# Single Equation Linear GMM

Eric Zivot

Winter 2013

**Single Equation Linear GMM**

Consider the linear regression model

$$\begin{aligned} y_t &= \mathbf{z}_t' \boldsymbol{\delta}_0 + \varepsilon_t,\ t = 1, \ldots, n \\ \mathbf{z}_t &= L \times 1 \text{ vector of explanatory variables} \\ \boldsymbol{\delta}_0 &= L \times 1 \text{ vector of unknown coefficients} \\ \varepsilon_t &= \text{random error term} \end{aligned}$$

Engodeneity

The model allows for the possibility that some or all of the elements of $z_t$ may be correlated with the error term $\varepsilon_t$ (i.e., $E[z_{tk}\varepsilon_t] \neq 0$ for some $k$).

If $E[z_{tk}\varepsilon_i] \neq 0$, then $z_{tk}$ is called an *endogenous variable*.

If $\mathbf{z}_t$ contains endogenous variables, then the least squares estimator of $\boldsymbol{\delta}_0$ in is biased and inconsistent.

Instruments

It is assumed that there exists a $K \times 1$ vector of *instrumental variables* $\mathbf{x}_t$ that may contain some or all of the elements of $\mathbf{z}_t$.

Let $\mathbf{w}_t$ represent the vector of unique and nonconstant elements of $\{y_t, \mathbf{z}_t, \mathbf{x}_t\}$.

It is assumed that $\{\mathbf{w}_t\}$ is a stationary and ergodic stochastic process.

## Moment Conditions for General Model

Define

$$\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0) = \mathbf{x}_t \varepsilon_t = \mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)$$

It is assumed that the instrumental variables $\mathbf{x}_t$ satisfy the set of $K$ orthogonality conditions

$$E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0)] = E[\mathbf{x}_t \varepsilon_t] = E[\mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)] = \mathbf{0}$$

Expanding gives the relation

$$\begin{aligned} E[\mathbf{x}_t y_t] - E[\mathbf{x}_t \mathbf{z}_t']\boldsymbol{\delta}_0 &= \mathbf{0} \\ \Rightarrow \underset{(K\times 1)}{\boldsymbol{\Sigma}_{xy}} &= \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}_0} \end{aligned}$$

where $\boldsymbol{\Sigma}_{xy} = E[\mathbf{x}_t y_t]$ and $\boldsymbol{\Sigma}_{xz} = E[\mathbf{x}_t \mathbf{z}_t']$.

Example: Demand-Supply model with supply shifter

$$\begin{aligned} \text{demand:} \qquad & q_i^d = \alpha_0 + \alpha_1 p_i + u_i \\ \text{supply:} \qquad & q_i^s = \beta_0 + \beta_1 p_i + \beta_2 \text{temp}_i + v_i \\ \text{equilibrium:} \qquad & q_i^d = q_i^s = q_i \\ E[\text{temp}_i u_i] \ = \ & E[\text{temp}_i v_i] = 0 \end{aligned}$$

Goal: Estimate demand equation; $\alpha_1 =$ demand elasticity if data are in logs

Remark

Simultaneity causes $p_i$ to be endogenous in demand equation: $E[p_i u_i] \neq 0$

Why? Use equilibrium condition, solve for $p_i$ and compute $E[p_i u_i]$

Demand/Supply model in Hayashi notation

$$
\begin{aligned}
y_i &= \mathbf{z}_i' \boldsymbol{\delta} + \varepsilon_i \\
y_i &= q_i, \\
\mathbf{z}_i &= (1, p_i)', \boldsymbol{\delta} = (\alpha_0, \alpha_1)' \\
\mathbf{x}_i &= (1, \mathsf{temp}_i)',\ \mathbf{w}_i = (q_i, p_i, \mathsf{temp}_i)' \\
L &= 2,\ K = 2
\end{aligned}
$$

Note: 1 is common to both $\mathbf{z}_i$ and $\mathbf{x}_i$.

Example: Wage Equation

$$\begin{aligned} \ln W_i &= \delta_1 + \delta_2 S_i + \delta_3 EXPR_i + \delta_4 IQ_i + \varepsilon_i \\ W_i &= \text{wages} \\ S_i &= \text{years of schooling} \\ EXPR_i &= \text{years of experience} \\ IQ_i &= \text{score on IQ test} \\ \delta_2 &= \text{rate of return to schooling} \end{aligned}$$

Assume

$$E[S_i \varepsilon_i] = E[EXPR_i \varepsilon_i] = 0$$

Endogeneity: $IQ_i$ is a proxy for unobserved ability but is measured with error

$$\begin{aligned} IQ_i &= ABILITY_i + error_i \\ &\Rightarrow E[IQ_i \varepsilon_i] \neq 0 \end{aligned}$$

Instruments:

$$
\begin{aligned}
AGE_i &= \text{age in years} \\
MED_i &= \text{years of mother's education} \\
E[AGE_i \varepsilon_i] &= E[MED_i \varepsilon_i] = 0
\end{aligned}
$$

Wage Equation in Hayashi notation

$$
\begin{aligned}
y_i &= \mathbf{z}_i'\boldsymbol{\delta} + \varepsilon_i \\
y_i &= \ln W_i, \\
\mathbf{z}_i &= (1, S_i, EXPR_i, IQ_i)' \\
\mathbf{x}_i &= (1, S_i, EXPR_i, AGE_i, MED_i)' \\
\mathbf{w}_i &= (\ln W_i, S_i, EXPR_i, IQ_i, AGE_i, MED_i)' \\
L &= 4,\ K = 5
\end{aligned}
$$

Remarks

1. $1, S_i, EXPR_i$ are often called included exogenous variables. That is, they are included in the behavioral wage equation.

