Chap.9 Generalized Method of Moments

(Chap.10 P201-P206; Chap.18 P525-P556, Greene, 5th)

9.1 Introduction

9.1.1 Classical Method of Moments (CMM)

Suppose $f(X,\theta)$ is the pdf of a univariate random variable X, the CMM estimator of θ can be obtained from following steps:

i) Compute population moments $E\left(X_i^K\right)$ under the model density $f(x,\theta)$, for example

$$E(X_i)=\int Xf(x,\theta)dX=\mu$$
$$E\left(X_i^2\right)=\int X^2 f(x,\theta)dX=\sigma^2+\mu^2$$

ii) Compute the sample moments from random sample $(X_1,\cdots,X_n)$, for example

$$\bar{m}_1=\frac{1}{n}\sum_{i=1}^{n}X_i$$
$$\bar{m}_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$$

iii) Match the sample moments and population moments.

$$\begin{cases}\bar{m}_1=\hat{\mu}\\ \bar{m}_2=\hat{\sigma}^2+\hat{\mu}^2\end{cases}$$

iv) Solve for the equations. The solution $\left(\hat{\mu},\hat{\sigma}^2\right)'$ is called the method of moment estimator for $\theta=\left(\mu,\sigma^2\right)'$.

In general, if θ is a $K\times 1$ vector, and we have the kth order population moment function

$$E\left[g\left(X_i^k\right)\right]=\gamma_k(\theta_1,\cdots,\theta_K)\quad(k=1,\cdots,K)\tag{9.1}$$

And its sample counterpart

$$\frac{1}{n}\sum_{i=1}^{n}g\left(X_i^k\right)=\bar{g}_k(\theta_1,\cdots,\theta_K)\tag{9.2}$$

Then the method of moment estimator for θ can be obtained by solving for the system of equations:

$$
\begin{cases} \bar{g}_1 = \gamma_1\left(\theta_1, \cdots, \theta_K\right) \\ \qquad \vdots \\ \bar{g}_K = \gamma_K\left(\theta_1, \cdots, \theta_K\right) \end{cases} \tag{9.3}
$$

i.e. $\hat{\theta}_k = \hat{\theta}_k\left(\bar{g}_1, \cdots, \bar{g}_K\right)$

9.1.2 The Statistical Properties of CMM Estimator

According to Theorem D.4, the corollary to Theorem D.4 and Slutsky Theorem, we have

$$
p\lim \bar{g}_k = p\lim \frac{1}{n}\sum_{i=1}^{n} g\left(X_i^k\right) = E\left[g\left(X_i^k\right)\right]
$$

$$
p\lim \hat{\theta}_k = \theta_k \quad \left(k = 1, \cdots, K\right)
$$

The CMM estimators are consistent, but in most cases, they are not efficient. The exception is in random sampling form exponential families of distributions. (See P529, Greene 5^{th})

9.2 Generalized Method of Moments Estimator

9.2.1 Why we need GMM estimation?

Example 1: The pdf of Gamma Distribution

$$
f\left(x; p, \lambda\right) = \frac{\lambda^p}{\Gamma\left(p\right)} e^{-\lambda x} x^{p-1}, \quad x > 0, \lambda > 0, p > 0 \tag{9.4}
$$

It can be shown that

$$
E\left(X^k\right) = \frac{\Gamma\left(p+k\right)}{\Gamma\left(p\right)} \frac{1}{\lambda^k} = \frac{\prod_{j=1}^{k}\left(p+k-j\right)}{\lambda^k} = \frac{\left(p+k-1\right)\left(p+k-2\right)\cdots p}{\lambda^k}
$$

$$
E\left(X^{-k}\right) = \frac{\lambda^k}{\prod_{j=1}^{k}\left(p-j\right)} = \frac{\lambda^k}{\left(p-1\right)\cdots\left(p-k\right)}
$$

(k is a positive integer)

So we have

$$
E\left(X\right) = \frac{p}{\lambda}
$$

$$
E\left(X^2\right) = \frac{p\left(p+1\right)}{\lambda^2}
$$

$$
E\left(\frac{1}{X}\right) = \frac{\lambda}{p-1}
$$

The log-likelihood function for (9.4) is

$$
\ln L\left(X; p, \lambda\right) = n\left[p\ln\left(\lambda\right) - \ln\left(\Gamma\left(p\right)\right)\right] - \lambda\sum_{i=1}^{n} X_i + \left(p-1\right)\sum_{i=1}^{n}\ln\left(X_i\right)
$$

$$\Rightarrow \begin{cases} \dfrac{\partial \ln L}{\partial p} = n\ln\lambda - n\dfrac{d\ln\left(\Gamma\left(p\right)\right)}{dp} + \sum_{i=1}^{n}\ln\left(X_i\right) = 0 \\ \dfrac{\partial \ln L}{\partial \lambda} = \dfrac{1}{\lambda}np - \sum_{i=1}^{n} X_i = 0 \end{cases}$$

$$\Rightarrow E\left[\ln\left(X_i\right)\right] = \Psi\left(p\right) - \ln\lambda$$

Where $\Psi\left(p\right) = \dfrac{d\ln\left(\Gamma\left(p\right)\right)}{dp}$.

So far, with 2 parameters to be estimated, we have 4 moment equations:

$$\begin{cases} \dfrac{1}{n}\sum X_i = \dfrac{p}{\lambda} \\ \dfrac{1}{n}\sum X_i^2 = \dfrac{p\left(p+1\right)}{\lambda^2} \\ \dfrac{1}{n}\sum \ln X_i = \Psi\left(p\right) - \ln\lambda \\ \dfrac{1}{n}\sum \dfrac{1}{X_i} = \dfrac{\lambda}{p-1} \end{cases} \tag{9.5}$$

How can we deal with this system of equations?

More important, in Econometrics, we need to solve the equation system that there are more orthogonal conditions than parameters (L>K).

Example 2: Hansen and Singleton's (1982) Asset Pricing Model

Suppose a representative agent has a constant relative risk aversion utility over his lifetime:

$$U = \sum_{t=0}^{T} \delta^t \frac{c_t^r - 1}{r}$$

Where $\delta > 0$ is the agent's time discount factorr, $r \geq 0$ is the risk aversion parameter and c_t is consumption during period t. Let the information available to the agent at time t be represented by the sigma-algebra I_t, and let $R_t = \dfrac{P_t}{P_{t-1}} = 1 + \dfrac{P_t - P_{t-1}}{P_{t-1}}$ be the gross return to an asset acquired at time t-1 at the price of P_{t-1}. The agent's optimization problem is to

$$\max_{\{c_t\}} E\left(U\right)$$

Subject to the budget constraint

$$c_t + p_t q_t = w_t + p_t q_{t-1}$$

Where q_t is the quantity of the asset purchased at time t and w_t is the agent's income at period t. The marginal rate of inter-temporal substitution is

$$MRS_t(\theta) = \frac{\partial U / \partial c_t}{\partial U / \partial c_{t-1}} = \delta \left(\frac{c_t}{c_{t-1}} \right)^{r-1}$$

Where $\theta = (\delta, r)'$. The first order conditions of the agent optimization problem are

$$E\left[MRS_t(\theta) R_t \middle| I_{t-1} \right] = 1$$

That is, the marginal rate of inter-temporal substitution discounts gross returns to unity.

