>

Dependent Variable: TSLSE
Method: Least Squares
Sample: 1 923
Included observations: 923

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007946	0.499782	0.015898	0.9873
ED	0.001979	0.052571	0.037648	0.9700
ED^2	7.79E-05	0.002075	0.037542	0.9701
EXPP	-0.008017	0.010503	-0.763331	0.4455
EXPP^2	0.000118	0.000319	0.368304	0.7127
AGE	0.004413	0.016663	0.264810	0.7912
AGE^2	-2.88E-05	0.000205	-0.140328	0.8884
OCCW	-0.038310	0.053183	-0.720330	0.4715
OCCB	0.011739	0.075392	0.155711	0.8763
UNEMPR	-0.834961	0.806703	-1.035029	0.3009
REGS	0.012327	0.045598	0.270328	0.7870
MINOR	-0.010894	0.048193	-0.226055	0.8212
INDUMG	0.007578	0.052618	0.144024	0.8855
UNION	0.086049	0.050459	1.705336	0.0885
URB	0.002664	0.043438	0.061332	0.9511
KIDS5	-0.004513	0.030248	-0.149200	0.8814
LOFINC	-0.004118	0.029266	-0.140712	0.8881

R-squared	**0.007080**	Mean dependent var	-6.22E-15
Adjusted R-squared	-0.010455	S.D. dependent var	0.552941
S.E. of regression	0.555824	Akaike info criterion	1.681516
Sum squared resid	279.8998	Schwarz criterion	1.770432
Log likelihood	-759.0196	F-statistic	0.403760
Durbin-Watson stat	1.631399	Prob(F-statistic)	0.981932

J_T = 0.007080*923 = 6.534 < c = 12.59 (df = 6).

The IVs are exogenous.

<Testing Exogeneity of LRATE>

- 2SLS for the LHWORK Equation Using [X, LRATE] as Instruments

Dependent Variable: LHWORK
Method: Two-Stage Least Squares
Sample: 1 923
Included observations: 923
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	-0.294699	0.055614	-5.299027	0.0000
C	0.994632	0.470753	2.112854	0.0349
ED	0.105456	0.051871	2.033064	0.0423
ED^2	-0.003979	0.002076	-1.917036	0.0555
AGE	0.054478	0.014130	3.855400	0.0001
AGE^2	-0.000583	0.000178	-3.276454	0.0011
REGS	-0.059622	0.043971	-1.355931	0.1755
MINOR	-0.051056	0.046553	-1.096743	0.2730
URB	0.020359	0.043634	0.466573	0.6409
KIDS5	0.112525	0.029977	3.753647	0.0002
LOFINC	0.056628	0.029258	1.935435	0.0532

R-squared	0.078925	Mean dependent var	2.909292
Adjusted R-squared	0.068825	S.D. dependent var	0.574604
S.E. of regression	0.554478	Sum squared resid	280.3904
F-statistic	7.814705	Durbin-Watson stat	1.622138
Prob(F-statistic)	0.000000		

GENR TSLSES = RESID

<EHS Test for the Exogeneity of LRATE>

Dependent Variable: TSLSES
Method: Least Squares
Sample: 1 923
Included observations: 923

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.008618	0.498819	-0.017277	0.9862
ED	0.002554	0.052610	0.048555	0.9613
ED^2	4.04E-05	0.002093	0.019304	0.9846
EXPP	-0.011537	0.010577	-1.090793	0.2757
EXPP^2	0.000173	0.000319	0.542686	0.5875
AGE	0.007156	0.016620	0.430563	0.6669
AGE^2	-5.34E-05	0.000205	-0.260755	0.7943
OCCW	-0.057669	0.053650	-1.074904	0.2827
OCCB	0.007591	0.075214	0.100926	0.9196
UNEMPR	-0.754840	0.805268	-0.937378	0.3488
REGS	0.008633	0.045512	0.189686	0.8496
MINOR	-0.008276	0.048271	-0.171444	0.8639
INDUMG	-0.011982	0.053106	-0.225620	0.8215
UNION	0.064854	0.051091	1.269372	0.2046
URB	0.000261	0.044181	0.005913	0.9953
KIDS5	-0.005353	0.030176	-0.177399	0.8592
LOFINC	-0.008303	0.030020	-0.276567	0.7822
LRATE	0.022382	0.060448	0.370270	0.7113

R-squared	0.008094	Mean dependent var	-7.47E-15
Adjusted R-squared	-0.010538	S.D. dependent var	0.551463
S.E. of regression	0.554361	Akaike info criterion	1.677307
Sum squared resid	278.1209	Schwarz criterion	1.771453
Log likelihood	-756.0771	F-statistic	0.434422
Durbin-Watson stat	1.622910	Prob(F-statistic)	0.977495

$J_T^* = 0.008094*923 = 7.467$

$D_T = J_T^* - J_T = 7.467 - 6.534 = 0.933 < 3.84$ (c at 95%).