2. $AGE_i, MED_i$ are often called excluded exogenous variables. That is, they are excluded from the behavioral wage equation.

$y_t = z_t{}' \delta_0 + \epsilon_t$ (1×L, L×1)    $x_t = (x_{1t}, \ldots, x_{Kt})$

$E[\underset{\sim}{x_t} \epsilon_t] = 0$ (K×1)    K moment conditions

## Identification

Identification means that $\boldsymbol{\delta}_0$ is the unique solution to the moment equations

$$E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0)] = 0$$

K×1

That is, $\boldsymbol{\delta}_0$ is identified provided

$$\begin{aligned} E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0)] &= 0 \text{ and } E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta})] \neq 0 \\ \text{for } \boldsymbol{\delta} &\neq \boldsymbol{\delta}_0 \end{aligned}$$

For the linear GMM model, this is equivalent to

$$\begin{aligned} \underset{(K\times 1)}{\boldsymbol{\Sigma}_{xy}} &= \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}_0} \\ \underset{(K\times 1)}{\boldsymbol{\Sigma}_{xy}} &\neq \underset{(K\times L)}{\boldsymbol{\Sigma}_{xz}} \underset{(L\times 1)}{\boldsymbol{\delta}} \text{ for } \boldsymbol{\delta} \neq \boldsymbol{\delta}_0 \end{aligned}$$

$$E\begin{bmatrix} x_{1t}\epsilon_t \\ x_{2t}\epsilon_t \\ \vdots \\ x_{Kt}\epsilon_t \end{bmatrix} = \begin{bmatrix} x_{1t}(y_t - z_t{}'\delta_0) \\ \vdots \\ x_{Kt}(y_t - z_t{}'\delta_0) \end{bmatrix}$$

## Rank Condition

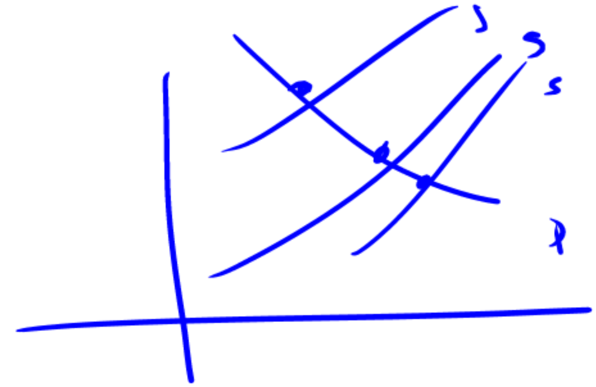
For identification of $\boldsymbol{\delta}_0$, it is required that the $K \times L$ matrix $E[\mathbf{x}_t \mathbf{z}_t'] = \boldsymbol{\Sigma}_{xz}$ be of full rank $L$. This *rank condition* is usually stated as

$$\text{rank}(\boldsymbol{\Sigma}_{xz}) = L$$

This condition ensures that $\boldsymbol{\delta}_0$ is the unique solution to $E[\mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)] = \mathbf{0}$.

Note: If $K = L$, then $\boldsymbol{\Sigma}_{xz}$ is invertible and $\boldsymbol{\delta}_0$ may be determined using

$$\boldsymbol{\delta}_0 = \boldsymbol{\Sigma}_{xz}^{-1} \boldsymbol{\Sigma}_{xy}$$



Example: Demand/Supply Equation

$$
\begin{aligned}
\mathbf{x}_i &= (1, \mathsf{temp}_i)', \mathbf{z}_i = (1, p_i)' \\
L &= 2, K = 2 \\
\boldsymbol{\Sigma}_{xz} &= E[\mathbf{x}_i \mathbf{z}_i'] = \begin{pmatrix} 1 & E[p_i] \\ E[\mathsf{temp}_i] & E[\mathsf{temp}_i p_i] \end{pmatrix}
\end{aligned}
$$

For the rank condition, we need $\mathsf{rank}(\boldsymbol{\Sigma}_{zx}) = L = 2$. Now, $\mathsf{rank}(\boldsymbol{\Sigma}_{xz}) = 2$ if $\mathsf{det}(\boldsymbol{\Sigma}_{xz}) \neq 0$. Further,

$$
\begin{aligned}
\mathsf{det}(\boldsymbol{\Sigma}_{xz}) &= E[\mathsf{temp}_i p_i] - E[\mathsf{temp}_i] E[p_i] \\
&= \mathsf{cov}(\mathsf{temp}_i, p_i)
\end{aligned}
$$

Hence, $\mathsf{rank}(\boldsymbol{\Sigma}_{xz}) = 2$ provided $\mathsf{cov}(\mathsf{temp}_i, p_i) \neq 0$. That is, $\boldsymbol{\delta}_0$ is identified provided the instrument is correlated with the endogenous variable.

## Order Condition

A necessary condition for the identification of $\boldsymbol{\delta}_0$ is the *order condition*

$$K \geq L$$

which simply states that the number of instrumental variables must be greater than or equal to the number of explanatory variables.

If $K = L$, then $\boldsymbol{\delta}_0$ is said to be (apparently) just identified;

if $K > L$ then $\boldsymbol{\delta}_0$ is said to be (apparently) overidentified;

if $K < L$ then $\boldsymbol{\delta}_0$ is not identified.

Note: the word "apparently" in parentheses is used to remind the reader that the rank condition $\text{rank}(\boldsymbol{\Sigma}_{xz}) = L$ must also be satisfied for identification.