How to estimate the unknown parameter θ in an asset pricing model? More generally, how to estimate θ from any linear or nonlinear econometric model which can be formulated as a set of moment conditions? From the Euler equation, we can induce the following conditional moment restrictions:

$$E\left[MRS_t(\theta) R_t - 1 \right] = 0$$

$$E\left\{ \frac{c_{t-1}}{c_{t-2}} \left[MRS_t(\theta) R_t - 1 \right] \right\} = 0$$

$$E\left\{ R_{t-1} \left[MRS_t(\theta) R_t - 1 \right] \right\} = 0$$

Therefore, we can consider the 3×1 sample moments.

$$\bar{m}(\theta) = \frac{1}{T} \sum_{t=1}^{T} m_t(\theta)$$

Where $m_t(\theta) = \left[MRS_t(\theta) R_t - 1 \right] \left[1, \frac{c_{t-1}}{c_{t-2}}, R_{t-1} \right]'$ can serves as the basis for estimation. The elements of vector

$$Z_t = \left[1, \frac{c_{t-1}}{c_{t-2}}, R_{t-1} \right]'$$

Are just instrumental variables which are a subset of information set I_{t-1}.

9.2.2 A Presentation of GMM

Suppose that the model involves K parameters $\theta = (\theta_1, \cdots \theta_K)'$, and that we have a set of L orthogonal conditions (or population conditions).

$$E\left[m_l(Y_i, X_i, Z_i, \theta) \right] = E\left[m_{il}(\theta) \right] = 0, \quad (i = 1, \cdots, n; l = 1, \cdots, L) \tag{9.6}$$

Where Y_i, X_i and Z_i are variables that appear in the model and the subscript i on $m_{il}(\theta)$ indicates the dependence on (Y_i, X_i, Z_i). Denote the corresponding sample means as

$$\overline{m}_l(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_l(Y_i, X_i, Z_i, \theta) = \frac{1}{n}\sum_{i=1}^{n} m_{il}(\theta), \quad (l = 1, \cdots L) \tag{9.7}$$

If we denote $m_i(\theta)$ as $\left[m_{i1}(\theta), \cdots m_{iL}(\theta)\right]'$, then

$$\overline{m}(\theta) = \begin{bmatrix} \overline{m}_1(\theta) \\ \vdots \\ \overline{m}_L(\theta) \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} m_{i1}(\theta) \\ \vdots \\ m_{iL}(\theta) \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta) \tag{9.8}$$

Definition [GMM Estimator]: The GMM estimator is

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left[\overline{m}(\theta)' \hat{W} \overline{m}(\theta)\right] \tag{9.9}$$

Where $\overline{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta)$, $\hat{W}$ is a $L \times L$ symmetric nonsingular matrix which is possibly data-dependent. Here, we assume L>K, i.e., the number of moments may be larger than the number of parameters.

$$F.O.C \quad \left(\frac{\partial \overline{m}(\theta)}{\partial \theta'}\right)' \hat{W} \overline{m}(\theta)\bigg|_{\hat{\theta}} = 0 \tag{9.10}$$

Order condition: $L \geq K$.

Rank condition: $\frac{\partial \overline{m}(\theta)}{\partial \theta'} = G(\theta), \quad rank\left[G(\theta)\right] = K \qquad (9.11)$

L=K, exactly identified
L>K, over identified
L<K, under identified

9.2.3 Asymptotic Properties of GMM Estimator

1) Consistency

Assumptions:

A9.1 Parameter Space Θ is compact (closed and bounded)

A9.2 (Y_i, X_i, Z_i) is jointly stationary and ergodic

A9.3 There exists some parameter θ^0 in Θ such that $m(\theta^0) = 0$

A9.4 θ^0 is the only solution of $m(\theta)\big|_{\theta^0} = 0$ over Θ

A9.5 $\hat{W} \xrightarrow{a.s} W$, where W is a non-stochastic symmetric and nonsingular matrix

Suppose Assumptions 9.1-9.5 hold, then $\hat{\theta}_{GMM} \xrightarrow{a.s} \theta^0$.

Sketch of proof:

i) By A9.2 and Ergodic Theorem

$$\overline{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta) \xrightarrow{a.s} E\left[m_i(\theta)\right]$$

ii)Define $\hat{q}(\theta) = -\overline{m}(\theta)' \cdot \hat{W} \cdot \overline{m}(\theta) \leq 0, \quad q(\theta) = -E\left[m(\theta)\right]' \cdot W \cdot E\left[m(\theta)\right] \leq 0$.

By A9.5 ,Ergodic Theorem and Slutsky theorem,

$$\hat{q}(\theta) \xrightarrow{a.s} q(\theta)$$

iii) To show this theorem, we need the following lemma

Lemma [extreme estimator]: Let $\hat{q}(\theta)$ be a stochastic real-valued function of $\theta \in \Theta$, and $q(\theta)$ be a non-stochastic real-valued function of θ, where Θ is a compact parameter space. Suppose that for each θ , $\hat{q}(\theta)$ is a measurable function of the data, and for each n, $\hat{q}(\theta)$ is continuous in $\theta \in \Theta$. Suppose

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \hat{q}(\theta)$$
$$\theta^0 = \arg\max_{\theta \in \Theta} q(\theta)$$

is the unique maximizer. Where $q(\theta)$ is continuous in $\theta \in \Theta$. Also suppose $\hat{q}(\theta) - q(\theta) \xrightarrow{a.s} 0$, then

$$\hat{\theta} - \theta^0 \xrightarrow{a.s} 0 \qquad (9.12)$$

2) Asymptotic Normality of GMM

A9.6 $\theta^0 \in \operatorname{int}(\Theta)$

A9.7 (i) For each i, $m_i(\theta)$ is continuously differentiable with respect to $\theta \in \Theta$.

(ii) $\sqrt{n}\overline{m}(\theta^0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_i(\theta^0) \xrightarrow{d} N(0, \Phi) \qquad (9.13)$

Where $\Phi = A\text{var}\left[\sqrt{n}\bar{m}\left(\theta^0\right)\right]$ is finite and positive definite, then

$$\hat{\theta}_{GMM} \xrightarrow{d} N\left(\theta^0, V_{GMM}\right) = N\left(\theta^0, \frac{1}{n}\Omega\right)$$

Where $\Omega = \left[G\left(\theta^0\right)' WG\left(\theta^0\right)\right]^{-1} G\left(\theta^0\right)' W\Phi WG\left(\theta^0\right)\left[G\left(\theta^0\right)' WG\left(\theta^0\right)\right]^{-1}$.

$$G\left(\theta^0\right) = \lim \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial m_i\left(\theta^0\right)}{\partial \theta^{0\prime}}\right] \qquad (9.14)$$

Proof:

$\because \quad \theta^0 \in \text{int}\left(\Theta\right) \quad \hat{\theta}_{GMM} \xrightarrow{a.s} \theta^0$

$\therefore \quad \hat{\theta} \in \text{int}\left(\Theta\right)$

By (9.10), we have

$$\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\right]' \hat{W}\bar{m}\left(\theta\right)\Bigg|_{\hat{\theta}} = 0 \qquad (9.15)$$

Using the Taylor series expansion, we have

$$\bar{m}\left(\hat{\theta}\right) = \bar{m}\left(\theta^0\right) + \frac{\partial \bar{m}\left(\bar{\theta}\right)}{\partial \bar{\theta}'}\left(\bar{\theta} - \theta^0\right) \qquad (9.16)$$

Where $\bar{\theta} = \lambda\hat{\theta}_{GMM} + \left(1-\lambda\right)\theta^0, \quad \lambda \in \left(0,1\right)$.

