Serially Correlated Errors (Time Series Data):

• Suppose that $u_1, u_2, \cdots, u_n$ are serially correlated.

Consider the case where the subscript represents time.

Remember that $\beta_{GMM} \sim N\Big(\beta, \ \sigma^2(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}\Big)$,

We need to consider evaluation of $\sigma^2 Z'\Omega Z = V(u^*)$, i.e.,

$$
\begin{aligned}
V(u^*) &= V(Z'u) = V(\sum_{i=1}^{n} z_i' u_i) = V(\sum_{i=1}^{n} v_i) \\
&= E\Big((\sum_{i=1}^{n} v_i)(\sum_{i=1}^{n} v_i)'\Big) = E\Big((\sum_{i=1}^{n} v_i)(\sum_{j=1}^{n} v_j)'\Big) \\
&= E(\sum_{i=1}^{n} \sum_{j=1}^{n} v_i v_j') = \sum_{i=1}^{n} \sum_{j=1}^{n} E(v_i v_j')
\end{aligned}
$$

where $v_i = z_i' u_i$ is a $r \times 1$ vector.

Define $\Gamma_\tau = \mathrm{E}(v_i v'_{i-\tau})$.

$\Gamma_0 = \mathrm{E}(v_i v'_i)$ represents the $r \times r$ variance-covariance matrix of v_i.

$\Gamma_{-\tau} = \mathrm{E}(v_{i-\tau} v'_i) = \mathrm{E}\big((v_i v'_{i-\tau})'\big) = \big(\mathrm{E}(v_i v'_{i-\tau})\big)' = \Gamma'_\tau.$

$$\mathrm{V}(u^*) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}(v_i v'_j)$$

$$= \mathrm{E}(v_1 v'_1) + \mathrm{E}(v_1 v'_2) + \mathrm{E}(v_1 v'_3) + \cdots + \mathrm{E}(v_1 v'_n)$$

$$+ \mathrm{E}(v_2 v'_1) + \mathrm{E}(v_2 v'_2) + \mathrm{E}(v_2 v'_3) + \cdots + \mathrm{E}(v_2 v'_n)$$

$$+ \mathrm{E}(v_3 v'_1) + \mathrm{E}(v_3 v'_2) + \mathrm{E}(v_3 v'_3) + \cdots + \mathrm{E}(v_3 v'_n)$$

$$\vdots$$

$$+ \mathrm{E}(v_n v'_1) + \mathrm{E}(v_n v'_2) + \mathrm{E}(v_n v'_3) + \cdots + \mathrm{E}(v_n v'_n)$$

$$= \Gamma_0 + \Gamma_{-1} + \Gamma_{-2} + \cdots + \Gamma_{1-n}$$

$$+ \Gamma_1 + \Gamma_0 + \Gamma_{-1} + \cdots + \Gamma_{2-n}$$

$$+ \Gamma_2 + \Gamma_1 + \Gamma_0 + \cdots + \Gamma_{3-n}$$

$$\vdots$$

$$+ \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0$$

$$= \Gamma_0 + \Gamma_1' + \Gamma_2' + \cdots + \Gamma_{n-1}'$$

$$+ \Gamma_1 + \Gamma_0 + \Gamma_1' + \cdots + \Gamma_{n-2}'$$

$$+ \Gamma_2 + \Gamma_1 + \Gamma_0 + \cdots + \Gamma_{n-3}'$$

$$\vdots$$

$$+ \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0$$

$$= n\Gamma_0 + (n-1)(\Gamma_1 + \Gamma_1') + (n-2)(\Gamma_2 + \Gamma_2') + \cdots + (\Gamma_{n-1} + \Gamma_{n-1}')$$

$$= n\Gamma_0 + \sum_{i=1}^{n-1}(n-i)(\Gamma_i + \Gamma_i')$$

$$= n\left(\Gamma_0 + \sum_{i=1}^{n-1}(1 - \frac{i}{n})(\Gamma_i + \Gamma_i')\right)$$

$$\approx n\left(\Gamma_0 + \sum_{i=1}^{q}(1 - \frac{i}{q+1})(\Gamma_i + \Gamma_i')\right).$$

242

In the last line, $n - 1$ is replaced by q, where $q < n - 1$.