## Conditional Heteroskedasticity

$\Rightarrow E[x_t \varepsilon_t] = 0$

In the regression model the error terms are allowed to be conditionally heteroskedastic and/or serially correlated. Serially correlated errors will be treated later on.

$E[x_t \varepsilon_t | I_{t-1}] = 0$, $I_t = \{(x_t, \varepsilon_t), (x_{t-1}, \varepsilon_{t-1}), \ldots (x_1, \varepsilon_1)\}$

For the case in which $\varepsilon_t$ is conditionally heteroskedastic, it is assumed that $\{\mathbf{g}_t\} = \{\mathbf{x}_t\varepsilon_t\}$ is a stationary and ergodic martingale difference sequence (MDS) satisfying

$$
\begin{aligned}
E[\mathbf{g}_t\mathbf{g}_t'] &= E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] = \mathbf{S} \\
\mathbf{S} &= \text{nonsingular } K \times K \text{ matrix}
\end{aligned}
$$

= asymptotic variance of moment conditions.

Note: if $\text{var}(\varepsilon_t|\mathbf{x}_t) = \sigma^2$ then $\mathbf{S} = \sigma^2\mathbf{\Sigma}_{xx}$.

this is the assumption used for 2SLS.

Remark:

The matrix $\mathbf{S}$ is the asymptotic variance-covariance matrix of the sample moments $\bar{\mathbf{g}} = n^{-1}\sum_{t=1}^{n}\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\delta}_0)$. This follows from the central limit theorem for ergodic stationary martingale difference sequences (see Hayashi, 2000, p. 106)

$$\begin{aligned}\sqrt{n}\bar{\mathbf{g}} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{x}_t\varepsilon_t \xrightarrow{d} N(\mathbf{0}, \mathbf{S}) \\ \mathbf{S} &= E[\mathbf{g}_t\mathbf{g}_t'] = \text{avar}(\sqrt{n}\bar{\mathbf{g}})\end{aligned}$$

## Definition of the GMM Estimator

The GMM estimator of $\boldsymbol{\delta}_0$ is constructed by exploiting the orthogonality conditions $E[\mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}_0)] = \mathbf{0}$. The idea is to create a set of estimating equations for $\boldsymbol{\delta}_0$ by making sample moments match the population moments.

The sample moments for an arbitrary value $\boldsymbol{\delta}$ are

$$\begin{aligned}\mathbf{g}_n(\boldsymbol{\delta}) &= \frac{1}{n}\sum_{t=1}^{n} g(\mathbf{w}_t, \boldsymbol{\delta}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t(y_t - \mathbf{z}_t'\boldsymbol{\delta}) \\ &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_{1t}(y_t - \mathbf{z}_t'\boldsymbol{\delta}) \\ \vdots \\ \frac{1}{n}\sum_{t=1}^{n} x_{Kt}(y_t - \mathbf{z}_t'\boldsymbol{\delta}) \end{pmatrix}\end{aligned}$$

These moment conditions are a set of $K$ linear equations in $L$ unknowns.

Equating these sample moments to the population moment $E[\mathbf{x}_t \varepsilon_t] = \mathbf{0}$ gives the estimating equations

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t (y_t - \mathbf{z}_t' \boldsymbol{\delta}) = \mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta} = \mathbf{0}$$

where

$$\begin{aligned} \mathbf{S}_{xy} &= n^{-1}\sum_{t=1}^{n} \mathbf{x}_t y_t \\ \mathbf{S}_{xz} &= n^{-1}\sum_{t=1}^{n} \mathbf{x}_t \mathbf{z}_t' \end{aligned}$$

are the sample moments.

$K = L$ : $\underset{L\times L}{S_{xz}}\hat{\delta} = \underset{L\times 1}{S_{xy}} \Rightarrow \hat{\delta} = S_{xz}^{-1} S_{xy}$

= ILS estimator

IV estimator

Analytical Solution: Just Identified Model

If $K = L$ ($\boldsymbol{\delta}_0$ is just identified) and $\mathbf{S}_{xz}$ is invertible, then the GMM estimator of $\boldsymbol{\delta}_0$ is

$$\hat{\boldsymbol{\delta}} = \mathbf{S}_{xz}^{-1}\mathbf{S}_{xy}$$

which is also known as the *indirect least squares* (ILS) or *instrumental variables* (IV) estimator.

Remark: It is straightforward to establish the consistency of $\hat{\boldsymbol{\delta}}$

Since $y_t = \mathbf{z}_t' \boldsymbol{\delta}_0 + \varepsilon_t$

$$
\begin{aligned}
S_{xy} &= n^{-1} \sum_{t=1}^{n} \mathbf{x}_t y_t = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t (\mathbf{z}_t' \boldsymbol{\delta}_0 + \varepsilon_t) \\
&= \mathbf{S}_{xz} \boldsymbol{\delta}_0 + \mathbf{S}_{x\varepsilon}
\end{aligned}
$$

Then

$$
\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = \mathbf{S}_{xz}^{-1} \mathbf{S}_{x\varepsilon}
$$

By the ergodic theorem and Slutsky's theorem

$$\mathbf{S}_{xz} = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{z}_t' \xrightarrow{p} E[\mathbf{x}_t \mathbf{z}_t'] = \mathbf{\Sigma}_{xz}$$

$$\mathbf{S}_{xz}^{-1} \xrightarrow{p} \mathbf{\Sigma}_{xz}^{-1}, \text{ provided } \text{rank}(\mathbf{\Sigma}_{xz}) = L$$

$$\mathbf{S}_{x\varepsilon} = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \boldsymbol{\varepsilon}_t \xrightarrow{p} E[\mathbf{x}_t \varepsilon_t] = \mathbf{0}$$

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = \mathbf{S}_{xz}^{-1} \mathbf{S}_{x\varepsilon} \xrightarrow{p} 0$$

Analytic Solution: Overidentified Model

If $K > L$, then are $K$ equations in $L$ unknowns. In general, there is no solution to the estimating equations

$$\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta} = \mathbf{0}$$

GMM Solution: Use the method of minium chi-square. The idea is to try to find a $\boldsymbol{\delta}$ that makes $\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta}$ as close to zero as possible. This is achieved by minimizing a weighted average of the $K$ moment equations.

