ApEc 8213: Econometric Analysis III -- Lecture #3

Generalized Method of Moments
Hansen Chapter 13

Generalized Method of Moments (GMM) estimation is used very often in econometrics. GMM includes OLS, 2SLS and many other estimation methods as special cases. GMM started with Lars Peter Hansen's 1982 *Econometrica* paper.

One advantage of GMM is that it can be applied to complex nonlinear models, but today we focus on linear models.

I. Method of Moments Estimation (13.2, 13.3)

Method of moments estimation has a long history in statistics. It is rarely used because it is often inefficient, but this is what GMM generalizes, and GMM has good efficiency properties.

As you may know, statistical distributions can be defined by their "moments", which can be defined (for some variable X) as $E[X^k]$, where k is some integer ≥ 1. These moments usually do not equal 0, but they are often "centered" so that they equal 0. For example, if $\mu_X = E[X]$ then $E[X - \mu_X] = 0$.

Almost all of the estimation methods up through Chapter 12 of Hansen's book can be expressed as **moment equation models**. These models have the property that ***their parameters are the solutions to a system of moment equations***.

The **most general form** for a system of moment equations is:

$$E[g_i(\beta)] = 0 \qquad (13.1)$$

where $g_i(\beta)$ is a known $\ell \times 1$ vector of functions of the data from the i^{th} observation, and β is a $k \times 1$ vector of parameters.

Example: The IV model has $g_i(\beta) = Z_i e_i = Z_i(Y_i - X_i'\beta)$.

In general, the parameter β is **identified** if there is a unique "mapping" of the data to β. Since (13.1) is a system of ℓ equations with k unknowns, it can be identified only if $\ell \geq k$. If $\boldsymbol{\ell = k}$, the model is **just identified**. If $\boldsymbol{\ell > k}$, the model is **overidentified**. If $\boldsymbol{\ell < k}$, the model is **under-identified** (insufficient information to identify β).

"Classic" method of moments (MM) estimation can be applied **only for "just identified" models** ($\ell = k$). The "sample analog" of $E[g_i(\beta)]$ in equation (13.1) is:

$$\bar{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) \qquad (13.2)$$

The **method of moments estimator (MME)**, $\boldsymbol{\hat{\beta}_{mm}}$, is the value of β that **sets equation (13.2) equal to 0**:

$$\bar{g}_n(\hat{\beta}_{mm}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta}_{mm}) = 0 \qquad (13.3)$$

The equations in (13.3) are called **estimating equations**. Sometimes they provide an explicit "formula" for $\hat{\beta}_{mm}$, but other times you need to iterate to get $\hat{\beta}_{mm}$.

Hansen gives many **examples**. Here are two of them:

OLS: $g_i(\beta) = X_i e_i = X_i(Y_i - X_i'\beta) = 0$

Sample analog: $\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\beta) = 0$. Solve for β: $\hat{\beta}_{mm} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.

IV: $g_i(\beta) = Z_i e_i = Z_i(Y_i - X_i'\beta) = 0$

Sample analog: $\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}\beta) = 0$. Solve for β: $\hat{\beta}_{mm} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$.

II. Generalized Method of Moments Estimation (13.4-13-7)

Method of moments estimation requires $\ell = k$. **Consider IV estimation** with $\ell > k$. The sample analog equation is:

$$\bar{g}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i'\beta) = \frac{1}{n}(\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\beta) \quad (13.4)$$

The method of moments estimator for β is the value of β that sets (13.4) equal to 0. But **if $\boldsymbol{\ell > k}$** there is, in general, **no possible $\boldsymbol{\beta}$ that sets this equal to 0**.

How about finding **a $\boldsymbol{\beta}$ that sets (13.4) as close to 0 as possible?** One way to do this is to **express** ($\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\beta$) **as**

a regression. Define $\mu = \boldsymbol{Z'Y}$ and $\boldsymbol{G} = \boldsymbol{Z'X}$, so that (13.4) becomes $\mu = \boldsymbol{G}\beta + \eta$, where $\eta = \mu - \boldsymbol{G}\beta$. OLS makes the (squares of the) "errors" small by setting $\beta_{\text{ols}} = (\boldsymbol{G'G})^{-1}\boldsymbol{G'}\mu$.