We need to estimate Γ_τ as: $\quad \hat{\Gamma}_\tau = \dfrac{1}{n} \displaystyle\sum_{i=\tau+1}^{n} \hat{v}_i \hat{v}'_{i-\tau}, \ $ where $\hat{v}_i = z'_i \hat{u}_i$ for $\hat{u}_i = y_i - x_i \beta_{GMM}$.

As τ is large, $\hat{\Gamma}_\tau$ is unstable.

Therefore, we choose the q which is less than $n - 1$.

Hansen's J Test: Is the model specification correct?

That is, is $E(z'u) = 0$ for $y = x\beta + u$ correct?

$H_0 :$ $E(z'u) = 0$ (The model is correct. Or, the instrumental variables are appropriate.)

$H_1 :$ $E(z'u) \neq 0$

The number of equations is r, while the number of parameters is k.

The degree of freedom is $r - k$.

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i'\hat{u}_i\right)'\left(V(\widehat{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i\hat{u}_i})\right)^{-1}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i'\hat{u}_i\right) \longrightarrow \chi(r - k),$$

where $\hat{u}_i = y_i - x_i\beta_{GMM}$.

$V(\widehat{\frac{1}{n} \sum_{i=1}^{n} z_i'\hat{u}_i})$ indicates the estimate of $V(\frac{1}{n} \sum_{i=1}^{n} z_i'u_i)$ for $u_i = y_i - x_i\beta$.

The J test is called a test for over-identifying restrictions (   ).

Remark 1: $X_1, X_2, \cdots, X_n$ are mutually independent.

$X_i \sim N(\mu, \sigma^2)$ are assumed.

244

Consider $\overline{X} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$.

Then, $\dfrac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1)$.

That is, $\sqrt{n}(\overline{X} - \mu) \longrightarrow N(0, \sigma^2)$.

Remark 2: $X_1, X_2, \cdots, X_n$ are mutually independent.

$X_i \sim N(\mu, \sigma^2)$ are assumed.

Then, $\left(\dfrac{X_i - \mu}{\sigma^2}\right)^2 \sim \chi^2(1)$ and $\displaystyle\sum_{i=1}^{n}\left(\dfrac{X_i - \mu}{\sigma^2}\right)^2 \sim \chi^2(n)$.

If μ is replaced by its estimator $\overline{X}$, then $\displaystyle\sum_{i=1}^{n}\left(\dfrac{X_i - \overline{X}}{\sigma^2}\right)^2 \sim \chi^2(n-1)$.

Note:

$$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma^2}\right)^2 = \begin{pmatrix} X_i - \overline{X} \\ X_i - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}' \begin{pmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} X_i - \overline{X} \\ X_i - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix} \sim \chi^2(n-1)$$

In the case of GMM,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' u_i \longrightarrow N(0, \Sigma),$$

where $\Sigma = V\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' u_i\right)$.

Therefore, we obtain: $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' u_i\right)' \Sigma^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' u_i\right) \longrightarrow \chi^2(r)$.

In order to obtain $\hat{u}_i$, we have to estimate β, which is a $k \times 1$ vector.

Therefore, replacing u_i by $\hat{u}_i$, we have: $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' \hat{u}_i\right)' \Sigma^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' \hat{u}_i\right) \longrightarrow \chi^2(r - k)$.

Moreover, from $\hat{\Sigma} \longrightarrow \Sigma$, we obtain: $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' \hat{u}_i\right)' \hat{\Sigma}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i' \hat{u}_i\right) \longrightarrow \chi^2(r - k)$, where $\hat{\Sigma}$ is a consistent estimator of Σ.