Substituting (9.16) into (9.15),

$$\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\hat{\theta}}\right]' \hat{W}\bar{m}\left(\theta^0\right) + \left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\hat{\theta}}\right]' \hat{W}\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\bar{\theta}}\right]\left(\hat{\theta} - \theta^0\right) = 0$$

It follows that

$$\sqrt{n}\left(\hat{\theta}_{GMM} - \theta^0\right) = -\left\{\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\hat{\theta}}\right]' \hat{W}\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\bar{\theta}}\right]\right\}^{-1}\left[\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\hat{\theta}}\right]' \hat{W}\sqrt{n}\bar{m}\left(\theta^0\right) \qquad (9.17)$$

By $\hat{\theta} \xrightarrow{a.s} \theta^0$ and A9.7, we have

$$\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\hat{\theta}} \xrightarrow{a.s} G\left(\theta^0\right)$$

$$\frac{\partial \bar{m}\left(\theta\right)}{\partial \theta'}\Bigg|_{\bar{\theta}} \xrightarrow{a.s} G\left(\theta^0\right)$$

$$\left\{\left[\left.\frac{\partial\overline{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right]'\hat{W}\left[\left.\frac{\partial\overline{m}(\theta)}{\partial\theta'}\right|_{\overline{\theta}}\right]\right\}^{-1}\left[\left.\frac{\partial\overline{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right]'\hat{W}\xrightarrow{a.s}\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W \quad (9.18)$$

$$\sqrt{n}\left(\hat{\theta}-\theta^0\right)\xrightarrow{d}N\left(0,\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W\Phi WG\left(\theta^0\right)\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}\right)$$

i.e. $\hat{\theta}\xrightarrow{d}N\left(\theta^0,V_{GMM}\right)$.

Where

$$V_{GMM}=\frac{1}{n}\Omega,\quad \Omega=\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W\Phi WG\left(\theta^0\right)\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}.$$

3）Asymptotic Efficiency

There are many possible choices of $\hat{W}$, which one is the optimal choice of $\hat{W}$?

If we define

$$\Omega_0=\left[G\left(\theta^0\right)'\Phi^{-1}G\left(\theta^0\right)\right]^{-1} \quad (9.20)$$

Which is obtained from Ω by choosing $W=\Phi^{-1}$. Then $\Omega-\Omega_0$ is positive semi-definite.

Proof:

$$\Omega-\Omega_0=\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W\Phi WG\left(\theta^0\right)\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}$$
$$-\left[G\left(\theta^0\right)'\Phi^{-1}G\left(\theta^0\right)\right]^{-1}$$
$$=\left\{\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W-\left[G\left(\theta^0\right)'\Phi^{-1}G\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'\Phi^{-1}\right\}\Phi$$
$$\left\{\left[G\left(\theta^0\right)'WG\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'W-\left[G\left(\theta^0\right)'\Phi^{-1}G\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'\Phi^{-1}\right\}'$$
$$\geq 0$$

4) Two-Stage GMM Estimator

Step 1: Find a consistent preliminary estimator $\tilde{\theta}$.

$$\tilde{\theta}=\arg\min_{\theta\in\Theta}\left[\overline{m}(\theta)'\tilde{W}\overline{m}(\theta)\right]$$

For some pre-specified $\tilde{W}$. For convenience, we can set $\tilde{W} = I$.

Step 2: Find a preliminary consistent estimator $\tilde{\Phi}$ for $\Phi = A\text{var}\left(\sqrt{n}\bar{m}\left(\theta^0\right)\right)$,

and choose $\hat{W} = \tilde{\Phi}^{-1}$.

Case i): If $\left\{m_i\left(\theta^0\right)\right\}$ is stationary MDS, then

$$\tilde{\Phi} = \frac{1}{n}\sum_{i=1}^{n} m_i\left(\tilde{\theta}\right) m_i\left(\tilde{\theta}\right)' \qquad (9.24)$$

Which will be consistent for

$$\Phi = E\left[m_i\left(\theta^0\right) m_i\left(\theta^0\right)'\right]$$

Case ii): If $\left\{m_i\left(\theta^0\right)\right\}$ is not MDS, then

$$\tilde{\Phi} = \sum_{j=-q_n}^{q_n} k\left(j / q_n \right) \tilde{\Gamma}_j \qquad (9.25)$$

$$\tilde{\Gamma}_j = \frac{1}{n}\sum_{i=j+1}^{n} m_i\left(\tilde{\theta}\right) m_{i-j}\left(\tilde{\theta}\right)' \quad \text{for} \quad j \geq 0 \quad \text{and} \quad \tilde{\Gamma}_j = \tilde{\Gamma}'_{-j} \quad \text{if} \quad j < 0 \qquad (9.26)$$

The $\tilde{\Phi}$ in (9.25) is consistent for

$$\Phi = \sum_{j=-\infty}^{+\infty} \Gamma_j$$

Where $\Gamma_j = \text{cov}\left[m_i\left(\theta^0\right), m_{i-j}\left(\theta^0\right)\right] = E\left[m_i\left(\theta^0\right) m_{i-j}\left(\theta^0\right)'\right]$.

Finally, find an asymptotically optimal estimator $\hat{\theta}$

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \bar{m}\left(\theta\right)' \tilde{\Phi}^{-1} \bar{m}\left(\theta\right)$$

This two-stage GMM estimator is asymptotically optimal. And it can be shown that

$$\hat{\Omega}_0 = \left[G\left(\hat{\theta}\right)\hat{\Phi}^{-1}G\left(\hat{\theta}\right)'\right]^{-1} \xrightarrow{a.s} \Omega_0 \qquad (9.27)$$

Where $\hat{\Phi}$ is calculated from (9.24) ~(9.26) , replacing $\tilde{\theta}$ by $\hat{\theta}$.

9.3 GMM Estimation for Econometric Model

Suppose we have model

$$Y_i = h\left(X_i, \beta^0\right) + \varepsilon_i \qquad (9.28)$$

Where it is possible that

$$Cov\left[\varepsilon_i, h\left(X_i, \beta^0\right)\right] \neq 0$$

Or even

$Cov\left(\varepsilon_i, X_j\right) \neq 0$ for all i and j,

$$E\left(\varepsilon\varepsilon' \middle| X\right) = \sigma^2 \Omega$$

Suppose there are K variables in X_i. Now we find out some set of L instrumental variables Z_i where $L \geq K$, and Z_i satisfies

$$E\left(Z_i \varepsilon_i \middle| X_i\right) = 0 \qquad (9.29)$$

Thus, we have orthogonality conditions of moment

$$E\left(Z_i \varepsilon_i\right) = 0$$

The sample moments will be

$$\overline{m}\left(\hat{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n} Z_i e_i\left(X_i, \hat{\beta}\right) = \frac{1}{n} Z' e\left(X, \hat{\beta}\right)$$