Do not reject the exogeneity of LRATE.

[5] Hausman-and-Taylor Panel Data 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) Example:

y_{it} : log of earnings.

α_i : unobservable talent or IQ.

x_{1it} includes age, health, regional variables, etc.

x_{2it} includes work experience.

z_{1i} includes race, union, etc.

z_{2i} includes schooling.

(3) 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 $u_{it}-u_{i,t-1} = \varepsilon_{it} - \varepsilon_{i,t-1}$.]

- Hausman and Taylor Assumption (HTA):

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

- Amemiya and MaCurdy Assumption (AMA):

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

- Breusch, Mizon and Schmidt (BMSA):

  $E(x_{1it}'\overline{u}_i) = 0; E(z_{1i}'\overline{u}_i) = 0; E((x_{2it} - \overline{x}_{2i})'\overline{u}_i) = 0$, for t = 1, 2, ..., T-1.

(4) Estimationa

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. [$\Sigma^{-1/2} = Q_T + \theta P_T$].

- Procedure

STEP 1: Consider the deviation-from-mean equation:

$$Q_T y_i = Q_T X_i\beta + Q_T\varepsilon_i .$$

a) Do OLS and get $\hat{\beta}_W$ (consistent).

b) Using Within residuals, get s^2.

c) $e_i = P_T y_i - P_T X_i\hat{\beta}_W$ is an estimate of $e_T z_i\gamma + e_T\alpha_i + \varepsilon_i$.

STEP 2: Consider the equation $e_i = e_T z_i + err$.

a) Do 2SLS on (7) with IV = [$P_V X_{1i}, e_T z_{1i}$], and get $\hat{\gamma}_W$.

[$\hat{\gamma}_W$ is consistent but not efficient.]

b) Let v_i be the residuals from 2SLS on (7) = $e_i - z_i\hat{\gamma}_W$.

c) Define $s_{BB}^2 = \frac{1}{N}\Sigma_i v_i v_i'$ (consistent for $T\sigma_\alpha^2+\sigma_\varepsilon^2$). Let $\hat{\theta} = [s^2/s_{BB}^2]^{1/2}$.

STEP 3: Consider the following quasi-differenced equation:

$$\Sigma^{-1/2}y_i = \Sigma^{-1/2}H_i\gamma + \text{error} . \qquad (*)$$

a) Do 2SLS on (*) using IV = $G_{HT,i} = (Q_TX_i, P_TX_{1i}, e_Tz_{1i})$, and get

$$\tilde{\delta} = \begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix} = [H'\Omega^{-1/2}P(G_{HT})\Omega^{-1/2}H]^{-1}H'\Omega^{-1/2}P(G_{HT})\Omega^{-1/2}y,$$

where $P(G_{HT}) = G_{HT}(G_{HT}'G_{HT})^{-1}G_{HT}'$.

b) $\text{Cov}(\tilde{\delta}) = s^2[H'\Omega^{-1/2}P(G_{HT})\Omega^{-1/2}H]^{-1}$.

- The Main Results in HT:
  - If $k_1 < g_2$; can't estimate γ.
  - If $k_1 = g_2$; $\tilde{\beta} = \hat{\beta}_W$, and $\tilde{\gamma}$ can be obtained.
  - $k_1 > g_2$; HT is more efficient than Within.
  - If $k_2 = g_2 = 0$; HT estimator = GLS estimator.

- Efficiency of HT estimator:
  - HT estimator is not efficient. [See Amemiya and MaCurdy (1986, AM), Breusch, Mizon and Schmidt (1989, BMS).]
  - BMS is better than AM, and AM is better than HT.

- Testing HT specification:
  - Using $\tilde{\beta}$ and $\hat{\beta}_W$, construct Hausman statistic:

    $$H_T = (\hat{\beta}_W - \tilde{\beta})'[Cov(\hat{\beta}_W) - Cov(\tilde{\beta})]^{+}(\hat{\beta}_W - \tilde{\beta}),$$

    which is χ^2 with df = Rank[$Cov(\hat{\beta}_W) - Cov(\tilde{\beta})$]. [(.)$^+$ means Moore-Penrose generalized inverse.]

    Caution!!! Many textbooks say that df = k_1. However, there are many cases in which df < k_1. For example, if x_{1it} includes time-dummy variables, df = k_1 - # of time dummy variables. In general, if there are time-varying regressors common to everybody, df = k_1 - # of such variables.

  - If H_T > critical value, it implies that the variables which you choose for x_{1it} and z_{1i} may in fact be correlated with α_i. If your model is rejected, then some variables in x_{1it} may have to be moved to x_{2it}. Note that if k_1 = g_2, then, HT = 0. In this case, you cannot use Hausman test statistic for testing specification.

2) AM and BMS 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 $x_{AM,1i} = [x_{1i1}, \ldots , x_{1iT}]$.