## GMM Objective Function

Let $\hat{\mathbf{W}}$ denote a $K \times K$ symmetric and positive definite (p.d.) weight matrix, possibly dependent on the data, such that $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ as $n \longrightarrow \infty$ with $\mathbf{W}$ symmetric and p.d. The GMM objective function is defined as

$$\begin{aligned} J(\boldsymbol{\delta}, \hat{\mathbf{W}}) &= n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\delta}) \\ &= n(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta})'\hat{\mathbf{W}}(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta}) \end{aligned}$$

$$y_t = z_t \delta_0 + \epsilon_t, \quad x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$$

$$g_t = x_t \epsilon_t = \begin{bmatrix} x_{1t}\epsilon_t \\ x_{2t}\epsilon_t \end{bmatrix} = \begin{bmatrix} x_{1t}(y_t - z_t\delta_0) \\ x_{2t}(y_t - z_t\delta_0) \end{bmatrix}, \quad E[g_t] = \underset{2\times 1}{0}$$

Remarks

1. The scaling by $n$ will be explained later.

2. The elements of $\hat{\mathbf{W}}$ determine the weights put on the $K$ moment equations. For example, let $K = 2, L = 1$ so that

$$\mathbf{g}_n(\delta) = \begin{pmatrix} g_{n,1}(\delta) \\ g_{n,2}(\delta) \end{pmatrix}, \hat{\mathbf{W}} = \begin{pmatrix} \hat{w}_{11} & \hat{w}_{12} \\ \hat{w}_{12} & \hat{w}_{22} \end{pmatrix}$$

$$g_{n,1}(\delta) = \frac{1}{n}\sum_1^n x_{1t}(y_t - z_t\delta)$$

$$g_{n,2}(\delta) = \frac{1}{n}\sum_1^n x_{2t}(y_t - z_t\delta)$$

Then

$$J(\boldsymbol{\delta}, \hat{\mathbf{W}}) = n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\delta})$$
$$= n \begin{pmatrix} g_{n,1}(\delta) & g_{n,2}(\delta) \end{pmatrix}' \begin{pmatrix} \hat{w}_{11} & \hat{w}_{12} \\ \hat{w}_{12} & \hat{w}_{22} \end{pmatrix} \begin{pmatrix} g_{n,1}(\delta) \\ g_{n,2}(\delta) \end{pmatrix}$$
$$n(\hat{w}_{11}g_{n,1}(\delta)^2 + 2\hat{w}_{12}g_{n,1}(\delta)g_{n,2}(\delta) + \hat{w}_{22}g_{n,2}(\delta)^2)$$

3. We will discuss how to determine the best $\hat{\mathbf{W}}$ later

## Definition of GMM estimator

The GMM estimator of $\boldsymbol{\delta}_0$, denoted $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$, is defined as

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= \arg\min_{\boldsymbol{\delta}} \; J(\boldsymbol{\delta}, \hat{\mathbf{W}}) \\
& \quad \arg\min_{\boldsymbol{\delta}} n(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta})'\hat{\mathbf{W}}(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta})
\end{aligned}
$$

Since $J(\boldsymbol{\delta}, \hat{\mathbf{W}})$ is a simple quadratic form in $\boldsymbol{\delta}$, straightforward calculus may be used to determine the analytic solution for $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$:

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xy}$$

$y_t = z_t' \delta_0 + \epsilon_t$ (1×L, L×1)

$K \geq L$

$E\{x_t \epsilon_t\} = 0$ (K×1)

$\frac{1}{T}\sum_1^T x_t \epsilon_t = \frac{1}{T}\sum_1^T x_t (y_t - z_t'\delta)$

## Derivation of GMM estimator

First, note that

$$
\begin{aligned}
J(\boldsymbol{\delta}, \hat{\mathbf{W}}) &= n(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta})'\hat{\mathbf{W}}(\mathbf{S}_{xy} - \mathbf{S}_{xz}\boldsymbol{\delta}) \\
&= n\left[\mathbf{S}'_{xy}\hat{\mathbf{W}}\mathbf{S}_{xy} - 2\mathbf{S}'_{xy}\hat{\mathbf{W}}\mathbf{S}_{xz}\boldsymbol{\delta} + \boldsymbol{\delta}'\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz}\boldsymbol{\delta}\right]
\end{aligned}
$$

Then

$$
\frac{J(\boldsymbol{\delta}, \hat{\mathbf{W}})}{\partial \boldsymbol{\delta}} = -2\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xy} + 2\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz}\boldsymbol{\delta}
$$

Solving for $\hat{\boldsymbol{\delta}}$ such that $\frac{J(\hat{\boldsymbol{\delta}}, \hat{\mathbf{W}})}{\partial \boldsymbol{\delta}} = \mathbf{0}$ gives

$$
\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = (\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xy}
$$

provided $\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz}$ is invertible (which requires rank$(\mathbf{S}_{xz}) = L$).