A more general approach, **if there is heteroscedasticity**, is to **use generalized least squares**, with a weighting matrix $\boldsymbol{W}$:

$$\hat{\beta} = (\boldsymbol{G'WG})^{-1}\boldsymbol{G'W}\mu = (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZWZ'Y}$$

This minimizes the weighted sum of squares $\eta' W \eta$. **This is the GMM** (generalized method of moments) **estimator**.

The **general case (not just IV)** for when $\ell > k$ is to apply the weighting matrix $\boldsymbol{W}$ **(> 0)** to the (square of the) $\bar{g}_n(\beta)$ function:

$$J(\beta) = n\bar{g}_n(\beta)'\boldsymbol{W}\bar{g}_n(\beta)$$

$J(\beta)$ is the GMM "criterion function" (n is added for later use in deriving the distribution theory). **The GMM estimator** is defined as **the $\boldsymbol{\beta}$ that minimizes $\boldsymbol{J(\beta)}$**.

Definition 13.1. The **Generalized Method of Moments** estimator, $\hat{\beta}_{\text{gmm}}$, is defined as:

$$\hat{\beta}_{\text{gmm}} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

When $\boldsymbol{\ell = k}$, then $\widehat{\boldsymbol{\beta}}_{\textbf{gmm}} = \widehat{\boldsymbol{\beta}}_{\textbf{mm}}$, and the weighting matrix $\boldsymbol{W}$ plays no role since β can be set so that $\bar{g}_n(\beta) = 0$.

Most of this lecture will **focus on linear models**, and the **overidentified IV model** is a useful general linear model. (For the general GMM case, see Wooldridge, 2010, Ch. 14.) The IV model can be defined in terms of its moment equations:

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) \qquad (13.5)$$

where Z_i is an $\ell \times 1$ vector, and X_i and β are $k \times 1$ vectors.

The GMM Estimator (for the IV model)

The $J(\beta)$ criterion for equation (13.5) can be written as:

$$J(\beta) = n(\boldsymbol{Z'Y} - \boldsymbol{Z'X}\beta)'\boldsymbol{W}(\boldsymbol{Z'Y} - \boldsymbol{Z'X}\beta)$$

The GMM estimator minimizes $J(\beta)$. The F.O.C. are:

$$0 = \partial J(\hat{\beta})/\partial\hat{\beta} \qquad \text{(general case)}$$

$$= -2n(\tfrac{1}{n}\boldsymbol{X'Z})\boldsymbol{W}(\tfrac{1}{n}\boldsymbol{Z'}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})) \qquad \text{(IV model)}$$

(Don't worry about the n terms, since they play no role in choosing the β that sets these equations equal to zero.)

The $\hat{\beta}$ that sets this expression (IV model) equal to zero is:

Theorem 13.1. For the overidentified IV model:

$$\hat{\beta}_{\text{gmm}} = (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZWZ'Y} \qquad (13.6)$$

Note: If $\boldsymbol{W}$ is replaced by $c\boldsymbol{W}$ for some constant c, this does not change: the c "cancels out". You can **choose your own** (non-random) $\boldsymbol{W}$; this is called a **one-step GMM** estimator.

Also, for the **just-identified model** ($\ell = k$), $\boldsymbol{X'Z}$ is a $k \times k$ matrix, so (13.6) **simplifies to**:

$$\hat{\beta}_{\text{gmm}} = (\boldsymbol{Z'X})^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X'Z})^{-1}\boldsymbol{X'ZWZ'Y} = (\boldsymbol{Z'X})^{-1}\boldsymbol{Z'Y} = \hat{\beta}_{\text{iv}}$$

Finally, if we choose $(\boldsymbol{Z'Z})^{-1}$ to be the weighting matrix $\boldsymbol{W}$, then $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{2sls}}$:

Theorem 13.2. If $\boldsymbol{W} = (\boldsymbol{Z'Z})^{-1}$ then $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{2sls}}$. In addition, if $\ell = k$ then $\hat{\beta}_{\text{gmm}} = \hat{\beta}_{\text{iv}}$.