And

$$\hat{q} = \overline{m}\left(\hat{\beta}\right)' \hat{W} \overline{m}\left(\hat{\beta}\right) \qquad (9.30)$$

$$= \frac{1}{n^2}\left[e\left(X, \hat{\beta}\right)' Z\right] \hat{W} \left[Z' e\left(X, \hat{\beta}\right)\right]$$

1）For linear model $Y_i = X_i' \beta^0 + \varepsilon_i$, $\overline{m}\left(\hat{\beta}\right) = \frac{1}{n} Z'\left(Y - X\hat{\beta}\right)$

$$F.O.C \quad -\frac{2}{n^2}\left(X'Z\right)\hat{W}\left[Z'\left(Y - X\hat{\beta}\right)\right] = 0$$

$$\hat{\beta} = \left[\left(X'Z\right)\hat{W}Z'X\right]^{-1}\left[\left(X'Z\right)\hat{W}Z'Y\right] \qquad (9.31)$$

The optimal choice of W for this estimator is $\hat{W} = \Phi^{-1}$, where

$$\Phi = A\text{var}\left[\sqrt{n}\,\overline{m}\left(\beta^0\right)\right] \qquad (9.32)$$

$$\overline{m}\left(\beta^0\right) = \frac{1}{n}\sum_{i=1}^{n} m_i\left(\beta^0\right)$$

$$........=\frac{1}{n}\sum_{i=1}^{n}Z_i\varepsilon_i\left(X_i,\beta^0\right)=\frac{1}{n}Z'\varepsilon\left(X,\beta^0\right) \tag{9.33}$$

$$\therefore\quad \Phi=A\operatorname{var}\left[\sqrt{n}\left(\frac{1}{n}Z'\varepsilon\right)\right]=\frac{1}{n}E\left(Z'\varepsilon\varepsilon'Z\right)=\frac{1}{n}Z'\Sigma Z \tag{9.34}$$

Obviously, for spherical disturbances $\Sigma=\sigma^2 I_n$, the GMM estimator in (9.31) $\hat{\beta}=\left[(X'Z)(Z'Z)^{-1}Z'X\right]^{-1}\left[(X'Z)(Z'Z)^{-1}Z'Y\right]$ is IV estimator, and

$$\begin{aligned}V_{GMM}&=\frac{1}{n}\Omega_0\\&=\frac{1}{n}\left(G\left(\beta^0\right)'\Phi^{-1}G\left(\beta^0\right)\right)^{-1}\\&=\frac{1}{n}\left\{\left(\frac{1}{n}X'Z\right)\left(\frac{\sigma^2}{n}(Z'Z)\right)^{-1}\left(\frac{1}{n}Z'X\right)\right\}^{-1}\\&=\sigma^2\left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}\end{aligned}$$

If the condition $E\left(X_i\varepsilon_i\right)=0$ is satisfied, we have $Z=X$, the GMM estimator in (9.31) is just OLS estimator.

In general, $\Sigma=\sigma^2\Omega\neq\sigma^2 I_n$

$$\begin{aligned}\Phi&=\frac{1}{n}Z'\Sigma Z\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}Z_iZ_j'C\text{ov}\left(\varepsilon_i,\varepsilon_j\right)\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}Z_iZ_j'C\text{ov}\left[\left(Y_i-X_i'\beta^0\right),\left(Y_j-X_j'\beta^0\right)\right]\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}Z_iZ_j'\end{aligned}$$

Now, How to estimate $\frac{1}{n}Z'\Sigma Z$?

ⅰ) If the disturbances are only heteroscedasticity, then white (1982) estimator can be use

$$\frac{1}{n}Z'\hat{\Sigma}Z=\frac{1}{n}Z'\begin{pmatrix}e_1^2&0&0\\0&\ddots&0\\0&0&e_n^2\end{pmatrix}Z \tag{9.35}$$

ⅱ) If the disturbances are heteroscedastic and autocorrelated ,and the process is stationary ergodic, then Newey-West (1987) estimator is available

$$\frac{1}{n}Z'\hat{\Sigma}Z = S_0 + \frac{1}{n}\sum_{j=1}^{P} w(j)\sum_{i=j+1}^{n} e_i e_{i-j}\left(Z_i Z'_{i-j} + Z_{i-j}Z'_i\right) \qquad (9.36)$$

Where $w(j) = 1 - \frac{j}{L+1}$, the maximum lag length L must be determined in advance.

To state succinctly,

$$\hat{\beta}_{GMM} = \left[X'Z\left(Z'\hat{\Sigma}Z\right)^{-1}Z'X\right]^{-1}\left[X'Z\left(Z'\hat{\Sigma}Z\right)^{-1}Z'Y\right] \qquad (9.37)$$

$$\hat{V}_{GMM} = \left[X'Z\left(Z'\hat{\Sigma}Z\right)^{-1}Z'X\right]^{-1} \qquad (9.38)$$

2) For nonlinear model $Y_i = h\left(X_i, \beta^0\right) + \varepsilon_i$, $\overline{m}\left(\beta^0\right) = \frac{1}{n}\sum_{i=1}^{n} Z_i\varepsilon_i\left(X_i, \beta^0\right)$,

$$\frac{\partial \overline{m}\left(\beta^0\right)}{\partial \beta^{0\prime}} = \frac{1}{n}\sum_{i=1}^{n} Z_i\left(\frac{\partial \varepsilon_i}{\partial \beta^{0\prime}}\right) = \frac{1}{n}\sum_{i=1}^{n} Z_i\left(-\frac{\partial h\left(X_i, \beta^0\right)}{\partial \beta^{0\prime}}\right)$$

The derivatives are the pseudo-regressors in the linearized regression model, using the notation

$$\frac{\partial \varepsilon_i}{\partial \beta'} = -X_{i0}$$

We have

$$\frac{\partial \overline{m}\left(\beta^0\right)}{\partial \beta^{0\prime}} = \frac{1}{n}\sum_{i=1}^{n} -Z_i X_{i0} = -\frac{1}{n}Z'X_0 \qquad (9.39)$$

The GMM estimator for β is

$$\hat{\beta}_{GMM} = \left[\left(X_0'Z\right)\hat{W}\left(Z'X_0\right)\right]^{-1}\left[\left(X_0'Z\right)\hat{W}\left(Z'Y\right)\right] \qquad (9.40)$$

The calculate procedure for weighting matrix $\hat{W}$ is the same as (9.33)-(9.36), replacing $\varepsilon_i = Y_i - X_i'\beta^0$ by $\varepsilon_i = Y_i - h\left(X_i, \beta^0\right)$, and $e_i = Y_i - X_i'\hat{\beta}$ by ....$e_i = Y_i - h\left(X_i, \hat{\beta}\right)$.