Recall the results from matrix algebra

$$\frac{\partial(\mathbf{a}'\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \mathbf{a}$$

$$\frac{\partial(\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = 2\mathbf{A}\boldsymbol{\beta} \text{ for } \mathbf{A} \text{ symmetric}$$

## Asymptotic Properties

Under standard regularity conditions (see Hayashi, 2000, C hap. 3), it can be shown that

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xy} \xrightarrow{p} \boldsymbol{\delta}_0$$

$$\sqrt{n}\left(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0\right) \xrightarrow{d} N(\mathbf{0}, \text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})))$$

where

$$\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = (\boldsymbol{\Sigma}_{xz}'\mathbf{W}\boldsymbol{\Sigma}_{xz})^{-1}\boldsymbol{\Sigma}_{xz}'\mathbf{W}\mathbf{S}\mathbf{W}\boldsymbol{\Sigma}_{xz}(\boldsymbol{\Sigma}_{xz}'\mathbf{W}\boldsymbol{\Sigma}_{xz})^{-1}$$

bread    meat    bread

Sandwich form.

robust to conditional heteroskedasticity.

A consistent estimate of $\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$, denoted $\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$, may be computed using

$$\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\hat{\mathbf{S}}\hat{\mathbf{W}}\mathbf{S}_{xz}(\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}$$

where $\hat{\mathbf{S}}$ is a consistent estimate for $\mathbf{S} = \text{avar}(\bar{\mathbf{g}})$.

Note: $\mathbf{S}$ can be consistently estimated using the White or heteroskedasticity consistent (HC) estimator

$$\begin{aligned} \hat{\mathbf{S}}_{HC} &= \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'\hat{\varepsilon}_t^2 \xrightarrow{p} S \text{ b/c } \hat{\delta}(\hat{W}) \xrightarrow{p} \delta_0 \\ \hat{\varepsilon}_t &= y_t - \mathbf{z}_t\hat{\delta}(\hat{\mathbf{W}}) \end{aligned}$$

because $\hat{\delta}(\hat{\mathbf{W}}) \xrightarrow{p} \boldsymbol{\delta}_0$.

$S_{xz} = \hat{\Sigma}_{xz}$, $\hat{W}$, $\hat{S}$.

$$S_{xy} = \frac{1}{n}\sum_{t=1}^{n} x_t y_t = \frac{1}{n}\sum_{t=1}^{n} x_t (z_t'\delta_0 + \epsilon_t) = \frac{1}{n}\sum_{t=1}^{n} x_t z_t'\delta_0 + \frac{1}{n}\sum_{t=1}^{n} x_t \epsilon_t = S_{xz}\delta_0 + S_{x\epsilon}$$

Sketch of Proof

For consistency, note that

$$\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xy} \\
&= (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}(\mathbf{S}_{xz}\boldsymbol{\delta}_0 + \mathbf{S}_{x\varepsilon}) \\
&= \boldsymbol{\delta}_0 + (\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\hat{\mathbf{W}}\mathbf{S}_{x\varepsilon}
\end{aligned}$$

By assumption $\hat{\mathbf{W}} \rightarrow \mathbf{W}$ p.d. By the ergodic theorem and Slutsky's theorem

$$\mathbf{S}_{xz} = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \mathbf{z}_t' \xrightarrow{p} E[\mathbf{x}_t \mathbf{z}_t'] = \mathbf{\Sigma}_{xz}$$

$$\mathbf{S}_{xz}' \hat{\mathbf{W}} \xrightarrow{p} \mathbf{\Sigma}_{xz}' \mathbf{W}$$

$$\left(\mathbf{S}_{xz}' \hat{\mathbf{W}} \mathbf{S}_{xz}\right)^{-1} \xrightarrow{p} \left(\mathbf{\Sigma}_{xz}' \mathbf{W} \mathbf{\Sigma}_{xz}\right)^{-1}, \text{ provided rank}(\mathbf{\Sigma}_{xz}) = L$$

$$\mathbf{S}_{x\varepsilon} = n^{-1} \sum_{t=1}^{n} \mathbf{x}_t \boldsymbol{\varepsilon}_t \xrightarrow{p} E[\mathbf{x}_t \varepsilon_t] = \mathbf{0}$$

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0 \xrightarrow{p} 0$$

For asymptotic normality, write

$$\sqrt{n}\left(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta}_0\right) = \left(\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz}\right)^{-1}\mathbf{S}'_{xz}\hat{\mathbf{W}}\sqrt{n}\mathbf{S}_{x\varepsilon}$$

By the ergodic theorem and Slutsky's theorem

$$\mathbf{S}_{xz} \xrightarrow{p} E[\mathbf{x}_t\mathbf{z}'_t] = \mathbf{\Sigma}_{xz}$$

$$\left(\mathbf{S}'_{xz}\hat{\mathbf{W}}\mathbf{S}_{xz}\right)^{-1} \xrightarrow{p} \left(\mathbf{\Sigma}'_{xz}\mathbf{W}\mathbf{\Sigma}_{xz}\right)^{-1}$$

Since $\{g_t\} = \{\mathbf{x}_t\varepsilon_t\}$ is a stationary and ergodic martingale difference sequence (MDS) satisfying

$$\begin{aligned} E[\mathbf{g}_t\mathbf{g}'_t] &= E[\mathbf{x}_t\mathbf{x}'_t\varepsilon_t^2] = \mathbf{S} \\ \mathbf{S} &= \text{nonsingular } K \times K \text{ matrix} \end{aligned}$$

the CLT for stationary-ergodic MDS gives

$$\sqrt{n}\mathbf{S}_{x\varepsilon} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{x}_t\varepsilon_t \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$$