The Distribution of the GMM Estimator (for IV model)

To use $\hat{\beta}_{\text{gmm}}$ in equation (13.6) to test for statistical significance and conduct hypothesis tests, we need to know its distribution. To do this, **define** $\boldsymbol{Q} = \mathrm{E}[ZX']$ and **define** $\boldsymbol{\Omega} = \mathrm{E}[ZZ'e^2]$ (e^2 is a scalar). Then the following hold:

$$\left(\frac{1}{n}\boldsymbol{X'Z}\right)\boldsymbol{W}\left(\frac{1}{n}\boldsymbol{Z'X}\right) \underset{p}{\rightarrow} \boldsymbol{Q'WQ}$$

$$\left(\frac{1}{n}\boldsymbol{X'Z}\right)\boldsymbol{W}\left(\frac{1}{\sqrt{n}}\boldsymbol{Z'e}\right) \underset{d}{\rightarrow} \boldsymbol{Q'W}\,\mathrm{N}(0, \boldsymbol{\Omega})$$

Combining these results gives:

Theorem 13.3. Asymptotic Distribution of GMM Estimator. Under Assumptions 12.2 (IV assumptions in Chapter 12), as $n \to \infty$, $\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \underset{d}{\to} \text{N}(0, \boldsymbol{V}_\beta)$, where:

$$\boldsymbol{V}_\beta = (\boldsymbol{Q'WQ})^{-1}(\boldsymbol{Q'W}\boldsymbol{\Omega}\boldsymbol{WQ})(\boldsymbol{Q'WQ})^{-1} \qquad (13.7)$$

Strictly speaking, this assumes that we know $\boldsymbol{W}$, but this theorem applies for any estimate of $\boldsymbol{W}$, denoted by $\widehat{\boldsymbol{W}}$, that converges in probability to a positive definite matrix $\boldsymbol{W}$.

III. Efficient GMM for the IV model (13.8 – 13.10)

Applying Theorem 13.3, we should **choose $\boldsymbol{W}$ to make $\boldsymbol{V}_\beta$ as small as possible**. It turns out that the $\boldsymbol{W}$ that minimizes $\boldsymbol{V}_\beta$ is: $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$. Thus the **efficient GMM estimator** is:

$$\hat{\beta}_{\text{gmm}} = (\boldsymbol{X'Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z'X})^{-1}\boldsymbol{X'Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z'Y} \qquad (13.6')$$

In general, you should **use this $\widehat{\boldsymbol{\beta}}_{\textbf{gmm}}$** instead of the one in (13.6). In other words, **you should always set $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$**. This $\hat{\beta}_{\text{gmm}}$ also has the advantage of having a simpler asymptotic distribution. Replacing $\boldsymbol{W}$ with $\boldsymbol{\Omega}^{-1}$ in (13.7):

$$\boldsymbol{V}_\beta = (\boldsymbol{Q}'\,\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1}(\boldsymbol{Q}'\,\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\,\boldsymbol{\Omega}^{-1}\boldsymbol{Q})(\boldsymbol{Q}'\,\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1} = (\boldsymbol{Q}'\,\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1} \quad (13.7')$$

To summarize:

Theorem 13.4. Asymptotic Distribution of GMM for the IV model with Efficient Weight Matrix. Under Assumptions 12.2 and $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$, as $n \to \infty$

$$\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \underset{d}{\rightarrow} \text{N}(0, \boldsymbol{V}_{\beta}),$$

where $\boldsymbol{V}_{\beta} = (\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1}$.

Theorem 13.5. Efficient GMM for IV model. Under Assumptions 12.2, for any $\boldsymbol{W} > 0$ (any positive definite $\boldsymbol{W}$),

$$(\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q})^{-1}(\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{Q})(\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q})^{-1} - (\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1} \geq 0$$

Note: "≥ 0" means that the difference in these matrices is a positive semi-definite matrix. **This means that $\hat{\boldsymbol{\beta}}_{\text{gmm}}$ as defined in (13.6′) is the efficient GMM estimator**, and for any other GMM estimator (an estimate with a different $\boldsymbol{W}$ matrix), denoted by $\tilde{\beta}_{\text{gmm}}$:

$$\text{avar}[\hat{\beta}_{\text{gmm}}] \leq \text{avar}[\tilde{\beta}_{\text{gmm}}]$$

Since we do not know $\boldsymbol{\Omega}^{-1}$, we must estimate it. But for any consistent estimate of $\boldsymbol{\Omega}^{-1}$ Theorem 13.4 still holds.