- $G_{BMS,i} = [G_{AM,i}, e_T x_{BMS,2i}]$, $G_{BMS} = [G_{AM}, VX_{BMS,2}]$,

  where $x_{BMS,2i} = [x_{2i1} - \bar{x}_{2i}, x_{t2} - \bar{x}_{2i}, \ldots, x_{2i,T-1} - \bar{x}_{2i}]$.

- Under the AM assumptions,

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

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

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

  $\rightarrow E(G_{AM,i}'u_i) = 0.$

- BMS additionally assume $E[x_{BMS,2i}'e_T'u_i] = 0$.
- Reconsider the whitened equation

  $$\Omega^{-1/2}y = \Omega^{-1/2}H\delta + \Omega^{-1/2}u \qquad (6)$$

  We can estimate δ by using more IV than HT use.

- Procedure:

  STEP 1: Consider the deviation-from-mean equation:

  $$Q_V y = Q_V X\beta + Q_V\varepsilon$$

  a) Do OLS and get $\hat{\beta}_W$ (consistent).

  b) Using within residuals, get s^2.

  c) $e_i = P_T y_i - P_T X_i\hat{\beta}_W$ is an estimate of $e_T z_i\gamma + e_T\alpha + \varepsilon_i$.

[Note that e_{it} is the same for all t.]

STEP 2: Set $e_i = e_T z_i \gamma$ + err:

[The err contains α_i and is not correlated with x_{1it} and z_{1i}.]

a) Do 2SLS on this equation with instruments = $[e_T x_{AM,1}, e_T z_1]$, and get the estimate $\hat{\gamma}_W$. (For BMS, use $[e_T x_{AM,1}, e_T z_1, e_T x_{BMS,2}]$.)

b) Let v_i be the residuals from this 2SLS: $v_i = e_i - e_T z_i \hat{\gamma}_W$.

c) $s_{BB}^2 = v'v/N$ (consistent for $T\sigma_\alpha^2 + \sigma_\varepsilon^2$) and $\hat{\theta} = [s^2/s_{BB}^2]^{1/2}$.

STEP 3: $\Sigma^{-1/2}y = \Sigma^{-1/2}H\delta$ + error . (*)

a) Do 2SLS on (*) using AMIV = $G_{AM,i} = [Q_T X_i, e_T x_{AM,1}, e_T z_1]$, and get

$$\tilde{\delta} = \begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix}$$ [AM estimator]. [BMSIV = $[G_{AM,i}, e_T x_{BMS,2i}]$.]

b) AM is more efficient than HT.

- Testing AM specification:
  - Using $\tilde{\beta}$ and $\hat{\beta}_W$, construct Hausman statistic:

$$H_T = (\hat{\beta}_W - \tilde{\beta})'[Cov(\hat{\beta}_W) - Cov(\tilde{\beta})]^+(\hat{\beta}_W - \tilde{\beta}),$$

which is chi-squared with df = min{k, $Tk_1 - g_2$}. If HT > critical value, it implies that the variables which you choose for X_{1it} and Z_{1i} may in fact be correlated with α_i. [For BMS, df = min{k,$Tk_1+(T-1)k_2-g_2$}.

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

- Modified Generalized Instrumental Variables (MGIV) Estimator (Im, Ahn, Schmidt and Wooldridge, 1999; Ahn and Schmidt, 1999)
  - $W_{HT,i} = [Q_T X_i, P_T X_{1i}, e_T z_{1i}]$;
  - $W_{AM,i} = [Q_T X_i, e_T x_{AM,1i}, e_T z_{1i}]$, $x_{AM,1i} = [x_{1i1}, ..., x_{1iT}]$ $(1\times k_1 T)$.
  - $W_{BMS,i}=[W_{AM,i}, e_T x_{BMS,2i}]$, $x_{BMS,2i}=[x_{2i1}-\bar{x}_{2i}, ..., x_{2i,T-1}-\bar{x}_{2i}]$

    $(1\times k_2(T-1))$
  - Assume that the ε_i are homoskedastic across i but heteroskedastic or autocorrelated over t.

Digression to 2SLS and 3SLS

- Notation:
  - For $T\times p$ matrix M_i or $T\times 1$ 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 $T\times q$ matrix of instruments such that $E(W_i'u_i) = 0$.

- Estimators:

1) 2SLS:

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

2) 3SLS:

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

$$\hat{\Omega} = I_N \otimes \hat{\Sigma};$$

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

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

End of Digression

- MGIV estimator:

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

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

$$\tilde{\delta}_{MGIV,BMS} \equiv \hat{\delta}_{3SLS}[\Sigma^{-1}(Q_T X_i, e_T x_{AM,1i}, e_T z_{1i}, e_T x_{BMS,2i})]$$

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

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

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

- The HT (AM, BMS) MGIV estimator is efficient under the HT (AM, BMS) assumptions.
- Hausman test for HT, AM and BMS:

  Based on KR-MGIV and MGIV estimators of β.