Furthermore, Slutsky's theorem gives

$$\mathbf{S}'_{xz}\hat{\mathbf{W}}\sqrt{n}\mathbf{S}_{x\varepsilon} \xrightarrow{d} \mathbf{\Sigma}'_{xz}\mathbf{W}\cdot N(\mathbf{0},\mathbf{S})$$

Therefore

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})-\boldsymbol{\delta}_0\right) &\xrightarrow{d} \left(\mathbf{\Sigma}'_{xz}\mathbf{W}\mathbf{\Sigma}_{xz}\right)^{-1}\mathbf{\Sigma}'_{xz}\mathbf{W}\cdot N(\mathbf{0},\mathbf{S}) \\ \equiv\ & N\left(\mathbf{0},\left(\mathbf{\Sigma}'_{xz}\mathbf{W}\mathbf{\Sigma}_{xz}\right)^{-1}\mathbf{\Sigma}'_{xz}\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{\Sigma}_{xz}\left(\mathbf{\Sigma}'_{xz}\mathbf{W}\mathbf{\Sigma}_{xz}\right)^{-1}\right)\end{aligned}$$

Note:

$$\mathbf{A}\cdot N(\mathbf{0},\mathbf{S}) \equiv N(\mathbf{0},\mathbf{A}\mathbf{S}\mathbf{A}')$$

$$A = (\Sigma_{xz}' W \Sigma_{xz})^{-1} \Sigma_{xz}' W S W \Sigma_{xz}$$

### The Efficient GMM Estimator

For a given set of instruments $\mathbf{x}_t$, the GMM estimator $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ is defined for an arbitrary p.d. and symmetric weight matrix $\hat{\mathbf{W}}$. The asymptotic variance of $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ depends on the chosen weight matrix $\hat{\mathbf{W}}$

$$\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = (\boldsymbol{\Sigma}'_{xz}\mathbf{W}\boldsymbol{\Sigma}_{xz})^{-1}\boldsymbol{\Sigma}'_{xz}\mathbf{W}\mathbf{S}\mathbf{W}\boldsymbol{\Sigma}_{xz}(\boldsymbol{\Sigma}'_{xz}\mathbf{W}\boldsymbol{\Sigma}_{xz})^{-1}$$

A natural question to ask is: What weight matrix $\mathbf{W}$ produces the smallest value of avar$(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$? The GMM estimator constructed with this weight matrix is called the *efficient GMM estimator*.

For TSLS $\hat{W} = S_{xx}^{-1}$, so

$$\hat{\text{avar}}(\hat{\delta}_{TSLS}) = (S_{xz}' S_{xx}^{-1} S_{xz})^{-1} S_{xz}' S_{xx}^{-1} \hat{S} S_{xx}^{-1} S_{xz} ( \quad )^{-1}$$

$$\hat{S} = \frac{1}{n} \sum_{t=1}^{n} x_t x_t' (y_t - z_t' \hat{\delta}_{TSLS})^2$$

**Result**: Hansen (1982) showed that efficient the GMM estimator results from setting $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$ such that $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S} = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] = \text{avar}\,(\sqrt{n}\mathbf{g}_n(\boldsymbol{\delta}_0))$. For this choice of $\hat{\mathbf{W}}$, the asymptotic variance formula reduces to

$$\text{avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})) = (\boldsymbol{\Sigma}_{xz}'\mathbf{S}^{-1}\boldsymbol{\Sigma}_{xz})^{-1}$$

of which a consistent estimate is

$$\widehat{\text{avar}}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1})) = (\mathbf{S}_{xz}'\hat{\mathbf{S}}^{-1}\mathbf{S}_{xz})^{-1}$$

$$\text{avar}(\hat{\delta}(\hat{W})) = (\Sigma_{xz}'W\Sigma_{xz})^{-1}\Sigma_{xz}'WSW\Sigma_{xz}(\Sigma_{xz}'W\Sigma_{xz})^{-1}$$

$$W = S^{-1} \Rightarrow \text{avar}(\hat{\delta}(\hat{S}^{-1})) = (\Sigma_{xz}'S^{-1}\Sigma_{xz})^{-1}$$

$$\text{avar}(\hat{\delta}(\hat{W})) - \text{avar}(\hat{\delta}(\hat{S}^{-1})) = \text{psd matrix}$$

for any $\hat{W} \neq \hat{S}^{-1}$

The efficient GMM estimator is defined as

$$\begin{aligned}\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}^{-1}) &= \arg\min_{\boldsymbol{\delta}} \; J(\boldsymbol{\delta}, \hat{\mathrm{S}}^{-1}) \\ &= \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}^{-1}\mathbf{g}_n(\boldsymbol{\delta})\end{aligned}$$

which requires a consistent estimate of $\mathbf{S} = E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2]$.

However, a consistent estimation of $\mathbf{S}$, in turn, requires a consistent estimate of $\boldsymbol{\delta}_0$. White (1982) showed that a consistent estimate of $\mathbf{S}$ has the form

$$\hat{\mathrm{S}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'\hat{\varepsilon}_t^2 = \frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_t\mathbf{x}_t'(y_t - \mathbf{z}_t'\hat{\boldsymbol{\delta}})^2$$

such that $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$.

## Two-Step Efficient GMM

The two-step efficient GMM estimator utilizes the result that a consistent estimate of $\boldsymbol{\delta}_0$ may be computed by GMM with an arbitrary positive definite and symmetric weight matrix $\hat{\mathbf{W}}$ such that $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$. Let $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$ denote such an estimate.

Common choices for $\hat{\mathbf{W}}$ are $\hat{\mathbf{W}} = \mathbf{I}_k$ and $\hat{\mathbf{W}} = \mathbf{S}_{xx}^{-1} = (n^{-1}\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X}$ is an $n \times k$ matrix with the $t$th row equal to $\mathbf{x}_t'$.