Chamberlain (1987) showed that $\hat{\beta}_{\text{gmm}}$ as defined in (13.6′) is also efficient in a "semi-parametric" sense, in that it is efficient for all estimators based on the moment condition that $\text{E}[g_i(\beta)] = 0$ for any specified (known) $g(\beta)$ function.

Hansen also points out (Section 13.9) that a consequence of Theorem 13.5 is that $\widehat{\boldsymbol{\beta}}_{\mathbf{gmm}}$ (as defined in (13.6′)) **is, in general, more efficient than $\widehat{\boldsymbol{\beta}}_{\mathbf{2sls}}$**. However, if conditional homoscedasticity ($\mathrm{E}[e^2 | Z] = \sigma^2$) holds, then $\hat{\beta}_{2\text{sls}}$, which has a weighting matrix $\boldsymbol{W} = (\mathrm{E}[ZZ'])^{-1}$, is equal to $\hat{\beta}_{\text{gmm}}$.

Estimation of the Efficient Weight Matrix (IV case)

We rarely know $\boldsymbol{\Omega}^{-1}$, so we must estimate it. To do this, estimate $\boldsymbol{\Omega}$, denote the estimate by $\widehat{\boldsymbol{\Omega}}$, and then set $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Omega}}^{-1}$.

The **two-step GMM estimator** is implemented as follows. **First**, estimate $\hat{\beta}_{2\text{sls}}$; this is not efficient, but it is consistent. **Then** define $\tilde{e}_i = Y_i - X_i'\hat{\beta}_{2\text{sls}}$, $\tilde{g}_i = \tilde{g}_i(\tilde{\beta}) = Z_i\tilde{e}_i$, and $\bar{g}_n = (1/n)\sum_{i=1}^{n} \tilde{g}_i$. **Two possible estimators of $\boldsymbol{\Omega}$ are:**

$$\widehat{\boldsymbol{\Omega}} = (1/n)\sum_{i=1}^{n} \tilde{g}_i\tilde{g}_i' \qquad (13.8)$$

$$\widehat{\boldsymbol{\Omega}}^* = (1/n)\sum_{i=1}^{n}(\tilde{g}_i - \bar{g}_n)(\tilde{g}_i - \bar{g}_n)' \qquad (13.9)$$

$\widehat{\boldsymbol{\Omega}}$ is an **"uncentered"** estimator of $\boldsymbol{\Omega}$, and $\widehat{\boldsymbol{\Omega}}^*$ is a **"centered"** estimator of $\boldsymbol{\Omega}$. If $\mathrm{E}[Ze] = 0$, then: 1. $\boldsymbol{\Omega} = \mathrm{Var}(Ze)$; and 2. Both $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ are consistent estimates of $\boldsymbol{\Omega}$. But if $\mathrm{E}[Ze] \neq 0$, then $\widehat{\boldsymbol{\Omega}}^*$ still estimates $\mathrm{Var}(Ze)$ while $\widehat{\boldsymbol{\Omega}}$ estimates $\mathrm{E}[ZZ'e^2]$, which is $\neq \mathrm{Var}(Ze)$. Hansen says that this makes $\widehat{\boldsymbol{\Omega}}^*$ more "robust" than $\widehat{\boldsymbol{\Omega}}$. But if $\mathrm{E}[Ze] \neq 0$, than neither 2SLS nor GMM is consistent, so this robustness may be of little use. **Second**, use $\widehat{\boldsymbol{\Omega}}$ or $\widehat{\boldsymbol{\Omega}}^*$ to estimate $\hat{\beta}_{\text{gmm}}$ using equation (13.6′).