9.4 Hypothesis Testing

9.4.1 Model Specification Testing

How to test whether the model as characterized by $E\left[m_i\left(\theta^0\right)\right] = 0$ for some

θ^0 is correctly specified? Use the sample moment

$$\bar{m}\left(\hat{\theta}\right)=\frac{1}{n}\sum_{i=1}^{n}m_i\left(\hat{\theta}\right)$$

And see if it is significantly different from zero. For this purpose, we need to know the asymptotic distribution of $\sqrt{n}\bar{m}\left(\hat{\theta}\right)$, then we can construct statistic as following

$$J=n\hat{q}=n\bar{m}\left(\hat{\theta}\right)'\hat{W}\bar{m}\left(\hat{\theta}\right)\xrightarrow{d}\chi^2\left(L-K\right) \qquad (9.41)$$

Proof:

Consider the test statistic

$$\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\hat{\theta}\right)=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)+\hat{\Phi}^{-1/2}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right) \qquad (9.42)$$

Which follows by a Taylor series expansion, and $\bar{\theta}$ lies between $\hat{\theta}$ and θ^0. On the other hand, from F.O.C we have obtained

$$\begin{aligned}&\sqrt{n}\left(\hat{\theta}-\theta^0\right)\\&=-\left[\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right)'\hat{\Phi}^{-1}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\right]^{-1}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right)\\&=-\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\bar{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right) \qquad (9.43)\end{aligned}$$

It follows that

$$\begin{aligned}&\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\hat{\theta}\right)\\&=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)+\hat{\Phi}^{-1/2}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right)\\&=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)-\hat{\Phi}^{-1/2}G\left(\bar{\theta}\right)\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\bar{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right)\\&=\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)\right] \qquad (9.44)\end{aligned}$$

Where

$$\hat{\Pi}=I-\hat{\Phi}^{-1/2}G\left(\bar{\theta}\right)\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\bar{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1/2} \qquad (9.45)$$

By Gordin's CLT for $\left\{m_i\left(\theta^0\right)\right\}$ and Slutsky theorem, we have

$$\sqrt{n}\overline{m}\left(\theta^0\right)\xrightarrow{d}N\left(0,\Phi\right) \tag{9.46}$$

$$\Phi^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\xrightarrow{d}N\left(0,I_L\right)$$

Also, we have

$$\hat{\Pi}=I-\hat{\Phi}^{-1/2}G\left(\overline{\theta}\right)\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\overline{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1/2}$$

$$\xrightarrow{a.s}I-\Phi^{-1/2}G\left(\theta^0\right)\left[G\left(\theta^0\right)'\Phi^{-1}G\left(\theta^0\right)\right]^{-1}G\left(\theta^0\right)'\Phi^{-1/2}=\Pi \tag{9.47}$$

Where Π is a $L\times L$ symmetric matrix which is also idempotent, by the following lemma:

Lemma: If $v\xrightarrow{d}N\left(0,I\right)$ and Π is a $L\times L$ idempotent matrix with rank J, then the quadratic form

$$v'\Pi v\xrightarrow{d}\chi^2\left(J\right)$$

Now, we have

$$\left[\Phi^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]'\Pi\left[\Phi^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]\xrightarrow{d}\chi^2\left(L-K\right) \tag{9.48}$$

By (9.44),

$$\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\hat{\theta}\right)=\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]$$

Since

$$\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]'\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]\xrightarrow{a.s}\left[\Phi^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]'\Pi\left[\Phi^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]$$

If follows that

$$\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]'\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]\xrightarrow{d}\chi^2\left(L-K\right) \tag{9.49}$$

So

$$J=n\hat{q}$$

$$=\left(\sqrt{n}\overline{m}\left(\hat{\theta}\right)\right)'\hat{\Phi}^{-1}\sqrt{n}\overline{m}\left(\hat{\theta}\right)$$

$$=\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\hat{\theta}\right)\right]'\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\hat{\theta}\right)\right]$$

$$=\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]'\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\overline{m}\left(\theta^0\right)\right]\xrightarrow{d}\chi^2\left(L-K\right) \tag{9.50}$$

This statistic is often called J-test, testing for over-identification in the GMM literature, in essence, this test is used to check if the model specified as $E\left[m_i\left(\theta^0\right)\right]=0$ is correctly specified.

Note: several papers in the July 1996 issue of the Journal of Business and

Economic Statistics report that small samples far exceed the nominal size (i.e. the test rejects too often).

9.4.2 GMM Counterparts to the Wald, LR and LM Tests

Hypothesis of interest

$$H_0 : R(\theta^0) = \gamma$$

Where $R(\bullet)$ is a $J \times 1$ continuously differentiable vector function, and the $J \times K$ matrix $\frac{\partial R(\theta^0)}{\partial \theta'} = R'(\theta^0)$ is of full row rank.

1) Wald Test

By Taylor series expansion and $R(\theta^0) = \gamma$ under H_0, we have

$$\sqrt{n}\left(R(\hat{\theta}) - \gamma\right) = \sqrt{n}\left(R(\theta^0) - \gamma\right) + R'(\bar{\theta})\sqrt{n}\left(\hat{\theta} - \theta^0\right)$$
$$= R'(\bar{\theta})\sqrt{n}\left(\hat{\theta} - \theta^0\right)$$

Where $\bar{\theta}$ lies between $\hat{\theta}$ and θ^0.

Because $R'(\bar{\theta}) \xrightarrow{a.s} R'(\theta^0)$ given continuity of $R'(\bullet)$, and

$$\bar{\theta} - \theta^0 \xrightarrow{a.s} 0,$$

$$\sqrt{n}\left(\hat{\theta} - \theta^0\right) \xrightarrow{d} N(0, \Omega_0)$$

We have

$$\sqrt{n}\left(R(\hat{\theta}) - \gamma\right) \xrightarrow{d} N\left[0, R'(\theta^0)\Omega_0 R'(\theta^0)'\right]$$

By Slutsky theorem, it follows that

$$\sqrt{n}\left[R(\hat{\theta}) - \gamma\right]'\left[R'(\theta^0)\Omega_0 R'(\theta^0)'\right]^{-1}\sqrt{n}\left[R(\hat{\theta}) - \gamma\right] \xrightarrow{d} \chi^2(J)$$

$$\sqrt{n}\left[R(\hat{\theta}) - \gamma\right]'\left[R'(\hat{\theta})\hat{\Omega}_0 R'(\hat{\theta})'\right]^{-1}\sqrt{n}\left[R(\hat{\theta}) - \gamma\right] \xrightarrow{d} \chi^2(J)$$

Thus, under H_0, we have Wald Test statistic:

$$n\left[R(\hat{\theta}) - \gamma\right]'\left[R'(\hat{\theta})\hat{\Omega}_0 R'(\hat{\theta})'\right]^{-1}\left[R(\hat{\theta}) - \gamma\right] \xrightarrow{d} \chi^2(J) \qquad (9.51)$$

2) LR Test

$$LR = -2\left(\ln L_R - \ln L_u\right)$$

Suppose, under H_0,

$$\hat{q}_R = \overline{m}\left(\hat{\theta}_R\right)' \hat{W}_R \overline{m}\left(\hat{\theta}_R\right) \tag{9.52}$$

And under H_1,

$$\hat{q}_u = \overline{m}\left(\hat{\theta}_u\right)' \hat{W}_u \overline{m}\left(\hat{\theta}_u\right) \tag{9.53}$$

Then

$$LR_{GMM} = n\left(\hat{q}_R - \hat{q}_u\right) \xrightarrow{d} \chi^2\left(J\right) \tag{9.54}$$

Note, it is necessary to use the same weighting matrix W, in both restricted and unrestricted estimators. Since the unrestricted estimator is consistent under both H_0 and H_1, a consistent unrestricted estimator of θ is use to compute W, that is

$$\hat{W} = \hat{\Phi}^{-1} = A\operatorname{var}\left[\sqrt{n}\overline{m}\left(\hat{\theta}_u\right)\right] \tag{9.55}$$

3) LM Test

Suppose $\alpha = \dfrac{\partial q_R}{\partial \theta'}$, then the counterpart to the LM statistic is equivalent to Wald test as below

$$H_0 : \alpha = 0$$

$$Wald = \alpha' \left[\operatorname{var}\left(\alpha\right)\right]^{-1} \alpha$$

See P550-551, Greene 5th.