10.3 Generalized Method of Moments (GMM,   ) II — Nonlinear Case —

Consider the general case:

$$E(h(\theta; w)) = 0,$$

which is the orthogonality condition.

A $k \times 1$ vector θ denotes a parameter to be estimated.

$h(\theta; w)$ is a $r \times 1$ vector for $r \geq k$.

Let $w_i = (y_i, x_i)$ be the ith observed data, i.e., the ith realization of w.

Define $g(\theta; W)$ as:

$$g(\theta; W) = \frac{1}{n} \sum_{i=1}^{n} h(\theta; w_i),$$

where $W = \{w_n, w_{n-1}, \cdots, w_1\}$.

$g(\theta; W)$ is a $r \times 1$ vector for $r \geq k$.

Let $\hat{\theta}$ be the GMM estimator which minimizes:

$$g(\theta; W)'S^{-1}g(\theta; W),$$

with respect to θ.

- Solve the following first-order condition:

$$\frac{\partial g(\theta; W)'}{\partial \theta}S^{-1}g(\theta; W) = 0,$$

with respect to θ. There are r equations and k parameters.

Computational Procedure:

Linearizing the first-order condition around $\theta = \hat{\theta}$,

$$\begin{aligned}
0 &= \frac{\partial g(\theta; W)'}{\partial \theta}S^{-1}g(\theta; W) \\
&\approx \frac{\partial g(\hat{\theta}; W)'}{\partial \theta}S^{-1}g(\hat{\theta}; W) + \frac{\partial g(\hat{\theta}; W)'}{\partial \theta}S^{-1}\frac{\partial g(\hat{\theta}; W)}{\partial \theta'}(\theta - \hat{\theta}) \\
&= \hat{D}'S^{-1}g(\hat{\theta}; W) + \hat{D}'S^{-1}\hat{D}(\theta - \hat{\theta}),
\end{aligned}$$

where $\hat{D} = \dfrac{\partial g(\hat{\theta}; W)}{\partial \theta'}$, which is a $r \times k$ matrix.

248

Note that in the second term of the second line the second derivative is ignored and omitted.

Rewriting, we have the following equation:

$$\theta - \hat{\theta} = -(\hat{D}'S^{-1}\hat{D})^{-1}\hat{D}'S^{-1}g(\hat{\theta}; W).$$

Replacing θ and $\hat{\theta}$ by $\hat{\theta}^{(i+1)}$ and $\hat{\theta}^{(i)}$, respectively, we obtain:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - (\hat{D}^{(i)\prime}S^{-1}\hat{D}^{(i)})^{-1}\hat{D}^{(i)\prime}S^{-1}g(\hat{\theta}^{(i)}; W),$$

where $\hat{D}^{(i)} = \dfrac{\partial g(\hat{\theta}^{(i)}; W)}{\partial \theta'}$.

Given S, repeat the iterative procedure for $i = 1, 2, 3, \cdots$, until $\hat{\theta}^{(i+1)}$ is equal to $\hat{\theta}^{(i)}$.

How do we derive the weight matrix S?

- In the case where $h(\theta; w_i)$, $i = 1, 2, \cdots, n$, are mutually independent, S is:

$$\begin{aligned}
S &= V\Big(\sqrt{n}g(\theta; W)\Big) = n E\Big(g(\theta; W)g(\theta; W)'\Big) \\
&= n E\Big(\Big(\frac{1}{n} \sum_{i=1}^{n} h(\theta; w_i)\Big)\Big(\frac{1}{n} \sum_{j=1}^{n} h(\theta; w_j)\Big)'\Big) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\Big(h(\theta; w_i)h(\theta; w_j)'\Big) \\
&= \frac{1}{n} \sum_{i=1}^{n} E\Big(h(\theta; w_i)h(\theta; w_i)'\Big),
\end{aligned}$$

which is a $r \times r$ matrix.