Then, a first step consistent estimate of $\mathbf{S}$ is given by

$$\hat{\mathbf{S}}(\hat{\mathbf{W}}) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t\mathbf{x}_t'(y_t - \mathbf{z}_t'\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))^2$$

The *two-step efficient GMM estimator* is then defined as

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}})) = \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}})\mathbf{g}_n(\boldsymbol{\delta})$$
$$\Rightarrow \hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}})) = (\mathbf{S}'_{xz}\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}})\mathbf{S}_{xz})^{-1}\mathbf{S}'_{xz}\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}})\mathbf{S}_{xy}$$

Drawback: The numerical value of $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}}))$ depends on $\hat{\mathbf{W}}$

$$\hat{\delta}(\hat{S}^{-1}(I_K)) \neq \hat{\delta}(\hat{S}^{-1}(S_{xx}^{-1}))$$ in finite samples!

However, it straightforward to show that

$$\hat{\delta}(\hat{S}^{-1}(I_K)) - \hat{\delta}(\hat{S}^{-1}(S_{xx}^{-1})) \xrightarrow{p} 0$$

asymptotically, $\hat{\delta}(\hat{S}^{-1}(I_K)) = \hat{\delta}(\hat{S}^{-1}(S_{xx}^{-1}))$

## Iterated Efficient GMM

The *iterated efficient GMM estimator* uses the two-step efficient GMM estimator $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}(\hat{\mathbf{W}}))$ to update the estimation of $\mathbf{S}$ and then recomputes the estimator. The process is repeated (iterated) until the estimates of $\boldsymbol{\delta}_0$ do not change significantly from one iteration to the next. Typically, only a few iterations are required. The resulting estimator is denoted $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}_{\text{iter}}^{-1})$.

The iterated efficient GMM estimator has the same asymptotic distribution as the two-step efficient estimator. However, in finite samples, the two estimators may differ.

The iterated GMM estimator has a practical advantage over the two-step estimator in that the resulting estimates are invariant with respect to the scale of the data and to the initial weighting matrix $\hat{\mathbf{W}}$.

## Continuous Updating Efficient GMM

This estimator simultaneously estimates $\mathbf{S}$, as a function of $\boldsymbol{\delta}$, and $\boldsymbol{\delta}$. It is defined as

$$\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}_{\mathsf{CU}}) &= \arg\min_{\boldsymbol{\delta}} \; J(\boldsymbol{\delta}, \hat{\mathbf{S}}^{-1}(\boldsymbol{\delta})) \\
&= \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathbf{S}}^{-1}(\boldsymbol{\delta})\mathbf{g}_n(\boldsymbol{\delta}) \\
\hat{\mathbf{S}}(\boldsymbol{\delta}) &= \frac{1}{n}\sum_{t=1}^{n} \mathbf{x}_t\mathbf{x}_t'(y_t - \mathbf{z}_t'\boldsymbol{\delta})^2
\end{aligned}$$

Hansen, Heaton, and Yaron (1996) call $\hat{\boldsymbol{\delta}}(\hat{\mathbf{S}}^{-1}_{\mathsf{CU}})$ the *continuous updating* (CU) *efficient GMM estimator*.

$$n(S_{xy} - S_{xz}\delta)'\left[\frac{1}{n}\sum_{1}^{n} x_t x_t'(y_t - z_t'\delta)^2\right]^{-1}(S_{xy} - S_{xz}\delta)$$

Remarks:

1. The CU estimator is asymptotically equivalent to the two-step and iterated estimators, but may differ in finite samples.

2. The CU estimator does not depend on an initial weight matrix $\mathbf{W}$, and like the iterated efficient GMM estimator, the numerical value of CU estimator is invariant to the scale of the data.

3. The CU estimator is a *non-linear* function of $\boldsymbol{\delta}$ and does not have a closed form expression like the two-step or iterated estimators. It is computationally more burdensome than the iterated estimator, especially for large nonlinear models, and is more prone to numerical instability.

4. Hansen, Heaton, and Yaron found that the finite sample performance of the CU estimator, and test statistics based on it, is often superior to the other estimators.

Remarks continued:

5. Under homoskedastic and normally distributed errors the CU estimator is equivalent to the (limited information) maximum likelihood estimator.

6. More generally, the CU estimator is in the class of *generalized empirical likelihood* (GEL) estimators, which are related to non-parametric maximum likelihood estimators, and have certain optimality properties.

7. The CU estimator is more robust to the presence of weak instruments than the two-step or interated GMM estimators. Here, "robust" loosely means that the asymptotic distribution of the CU estimator does not get as screwed up as the asymptotic distributions of the two-step and iterated GMM estimators when instruments are characterized as "weak". We will more formally characterize weak instruments later on in the course.

## Two-Stage Least Squares as an Efficient GMM Estimator

If the errors are conditionally homoskedastic, then

$$E[\mathbf{x}_t \mathbf{x}_t' \varepsilon_t^2] = \sigma^2 \mathbf{\Sigma}_{xx} = \mathbf{S}$$

A consistent estimate of $\mathbf{S}$ has the form $\hat{\mathbf{S}} = \hat{\sigma}^2 \mathbf{S}_{xx}$, where $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. Typically,

$$\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} (y_t - \mathbf{z}_t' \hat{\boldsymbol{\delta}})^2$$

$$\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$$

The efficient GMM estimator becomes

$$\begin{aligned} \hat{\boldsymbol{\delta}}(\hat{\sigma}^{-2}\mathbf{S}_{xx}^{-1}) &= (\mathbf{S}_{xz}' \hat{\sigma}^{-2} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xz})^{-1} \mathbf{S}_{xz}' \hat{\sigma}^{-2} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \\ &= (\mathbf{S}_{xz}' \mathbf{S}_{xx}^{-1} \mathbf{S}_{xz})^{-1} \mathbf{S}_{xz}' \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \\ &= \hat{\boldsymbol{\delta}}(\mathbf{S}_{xx}^{-1}) \end{aligned}$$

which does not depend on $\hat{\sigma}^2$.