9.4 Hypothesis Testing

9.4.1 Model Specification Testing

How to test whether the model as characterized by $E\left[m_i\left(\theta^0\right)\right]=0$ for some θ^0 is correctly specified? Use the sample moment

$$\bar{m}\left(\hat{\theta}\right)=\frac{1}{n}\sum_{i=1}^{n}m_i\left(\hat{\theta}\right)$$

and see if it is significantly different from zero. For this purpose, we need to know the asymptotic distribution of $\sqrt{n}\bar{m}\left(\hat{\theta}\right)$, then we can construct statistic as following

$$J=n\hat{q}=n\bar{m}\left(\hat{\theta}\right)'\hat{W}\bar{m}\left(\hat{\theta}\right)\xrightarrow{d}\chi^2\left(L-K\right) \qquad (9.41)$$

Proof:

Consider the test statistic

$$\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\hat{\theta}\right)=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)+\hat{\Phi}^{-1/2}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right) \qquad (9.42)$$

Which follows by a Taylor series expansion, and $\bar{\theta}$ lies between $\hat{\theta}$ and θ^0. On the other hand, from F.O.C we have obtained

$$\begin{aligned}&\sqrt{n}\left(\hat{\theta}-\theta^0\right)\\&=-\left[\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right)'\hat{\Phi}^{-1}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\right]^{-1}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\hat{\theta}}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right)\\&=-\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\bar{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right) \qquad (9.43)\end{aligned}$$

It follows that

$$\begin{aligned}&\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\hat{\theta}\right)\\&=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)+\hat{\Phi}^{-1/2}\left(\left.\frac{\partial\bar{m}(\theta)}{\partial\theta'}\right|_{\bar{\theta}}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right)\\&=\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)-\hat{\Phi}^{-1/2}G\left(\bar{\theta}\right)\left[G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}G\left(\bar{\theta}\right)\right]^{-1}G\left(\hat{\theta}\right)'\hat{\Phi}^{-1}\sqrt{n}\bar{m}\left(\theta^0\right)\\&=\hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}\left(\theta^0\right)\right] \qquad (9.44)\end{aligned}$$

Where

$$\hat{\Pi} = I - \hat{\Phi}^{-1/2} G(\bar{\theta}) \left[G(\hat{\theta})' \hat{\Phi}^{-1} G(\bar{\theta}) \right]^{-1} G(\hat{\theta})' \hat{\Phi}^{-1/2} \tag{9.45}$$

By Gordin's CLT for $\{m_i(\theta^0)\}$ and Slutsky theorem, we have

$$\sqrt{n}\bar{m}(\theta^0) \xrightarrow{d} N(0, \Phi) \tag{9.46}$$

$$\Phi^{-1/2}\sqrt{n}\bar{m}(\theta^0) \xrightarrow{d} N(0, I_L)$$

Also, we have

$$\hat{\Pi} = I - \hat{\Phi}^{-1/2} G(\bar{\theta}) \left[G(\hat{\theta})' \hat{\Phi}^{-1} G(\bar{\theta}) \right]^{-1} G(\hat{\theta})' \hat{\Phi}^{-1/2}$$

$$\xrightarrow{a.s} I - \Phi^{-1/2} G(\theta^0) \left[G(\theta^0)' \Phi^{-1} G(\theta^0) \right]^{-1} G(\theta^0)' \Phi^{-1/2} = \Pi \tag{9.47}$$

Where Π is a $L \times L$ symmetric matrix which is also idempotent with $rank(\Pi) = L - K$, by the following lemma:

Lemma: If $v \xrightarrow{d} N(0, I)$ and Π is a $L \times L$ idempotent matrix with rank J, then the quadratic form

$$v'\Pi v \xrightarrow{d} \chi^2(J)$$

Now, we have

$$\left[\Phi^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]' \Pi \left[\Phi^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right] \xrightarrow{d} \chi^2(L-K) \tag{9.48}$$

By (9.44),

$$\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\hat{\theta}) = \hat{\Pi}\left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]$$

Since

$$\left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]' \hat{\Pi} \left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right] \xrightarrow{a.s} \left[\Phi^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]' \Pi \left[\Phi^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]$$

If follows that

$$\left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right]' \hat{\Pi} \left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\theta^0) \right] \xrightarrow{d} \chi^2(L-K) \tag{9.49}$$

So

$$J = n\hat{q}$$

$$= \left(\sqrt{n}\bar{m}(\hat{\theta}) \right)' \hat{\Phi}^{-1} \sqrt{n}\bar{m}(\hat{\theta})$$

$$= \left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\hat{\theta}) \right]' \left[\hat{\Phi}^{-1/2}\sqrt{n}\bar{m}(\hat{\theta}) \right]$$

$$= \left[\hat{\Phi}^{-1/2} \sqrt{n} \bar{m}\left(\theta^0\right) \right]' \hat{\Pi} \left[\hat{\Phi}^{-1/2} \sqrt{n} \bar{m}\left(\theta^0\right) \right] \xrightarrow{d} \chi^2\left(L-K\right) \qquad (9.50)$$

This statistic is often called J-test, testing for over-identification in the GMM literature, in essence, this test is used to check if the model specified as $E\left[m_i\left(\theta^0\right) \right] = 0$ is correctly specified.

Note: several papers in the July 1996 issue of the Journal of Business and Economic Statistics report that small samples far exceed the nominal size (i.e. the test rejects too often).

9.4.2 Testing Subsets of Orthogonal Conditions

Suppose we can divide the L instruments into two groups: the vector Z_{i1} of L_1 variables that are known to satisfy the orthogonal conditions, and the vector Z_{i2} of remaining $\mathrm{L\text{-}L_1}$ variables that are suspect. Since the ordering of instruments does not change the numerical values of the estimator and test statistics, we can assume without loss of generality that the last $\mathrm{L - L_1}$ elements of Z_i are the suspect instruments:

$$\mathrm{Z_i} = \begin{pmatrix} Z_{i1} \\ Z_{i2} \end{pmatrix} \begin{matrix} \mathrm{L_1 \text{ rows}} \\ \mathrm{L - L_1 \text{ rows}} \end{matrix} \qquad (9.51)$$

The part of the model we wish to test is

$$E(Z_{i2}\varepsilon_i) = 0$$

This restriction is testable if there are at least as many nonsuspect instruments as there are coefficients so that $L_1 \geq K$. The basic idea is to compare two J statistics from two separate GMM estimators of the same coefficients θ, one using only the instruments included in Z_{i1}, and the other using also the suspect instruments Z_{i2}. In these contexts, we can use a difference-in-Sargan test, the **C statistic**.