Note that

(i) $E\Big(h(\theta; w_i)\Big) = 0$ for all i and accordingly $E\Big(g(\theta; W)\Big) = 0$,

(ii) $g(\theta; W) = \dfrac{1}{n} \sum_{i=1}^{n} h(\theta; w_i) = \dfrac{1}{n} \sum_{j=1}^{n} h(\theta; w_j)$,

(iii) $E\Big(h(\theta; w_i)h(\theta; w_j)'\Big) = 0$ for $i \neq j$.

The estimator of S, denoted by $\hat{S}$ is given by: $\quad \hat{S} = \dfrac{1}{n} \sum_{i=1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_i)' \longrightarrow S$.

250

- Taking into account serial correlation of $h(\theta; w_i)$, $i = 1, 2, \cdots, n$, S is given by:

$$\begin{aligned} S &= V\left(\sqrt{n}g(\theta; W)\right) = nE\left(g(\theta; W)g(\theta; W)'\right) \\ &= nE\left(\left(\frac{1}{n}\sum_{i=1}^{n} h(\theta; w_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n} h(\theta; w_j)\right)'\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(h(\theta; w_i)h(\theta; w_j)'\right). \end{aligned}$$

Note that $E\left(\sum_{i=1}^{n} h(\theta; w_i)\right) = 0$.

Define $\Gamma_\tau = E\left(h(\theta; w_i)h(\theta; w_{i-\tau})'\right) < \infty$, i.e., $h(\theta; w_i)$ is stationary.

Stationarity:

(i) $E\left(h(\theta; w_i)\right)$ does not depend on i,

(ii) $E\left(h(\theta; w_i)h(\theta; w_{i-\tau})'\right)$ depends on time difference τ.

$$\implies E\left(h(\theta; w_i)h(\theta; w_{i-\tau})'\right) = \Gamma_\tau$$

$$S = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(h(\theta; w_i)h(\theta; w_j)'\right)$$

251

$$
\begin{aligned}
&= \frac{1}{n}\Big(\mathrm{E}\big(h(\theta;w_1)h(\theta;w_1)'\big) + \mathrm{E}\big(h(\theta;w_1)h(\theta;w_2)'\big) + \cdots + \mathrm{E}\big(h(\theta;w_1)h(\theta;w_n)'\big) \\
&\qquad\quad \mathrm{E}\big(h(\theta;w_2)h(\theta;w_1)'\big) + \mathrm{E}\big(h(\theta;w_2)h(\theta;w_2)'\big) + \cdots + \mathrm{E}\big(h(\theta;w_2)h(\theta;w_n)'\big) \\
&\qquad\qquad\qquad \vdots \\
&\qquad\quad \mathrm{E}\big(h(\theta;w_n)h(\theta;w_1)'\big) + \mathrm{E}\big(h(\theta;w_n)h(\theta;w_2)'\big) + \cdots + \mathrm{E}\big(h(\theta;w_n)h(\theta;w_n)'\big)\Big) \\
&= \frac{1}{n}(\Gamma_0 \quad + \Gamma_1' \quad + \Gamma_2' \quad + \cdots + \Gamma_{n-1}' \\
&\qquad\ \ \Gamma_1 \quad + \Gamma_0 \quad + \Gamma_1' \quad + \cdots + \Gamma_{n-2}' \\
&\qquad\qquad \vdots \\
&\qquad\ \ \Gamma_{n-1} + \Gamma_{n-2} + \Gamma_{n-3} + \cdots + \Gamma_0) \\
&= \frac{1}{n}\Big(n\Gamma_0 + (n-1)(\Gamma_1 + \Gamma_1') + (n-2)(\Gamma_2 + \Gamma_2') + \cdots + (\Gamma_{n-1} + \Gamma_{n-1}')\Big) \\
&= \Gamma_0 + \sum_{i=1}^{n-1} \frac{n-i}{n}(\Gamma_i + \Gamma_i') = \Gamma_0 + \sum_{i=1}^{n-1}\Big(1 - \frac{i}{n}\Big)(\Gamma_i + \Gamma_i') \\
&= \Gamma_0 + \sum_{i=1}^{q}\Big(1 - \frac{i}{q+1}\Big)(\Gamma_i + \Gamma_i').
\end{aligned}
$$

Note that $\Gamma'_\tau = E\Big(h(\theta; w_{i-\tau})h(\theta; w_i)'\Big) = \Gamma(-\tau)$, because $\Gamma_\tau = E\Big(h(\theta; w_i)h(\theta; w_{i-\tau})'\Big)$.