$\hat{S}_{2SLS} = \hat{\sigma}^2 \cdot S_{xx}$ , $\hat{S}_{2SLS}^{-1} = \{ \hat{\sigma}^2 S_{xx} \}^{-1}$ efficient weight matrix under homoskedasticity.

The estimator $\hat{\boldsymbol{\delta}}(\mathbf{S}_{xx}^{-1})$ is, in fact, identical to the *two-stage least squares* (TSLS) estimator of $\boldsymbol{\delta}_0$:

$$\begin{aligned}
\hat{\boldsymbol{\delta}}(\mathbf{S}_{xx}^{-1}) &= (\mathbf{S}_{xz}'\mathbf{S}_{xx}^{-1}\mathbf{S}_{xz})^{-1}\mathbf{S}_{xz}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{xy} \\
&= (\mathbf{Z}'\mathbf{P}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_X\mathbf{y} \\
&= (\hat{\mathbf{Z}}'\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'\mathbf{y} \\
&= \hat{\boldsymbol{\delta}}_{\mathsf{TSLS}}
\end{aligned}$$

$P_x = P_x' , \; P_x \cdot P_x = P_x$

$Z'P_xZ = Z'P_x'P_xZ$

$(P_xZ)'P_xZ$

$= \hat{Z}'\hat{Z}$

where

$$\begin{aligned}
\mathbf{Z} &= n \times L \text{ matrix of observations with tth row } \mathbf{z}_t' \\
\mathbf{X} &= n \times K \text{ matrix of observations with tth row } \mathbf{x}_t' \\
\mathbf{P}_X &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
\hat{\mathbf{Z}} &= \mathbf{P}_X\mathbf{Z}
\end{aligned}$$

$S_{xz} = \frac{1}{n}\sum_1^n x_t z_t' = X'Z\frac{1}{n}$, $\quad s_{xy} = \frac{1}{n}\sum_1^n x_t y_t = X'y$

$S_{xx} = \frac{1}{n}\sum_1^n x_t x_t' = X'X \cdot \frac{1}{n}$

$S^{-1} = \sigma^{-2} \Sigma_{xx}^{-1}$

The asymptotic variance of $\hat{\boldsymbol{\delta}}(\mathbf{S}_{xx}^{-1}) = \hat{\boldsymbol{\delta}}_{\mathsf{TSLS}}$ is

$$\mathsf{avar}(\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}}) = (\boldsymbol{\Sigma}'_{xz}\mathbf{S}^{-1}\boldsymbol{\Sigma}_{xz})^{-1} = \sigma^2(\boldsymbol{\Sigma}'_{xz}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xz})^{-1}$$

Although $\hat{\boldsymbol{\delta}}(\mathbf{S}_{xx}^{-1})$ does not depend on $\hat{\sigma}^2$, a consistent estimate of the asymptotic variance does:

$$\begin{aligned} \widehat{\mathsf{avar}}(\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}}) &= \hat{\sigma}^2_{\mathsf{TSLS}}(\mathbf{S}'_{xz}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xz})^{-1} \\ \hat{\sigma}^2_{\mathsf{TSLS}} &= n^{-1}\sum_{t=1}^{n}(y_t - \mathbf{z}'_t\hat{\boldsymbol{\delta}}_{\mathsf{TSLS}})^2 \end{aligned}$$

$$SE(\hat{\delta}_{i,TSLS}) = \left( \left[ \hat{\sigma}^2_{TSLS} (S'_{xz} S_{xx}^{-1} S_{xz})^{-1} \right]_{ii} \right)^{1/2}$$

## Continuous Updating Efficient GMM under Homoskedasticity: LIML

Under conditional homoskedasticity, $\mathbf{S} = \sigma^2 \mathbf{\Sigma}_{xx}$, and the CU efficient GMM estimator becomes

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}_{\mathsf{CU}}^{-1}) &= \arg\min_{\boldsymbol{\delta}} \; n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}^{-1}(\boldsymbol{\delta})\mathbf{g}_n(\boldsymbol{\delta}) \\
\hat{\mathrm{S}}(\boldsymbol{\delta}) &= \hat{\sigma}^2(\boldsymbol{\delta})\mathbf{S}_{xx} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \mathbf{z}_t'\boldsymbol{\delta})^2 \times \mathbf{S}_{xx}
\end{aligned}
$$

A little algebra shows that

$$
n\mathbf{g}_n(\boldsymbol{\delta})'\hat{\mathrm{S}}^{-1}(\boldsymbol{\delta})\mathbf{g}_n(\boldsymbol{\delta}) = n\frac{(\mathbf{y} - \mathbf{Z}\boldsymbol{\delta})'\mathbf{P}_{\mathbf{X}}(\mathbf{y} - \mathbf{Z}\boldsymbol{\delta})}{(\mathbf{y} - \mathbf{Z}\boldsymbol{\delta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\delta})}
$$

**Result**: $\hat{\boldsymbol{\delta}}(\hat{\mathrm{S}}_{\mathsf{CU}}^{-1}) = \hat{\boldsymbol{\delta}}_{\mathsf{LIML}} =$ *limited information maximum likelihood estimator.*