In accordance with the partition of Z_i, the sample orthogonal conditions $\bar{m}\left(\hat{\theta}\right)$ and Φ can be written as

$$\bar{m}\left(\hat{\theta}\right) = \begin{bmatrix} \bar{m}_1\left(\hat{\theta}\right) \\ \bar{m}_2\left(\hat{\theta}\right) \end{bmatrix}, \quad \hat{\Phi} = \begin{bmatrix} \hat{\Phi}_{11} & \hat{\Phi}_{12} \\ \hat{\Phi}_{21} & \hat{\Phi}_{22} \end{bmatrix} \qquad (9.52)$$

For a consistent estimator $\hat{\Phi}$ of Φ, the efficient GMM estimator using all L instruments and its associated J statistic is (9.50). The efficient GMM estimator of the same coefficient vector θ using only the first L_1 instruments and its associated J statistic is:

$$J_1 = n\hat{q}_1 = n\left[\bar{m}_1\left(\hat{\theta}\right)'\hat{\Phi}_{11}^{-1}\bar{m}_1\left(\hat{\theta}\right)\right] = n\left[\bar{m}\left(\hat{\theta}\right)' R\left(R'\hat{\Phi}R\right)^{-1} R'\bar{m}\left(\hat{\theta}\right)\right]$$

$$\xrightarrow{d}\left[\sqrt{n}\bar{m}\left(\theta^0\right)'\Phi^{-\frac{1}{2}}\right]\Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}}\left[\Phi^{-\frac{1}{2}}\sqrt{n}\bar{m}\left(\theta^0\right)\right]$$

$$\xrightarrow{d}\left[\sqrt{n}\bar{m}\left(\theta^0\right)'\Phi^{-\frac{1}{2}}\right]\mathrm{M}\left[\Phi^{-\frac{1}{2}}\sqrt{n}\bar{m}\left(\theta^0\right)\right]$$

Where $R = \begin{pmatrix} I_1 \\ 0 \end{pmatrix}$, $\Pi_1 = I_L - \left(R'\Phi R\right)^{-\frac{1}{2}} R'G\left[G'R\left(R'\Phi R\right)^{-1}R'G\right]^{-1}G'R\left(R'\Phi R\right)^{-\frac{1}{2}}$,

$$M = \Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}}.$$

Both Π and M are symmetric and idempotent matrices, and

$$\begin{aligned}
\Pi\mathrm{M} &= \Pi\cdot\Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}} \\
&= \left[I_L - \Phi^{-\frac{1}{2}}G\left(G'\Phi^{-1}G\right)^{-1}G'\Phi^{-\frac{1}{2}}\right]\Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}} \\
&= \Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}} \\
&\qquad - \Phi^{-\frac{1}{2}}G\left(G'\Phi^{-1}G\right)^{-1}G'R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}} \\
&= \Phi^{\frac{1}{2}}R\left(R'\Phi R\right)^{-\frac{1}{2}}\Pi_1\left(R'\Phi R\right)^{-\frac{1}{2}}R'\Phi^{\frac{1}{2}} = \mathrm{M}
\end{aligned}$$

So, $\Pi - \mathrm{M}$ is also a symmetric and idempotent matrix, with $rank\left(\Pi - \mathrm{M}\right) = rank\left(\Pi\right) - rank\left(\mathrm{M}\right) = L - L_1$

Thus, we have

$$\begin{aligned}
C &= J - J_1 = n\left(\hat{q} - \hat{q}_1\right) \\
&\xrightarrow{d}\left[\sqrt{n}\bar{m}\left(\theta^0\right)'\Phi^{-\frac{1}{2}}\right]\left(\Pi - \mathrm{M}\right)\left[\Phi^{-\frac{1}{2}}\sqrt{n}\bar{m}\left(\theta^0\right)\right]\xrightarrow{d}\chi^2\left(L - L_1\right)
\end{aligned} \tag{9.53}$$

The C statistic is computed as the difference between two J statistics. The first is computed from the fully efficient regression using the entire set of overidentifying restrictions. The second is that of the inefficient but consistent regression using a smaller set of restrictions in which a specified set of instruments are removed from the instrument list. For excluded instruments, the step is equivalent to dropping them from the instrument list. For included instruments, the C test places them in the list of included endogenous variables, treating them as endogenous regressors. The order condition must still be satisfied for this form of the equation.

9.4.3 Identification / IV Relevance Test

1. Anderson's Likelihood-ratio Test

A general approach to the problem of instrument relevance was proposed by Anderson (1984) and discussed in Hall, Rudebusch, and Wilcox (1996).

Anderson's approach considers the canonical correlations of the **X** and **Z** matrices. These measures, $r_i, i=1,...,K$ represent the correlations between linear combinations of the K columns of **X** and linear combinations of the L columns of **Z**. If an equation to be estimated by instrumental variables is identified from a numerical standpoint, all K of the canonical correlations must be significantly different from zero. Anderson's likelihood-ratio test has the null hypothesis that the smallest canonical correlation is zero and assumes that the regressors are distributed multivariate normal. Under the null, the test statistic is distributed χ^2 with $(L-K+1)$ degrees of freedom, so that it may be calculated even for an exactly identified equation. A failure to reject the null hypothesis calls the identification status of the estimated equation into question. The squared canonical correlations may be calculated as the eigenvalues of $(X'X)^{-1}(X'Z)(Z'Z)^{-1}(Z'X)$; see Hall, Rudebusch, and Wilcox (1996, 287).

2. **Kleibergen-Paap rk statisti**c（Appendix）

The Kleibergen-Paap (2005) rk statistic is a generalization of the Anderson canonical correlation rank test to the case of a non-Kronecker covariance matrix. The implementation in ranktest will calculate rk statistics that are robust to various forms of heteroskedasticity, autocorrelation, and clustering.

We are concerned with testing the rank of the $k \times m$ matrix of parameters Π. Under null hypothesis, the rank of the matrix Π is equal to q with $q < \min(k,m)$, that is, H0: rank(Π)=q. Let $\hat{\Pi}$ be an estimator of the unrestricted value of Π, $\hat{\pi} = vec(\hat{\Pi})$ and $\pi = vec(\Pi)$.

Assumption 1. The limiting behavior of the estimator of the matrix Π is characterized by

$$\sqrt{T}(\hat{\pi}-\pi) \xrightarrow{d} N(0,V), \tag{9.54}$$

where T is the sample size and V is a $km \times km$ covariance matrix.

Let $\hat{\Theta} = G\hat{\Pi}F'$, in which G and F are two non-singular matrices with dimension $k \times k$ and $m \times m$, and such that $\hat{\theta} = vec\left(\hat{\Theta}\right) = \left(F \otimes G\right)\hat{\pi}$, then

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(0, W\right), \tag{9.55}$$

where $W = \left(F \otimes G\right)V\left(F \otimes G\right)'$.

To test H$_0$: $rank\left(\Pi\right) = q$, which is equivalent to H$_0$: $rank\left(\Theta\right) = q$. The matrix Θ can be decomposed as

$$\Theta = A_q B_q + A_{q,\perp}\Lambda_q B_{q,\perp}, \tag{9.56}$$

with A_q a $k \times q$ matrix, B_q a $q \times m$ matrix, Λ_q a $\left(k-q\right)\times\left(m-q\right)$ matrix, $A_{q,\perp}$ a $k \times \left(k-q\right)$ matrix, $B_{q,\perp}$ a $\left(m-q\right)\times m$ matrix, and where $A_q' A_{q,\perp} = 0, B_q' B_{q,\perp} = 0, A_{q,\perp}' A_{q,\perp} = I_{k-q}, B_{q,\perp}' B_{q,\perp} = I_{m-q}$. Under H0, the matrix Λ_q is identical to zero. The null hypothesis H$_0$: $rank\left(\Theta\right) = q$ is identical to H$_0$: $\Lambda_q = 0$.