In the last line, n is replaced by $q + 1$, where $q < n$.

We need to estimate Γ_τ as: $\quad \hat{\Gamma}_\tau = \dfrac{1}{n} \sum_{i=\tau+1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_{i-\tau})'$.

As τ is large, $\hat{\Gamma}_\tau$ is unstable.

Therefore, we choose the q which is less than n.

S is estimatated as:

$$\hat{S} = \hat{\Gamma}_0 + \sum_{i=1}^{q}\left(1 - \frac{i}{q + 1}\right)(\hat{\Gamma}_i + \hat{\Gamma}_i'),$$

$\implies$ the Newey-West Estimator

Note that $\hat{S} \longrightarrow S$, because $\hat{\Gamma}_\tau \longrightarrow \Gamma_\tau$ as $n \longrightarrow \infty$.

Asymptotic Properties of GMM:

GMM is consistent and asymptotic normal as follows:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left(0, (D'S^{-1}D)^{-1}\right),$$

where D is a $r \times k$ matrix, and $\hat{D}$ is an estimator of D, defined as:

$$D = \frac{\partial g(\theta; W)}{\partial \theta'}, \qquad \hat{D} = \frac{\partial g(\hat{\theta}; W)}{\partial \theta'}.$$

Proof of Asymptotic Normality:

Assumption 1: $\hat{\theta} \longrightarrow \theta$

Assumption 2: $\sqrt{n}g(\theta; W) \longrightarrow N(0, S)$, i.e., $S = \lim_{n \to \infty} V\left(\sqrt{n}g(\theta; W)\right)$.

The first-order condition of GMM is:

$$\frac{\partial g(\theta; W)'}{\partial \theta} S^{-1} g(\theta; W) = 0.$$

The GMM estimator, denote by $\hat{\theta}$, satisfies the above equation.

Therefore, we have the following:

$$\frac{\partial g(\hat{\theta}; W)'}{\partial \theta} \hat{S}^{-1} g(\hat{\theta}; W) = 0.$$

Linearize $g(\hat{\theta}; W)$ around $\hat{\theta} = \theta$ as follows:

$$g(\hat{\theta}; W) = g(\theta; W) + \frac{\partial g(\overline{\theta}; W)}{\partial \theta'}(\hat{\theta} - \theta) = g(\theta; W) + \overline{D}(\hat{\theta} - \theta),$$

where $\overline{D} = \dfrac{\partial g(\overline{\theta}; W)}{\partial \theta'}$, and $\overline{\theta}$ is between $\hat{\theta}$ and θ.

⟹ **Theorem of Mean Value (   )**

Substituting the linear approximation at $\hat{\theta} = \theta$, we obtain:

$$\begin{aligned}
0 &= \hat{D}'\hat{S}^{-1}g(\hat{\theta}; W) \\
&= \hat{D}'\hat{S}^{-1}\Big(g(\theta; W) + \overline{D}(\hat{\theta} - \theta)\Big) \\
&= \hat{D}'\hat{S}^{-1}g(\theta; W) + \hat{D}'\hat{S}^{-1}\overline{D}(\hat{\theta} - \theta),
\end{aligned}$$

which can be rewritten as:

$$\hat{\theta} - \theta = -(\hat{D}'\hat{S}^{-1}\overline{D})^{-1}\hat{D}'\hat{S}^{-1}g(\theta; W).$$

Note that $\overline{D} = \dfrac{\partial g(\overline{\theta}; W)}{\partial \theta'}$, where $\overline{\theta}$ is between $\hat{\theta}$ and θ.

From Assumption 1, $\hat{\theta} \longrightarrow \theta$ implies $\overline{\theta} \longrightarrow \theta$

256

Therefore,

$$\sqrt{n}(\hat{\theta} - \theta) = -(\hat{D}'\hat{S}^{-1}\overline{D})^{-1}\hat{D}'S^{-1} \times \sqrt{n}g(\theta; W).$$

Accordingly , the GMM estimator $\hat{\theta}$ has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left(0, (D'S^{-1}D)^{-1}\right).$$

Note that $\hat{D} \longrightarrow D$, $\overline{D} \longrightarrow D$, $\hat{S} \longrightarrow S$ and Assumption 2 are utilized.

Computational Procedure:

(1) Compute $\hat{S}^{(i)} = \hat{\Gamma}_0 + \sum_{i=1}^{q} \left(1 - \dfrac{i}{q+1}\right)(\hat{\Gamma}_i + \hat{\Gamma}'_i)$, where $\hat{\Gamma}_\tau = \dfrac{1}{n} \sum_{i=\tau+1}^{n} h(\hat{\theta}; w_i)h(\hat{\theta}; w_{i-\tau})'$. q is set by a researcher.

(2) Use the following iterative procedure:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - (\hat{D}^{(i)\prime}\hat{S}^{(i)-1}\hat{D}^{(i)})^{-1}\hat{D}^{(i)\prime}\hat{S}^{(i)-1}g(\hat{\theta}^{(i)}; W).$$

(3) Repeat (1) and (2) until $\hat{\theta}^{(i+1)}$ is equal to $\hat{\theta}^{(i)}$.

In (2), remember that when S is given we take the following iterative procedure:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - (\hat{D}^{(i)\prime}S^{-1}\hat{D}^{(i)})^{-1}\hat{D}^{(i)\prime}S^{-1}g(\hat{\theta}^{(i)}; W),$$

where $\hat{D}^{(i)} = \dfrac{\partial g(\hat{\theta}^{(i)}; W)}{\partial \theta'}$. S is replaced by $\hat{S}^{(i)}$.

258

- If the assumption $E\Big(h(\theta; w)\Big) = 0$ is violated, the GMM estimator $\hat{\theta}$ is no longer consistent.

Therefore, we need to check if $E\Big(h(\theta; w)\Big) = 0$.

From Assumption 2, note as follows:

$$J = \Big(\sqrt{n}g(\hat{\theta}; W)\Big)' \hat{S}^{-1}\Big(\sqrt{n}g(\hat{\theta}; W)\Big) \longrightarrow \chi^2(r - k),$$

which is called Hansen's J test.

Because of r equations and k parameters, the degree of freedom is given by $r - k$.

If J is small enough, we can judge that the specified model is correct.

259

Testing Hypothesis:

Remember that the GMM estimator $\hat{\theta}$ has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\Big(0, (D'S^{-1}D)^{-1}\Big).$$

Consider testing the following null and alternative hypotheses:

- The null hypothesis: $\quad H_0 : R(\theta) = 0$,

- The alternative hypothesis: $\quad H_1 : R(\theta) \neq 0$,

where $R(\theta)$ is a $p \times 1$ vector function for $p \leq k$.

p denotes the number of restrictions.

$R(\theta)$ is linearized as: $R(\hat{\theta}) = R(\theta) + R_{\bar{\theta}}(\hat{\theta} - \theta)$, where $R_{\bar{\theta}} = \dfrac{\partial R(\bar{\theta})}{\partial \theta'}$, which is a $p \times k$ matrix.

Note that $\bar{\theta}$ is bewteen $\hat{\theta}$ and θ.   If $\hat{\theta} \longrightarrow \theta$, then $\bar{\theta} \longrightarrow \theta$ and $R_{\bar{\theta}} \longrightarrow R_\theta$.