Assumption2. The $\left(k-q\right)\left(m-q\right)\times\left(k-q\right)\left(m-q\right)$ covariance matrix

$$\Omega_q = \left(B_{q,\perp} \otimes A_{q,\perp}'\right) W \left(B_{q,\perp} \otimes A_{q,\perp}'\right)' \tag{9.57}$$

is non-singular.

Decomposition (9.56) is also applied to the estimator $\hat{\Theta}$,

$$\hat{\Theta} = \hat{A}_q \hat{B}_q + \hat{A}_{q,\perp}\hat{\Lambda}_q \hat{B}_{q,\perp}. \tag{9.58}$$

Theorem1. Under H$_0$ and Assumptions 1-2, the limiting behavior of the elements of $\hat{\Theta}$ in (9.60) is such that $\hat{A}_q \hat{B}_q$ resulting from the singular value decomposition of $\hat{\Theta}$ is a $\sqrt{T}$ consistent estimator of $A_q B_q$ and

$$\sqrt{T}\hat{\lambda}_q \xrightarrow{d} N\left(0, \Omega_q\right) \tag{9.59}$$

where $\hat{\lambda}_q = vec\left(\hat{\Lambda}_q\right)$ and $\hat{\Lambda}_q = \hat{A}_{q,\perp}'\hat{\Theta}\hat{B}_{q,\perp}'$.

Kleibergen-Paap rk statistic. Under Assumption 1-2, the statistic

$$rk\left(q\right) = T\lambda_q'\hat{\Omega}_q\lambda_q \tag{9.60}$$

converges under H_0: rank(Θ)=q in distribution to a $\chi^2\left((k-q)(m-q)\right)$ random variable.

Proposition 1.: When $V=\left((F'F)^{-1}\otimes(G'G)^{-1}\right)$ and $\hat{\Pi}$ is a least squares estimator, the rank statistic $rk(q)$ is identical to the canonical correlation rank statistic of Anderson (1951). When $k>m$, it is equal to the sum of the $m-q$ smallest eigenvalues of $\hat{\Theta}'\hat{\Theta}$ divided by T, and when $m>k$, it is equal to the sum of the $k-q$ smallest eigenvalues of $\hat{\Theta}\hat{\Theta}'$ divided by T. The smallest eigenvalues of $\hat{\Theta}'\hat{\Theta}$ and $\hat{\Theta}\hat{\Theta}'$ represent the smallest canonical correlations when $\hat{\Pi}$ is a least squares estimator.

In many econometric models, the rank of a matrix governs the identification of the parameters. The limiting distribution of estimators of these parameters are only valid if this matrix has full rank. For example, in order to obtain the limiting distributions of GMM estimators, it is assumed that a matrix of derivatives has full rank. In such case, rank statistics can be used to test for the identification of the parameters.

9.4.4 GMM Counterparts to the Wald, LR and LM Tests

Hypothesis of interest

$$H_0 : R\left(\theta^0\right)=\gamma$$

Where $R(\bullet)$ is a $J\times 1$ continuously differentiable vector function, and the $J\times K$ matrix $\frac{\partial R\left(\theta^0\right)}{\partial\theta'}=R'\left(\theta^0\right)$ is of full row rank.

1) Wald Test

By Taylor series expansion and $R\left(\theta^0\right)=\gamma$ under H_0, we have

$$\begin{aligned}\sqrt{n}\left(R\left(\hat{\theta}\right)-\gamma\right)&=\sqrt{n}\left(R\left(\theta^0\right)-\gamma\right)+R'\left(\bar{\theta}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right)\\&=R'\left(\bar{\theta}\right)\sqrt{n}\left(\hat{\theta}-\theta^0\right)\end{aligned}$$

Where $\bar{\theta}$ lies between $\hat{\theta}$ and θ^0.

Because $R'\left(\bar{\theta}\right)\xrightarrow{a.s}R'\left(\theta^0\right)$ given continuity of $R'(\bullet)$, and

$$\bar{\theta} - \theta^0 \xrightarrow{a.s} 0,$$

$$\sqrt{n}\left(\hat{\theta} - \theta^0\right) \xrightarrow{d} N\left(0, \Omega_0\right)$$

We have

$$\sqrt{n}\left(R\left(\hat{\theta}\right) - \gamma\right) \xrightarrow{d} N\left[0, R'\left(\theta^0\right)\Omega_0 R'\left(\theta^0\right)'\right]$$

By Slutsky theorem, it follows that

$$\sqrt{n}\left[R\left(\hat{\theta}\right) - \gamma\right]'\left[R'\left(\theta^0\right)\Omega_0 R'\left(\theta^0\right)'\right]^{-1}\sqrt{n}\left[R\left(\hat{\theta}\right) - \gamma\right] \xrightarrow{d} \chi^2\left(J\right)$$

$$\sqrt{n}\left[R\left(\hat{\theta}\right) - \gamma\right]'\left[R'\left(\hat{\theta}\right)\hat{\Omega}_0 R'\left(\hat{\theta}\right)'\right]^{-1}\sqrt{n}\left[R\left(\hat{\theta}\right) - \gamma\right] \xrightarrow{d} \chi^2\left(J\right)$$

Thus, under H_0, we have Wald Test statistic:

$$n\left[R\left(\hat{\theta}\right) - \gamma\right]'\left[R'\left(\hat{\theta}\right)\hat{\Omega}_0 R'\left(\hat{\theta}\right)'\right]^{-1}\left[R\left(\hat{\theta}\right) - \gamma\right] \xrightarrow{d} \chi^2\left(J\right) \qquad (9.61)$$

2) LR Test

$$LR = -2\left(\ln L_R - \ln L_u\right)$$

Suppose, under H_0,

$$\hat{q}_R = \overline{m}\left(\hat{\theta}_R\right)'\hat{W}_R\overline{m}\left(\hat{\theta}_R\right) \qquad (9.62)$$

And under H_1,

$$\hat{q}_u = \overline{m}\left(\hat{\theta}_u\right)'\hat{W}_u\overline{m}\left(\hat{\theta}_u\right) \qquad (9.63)$$

Then

$$LR_{GMM} = n\left(\hat{q}_R - \hat{q}_u\right) \xrightarrow{d} \chi^2\left(J\right) \qquad (9.64)$$

Note, it is necessary to use the same weighting matrix W, in both restricted and unrestricted estimators. Since the unrestricted estimator is consistent under both H_0 and H_1, a consistent unrestricted estimator of θ is use to compute W, that is

$$\hat{W} = \hat{\Phi}^{-1} = A\operatorname{var}\left[\sqrt{n}\overline{m}\left(\hat{\theta}_u\right)\right] \qquad (9.65)$$

3) LM Test

Suppose $\alpha = \frac{\partial q_R}{\partial \theta'}$, then the counterpart to the LM statistic is equivalent to

Wald test as below

$$H_0 : \alpha = 0$$

$$Wald = \alpha' \left[\text{var}\left(\alpha \right) \right]^{-1} \alpha$$

See P550-551, Greene 5th.