Under the null hypothesis $R(\theta) = 0$, we have $R(\hat{\theta}) = R_{\bar{\theta}}(\hat{\theta} - \theta)$, which implies that the distribution of $R(\hat{\theta})$ is equivalent to that of $R_{\bar{\theta}}(\hat{\theta} - \theta)$.

The distribution of $\sqrt{n}R(\hat{\theta})$ is given by:

$$\sqrt{n}R(\hat{\theta}) = \sqrt{n}R_{\bar{\theta}}(\hat{\theta} - \theta) \longrightarrow N\Big(0, R_\theta(D'S^{-1}D)^{-1}R_\theta'\Big).$$

Therefore, under the null hypothesis, we have the following distribution:

$$nR(\hat{\theta})\Big(R_\theta(D'S^{-1}D)^{-1}R_\theta'\Big)^{-1}R(\hat{\theta})' \longrightarrow \chi^2(p).$$

Practically, replacing θ by $\hat{\theta}$ in R_θ, D and S, we use the following test statistic:

$$nR(\hat{\theta})\Big(R_{\hat{\theta}}(\hat{D}'\hat{S}^{-1}\hat{D})^{-1}R_{\hat{\theta}}'\Big)^{-1}R(\hat{\theta})' \longrightarrow \chi^2(p).$$

$\implies$ Wald type test

Examples of $h(\theta; w)$:

1. **OLS:**

   Regression Model: $\quad y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ is taken as:

   $$h(\theta; w_i) = x_i'(y_i - x_i\beta).$$

2. **IV (Instrumental Variable,   ):**

   Regression Model: $\quad y_i = x_i\beta + \epsilon_i, \quad \mathrm{E}(x_i'\epsilon_i) \neq 0, \quad \mathrm{E}(z_i'\epsilon_i) = 0$

   $h(\theta; w_i)$ is taken as:

   $$h(\theta; w_i) = z_i'(y_i - x_i\beta),$$

   where z_i is a vector of instrumental variables.

   When z_i is a $1 \times k$ vector, the GMM of β is equivalent to the instrumental variable (IV) estimator.

   When z_i is a $1 \times r$ vector for $r > k$, the GMM of β is equivalent to the two-stage least squares (2SLS) estimator.

3. **NLS (Nonlinear Least Squares,   )**:

Regression Model: $f(y_i, x_i, \beta) = \epsilon_i, \quad E(x_i'\epsilon_i) \neq 0, \quad E(z_i'\epsilon_i) = 0$

$h(\theta; w_i)$ is taken as:

$$h(\theta; w_i) = z_i' f(y_i, x_i, \beta)$$

where z_i is a vector of instrumental variables.

Example: Demand function using STATA

```
year
y
q1
p1                                    CPI      CPI
p2                                     CPI       CPI
p3                                    CPI      CPI

year    y      q1      p1       p2       p3
2000 567865 7087.0 1.043390 0.884965 0.818365
2001 561722 6993.1 1.032520 0.886179 0.822154
2002 553768 6934.4 1.031800 0.891282 0.834872
2003 539928 6816.8 1.050410 0.876543 0.843621
2004 547006 6651.6 1.089510 0.865226 0.868313
2005 541367 6615.8 1.020640 0.862745 0.887513
2006 540863 6523.7 1.000000 0.878601 0.891975
2007 543994 6680.5 0.994856 0.886831 0.908436
2008 541821 6494.7 1.043610 0.894523 0.932049
2009 533154 6477.3 1.066870 0.898148 0.934156
2010 539577 6458.2 1.040420 0.889119 0.924352
2011 529750 6448.4 1.025960 0.894081 0.925234
2012 538988 6377.6 1.057170 0.904366 0.917879
2013 542018 6360.7 1.047620 0.909938 0.916149
2014 523953 6174.6 1.016130 0.971774 0.960685
2015 525669 6268.0 1.000000 1.000000 1.000000
2016 527501 6244.8 1.018020 1.019020 1.017020
```

```
2017 531693 6106.6 1.027890 1.066730 1.025900


. tsset year
        time variable:  year, 2000 to 2017
                delta:  1 unit

. reg q1 y p1 p2 p3 if year>2000.5

      Source |       SS           df       MS      Number of obs   =        17
-------------+----------------------------------   F(4, 12)        =     25.83
       Model |  913640.443         4  228410.111   Prob > F        =    0.0000
    Residual |  106100.077        12  8841.67308   R-squared       =    0.8960
-------------+----------------------------------   Adj R-squared   =    0.8613
       Total |  1019740.52        16  63733.7825   Root MSE        =     94.03

------------------------------------------------------------------------------
         q1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          y |   .0067843   .0045443     1.49   0.161    -.003117    .0166856
         p1 |  -1128.834   998.7698    -1.13   0.280   -3304.966    1047.299
         p2 |   356.8095   806.2301     0.44   0.666   -1399.815    2113.434
         p3 |  -3442.221   1130.078    -3.05   0.010   -5904.448   -979.9931
      _cons |   6850.563   3179.316     2.15   0.052   -76.57278     13777.7
------------------------------------------------------------------------------

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(y p1
> p2 p3)
```

265

```
Step 1
Iteration 0:   GMM criterion Q(b) =   42400764
Iteration 1:   GMM criterion Q(b) =  6.781e-12
Iteration 2:   GMM criterion Q(b) =  6.781e-12  (backed up)

Step 2
Iteration 0:   GMM criterion Q(b) =  1.966e-15
Iteration 1:   GMM criterion Q(b) =  1.963e-15  (backed up)
convergence not achieved
    The Gauss-Newton stopping criterion has been met but missing standard errors
    indicate some of the parameters are not identified.

GMM estimation

Number of parameters =    5
Number of moments    =    5
Initial weight matrix: Unadjusted              Number of obs   =         17
GMM weight matrix:     Robust


------------------------------------------------------------------------------
             |                Robust
             |     Coef.    Std. Err.     z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   6850.563   17645.71     0.39   0.698     -27734.4    41435.53
         /b1 |   .0067843   .0282325     0.24   0.810    -.0485504     .062119
         /b2 |  -1128.834   1057.915    -1.07   0.286    -3202.309    944.6415
         /b3 |   356.8095   1565.86      0.23   0.820    -2712.219    3425.838
         /b4 |  -3442.221   5085.561    -0.68   0.498    -13409.74    6525.296
------------------------------------------------------------------------------
```

```
Instruments for equation 1: y p1 p2 p3 _cons
Warning: convergence not achieved

. gmm (q1-{b0}-{b1}*y-{b2}*p1-{b3}*p2-{b4}*p3) if year>2000.5, instruments(p1 p2
> p3 l.p1 l.p2 l.p3)

Step 1
Iteration 0:   GMM criterion Q(b) =   42404066
Iteration 1:   GMM criterion Q(b) = 2790.3146
Iteration 2:   GMM criterion Q(b) = 2790.3146

Step 2
Iteration 0:   GMM criterion Q(b) =   .3201826
Iteration 1:   GMM criterion Q(b) =   .2469289
Iteration 2:   GMM criterion Q(b) =   .2469289

GMM estimation

Number of parameters =   5
Number of moments    =   7
Initial weight matrix: Unadjusted              Number of obs   =        17
GMM weight matrix:    Robust

--------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
         /b0 |  -1192.466   4669.012    -0.26   0.798    -10343.56     7958.63
         /b1 |   .0186312   .0067682     2.75   0.006     .0053657    .0318967
```

```
         /b2 |   -1016.864      780.979     -1.30    0.193    -2547.554     513.8271
         /b3 |   -905.5585     598.0885     -1.51    0.130     -2077.79     266.6734
         /b4 |   -499.8064     1147.985     -0.44    0.663    -2749.815     1750.202
--------------------------------------------------------------------------------
Instruments for equation 1: p1 p2 p3 L.p1 L.p2 L.p3 _cons

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(2) = 4.19779 (p = 0.1226)
```