To summarize estimation of $\mathbf{\Omega}$:

Theorem 13.7. Under Assumption 12.2 and assuming $\mathbf{\Omega}$ is positive definite (> 0), if $\widehat{\boldsymbol{W}} = \widehat{\mathbf{\Omega}}^{-1}$ or $\widehat{\boldsymbol{W}} = \widehat{\mathbf{\Omega}}^{*-1}$, as defined in (13.8) and (13.9), then as $n \to \infty$

$$\sqrt{n}(\hat{\beta}_{\text{gmm}} - \beta) \underset{d}{\to} \mathrm{N}(0, \boldsymbol{V}_\beta)$$

where $\hat{\beta}_{\text{gmm}}$ is from equation (13.6′) and $\boldsymbol{V}_\beta = (\boldsymbol{Q}'\mathbf{\Omega}^{-1}\boldsymbol{Q})^{-1}$.

While this two-step GMM estimator is asymptotically efficient, it may not be efficient in finite samples and it may be better to iterate a few times. See Section 13.11 for a discussion, and the Stata command for "iterated GMM".

IV. Covariance Matrix Estimation for IV (13.12 – 13.14)

To apply Theorems 13.3 and 13.4, replace the (unknown) matrices in equations (13.7) or (13.7′) with their estimates. For the **one-step or two-step GMM estimators**, use:

$$\widehat{\boldsymbol{V}}_\beta = (\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}})^{-1}(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\mathbf{\Omega}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}})(\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Q}})^{-1} \qquad (13.10)$$

where $\widehat{\boldsymbol{Q}} = \frac{1}{n}\sum_{i=1}^{n} Z_i X_i'$. If you want to use the "centered" estimate for $\mathbf{\Omega}$, you can replace $\widehat{\mathbf{\Omega}}$ with $\widehat{\mathbf{\Omega}}^*$.

Efficient GMM estimation sets $\widehat{\boldsymbol{W}} = \widehat{\mathbf{\Omega}}^{-1}$, which gives:

$$\widehat{\boldsymbol{V}}_\beta = (\widehat{\boldsymbol{Q}}'\widehat{\mathbf{\Omega}}^{-1}\widehat{\boldsymbol{Q}})^{-1} \qquad (13.11)$$

For $\widehat{\mathbf{\Omega}}$, you can use (13.8), (13.9). or the "final" residuals $\hat{e}_i = Y_i - X_i'\hat{\beta}_{\text{gmm}}$. Lastly, the asymptotic std errors of $\hat{\beta}_{\text{gmm}}$ are square roots of diagonals of $\widehat{\boldsymbol{V}}_\beta/n$ (note the division by n!).

Equation (13.11) is also the estimator of the covariance matrix of the iterated GMM estimator.

Hansen says that using (13.9) instead of (13.8) should give a smaller covariance matrix, so this may be useful for finite samples (asymptotically there is no difference).

Clustered Standard Errors (IV model)

If you have a clustered (2-stage) sample design, you need to adjust equations (13.10) and (13.11) for clustering.

As in Chapter 12, the (structural) **equation for the g^{th} cluster can be written** (in matrix form) **as** $\boldsymbol{Y}_g = \boldsymbol{X}_g\beta + \boldsymbol{e}_g$. Subtracting β from the (centered) GMM estimator gives:

$$\hat{\beta}_{\text{gmm}} - \beta = (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZW}(\textstyle\sum_{g=1}^{G} \boldsymbol{Z}_g{}'\boldsymbol{e}_g)$$

The cluster-robust covariance matrix estimate for $\hat{\beta}_{\text{gmm}}$ is:

$$\widehat{\boldsymbol{V}}_\beta = (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZW}\widehat{\boldsymbol{S}}\boldsymbol{WZ'X}(\boldsymbol{X'ZWZ'X})^{-1} \qquad (13.12)$$

where

$$\widehat{\boldsymbol{S}} = \textstyle\sum_{g=1}^{G} \boldsymbol{Z}_g{}'\hat{\boldsymbol{e}}_g\hat{\boldsymbol{e}}_g{}'\boldsymbol{Z}_g \qquad (13.13)$$

$$\hat{\boldsymbol{e}}_g = \boldsymbol{Y}_g - \hat{\beta}_{\text{gmm}}\boldsymbol{X}_g \qquad (13.14)$$

This is what most economists use. Yet **Hansen says** that, **for clustered errors**, it is "more natural" to **use the cluster-robust weight matrix $\widehat{\boldsymbol{S}}$, not $\boldsymbol{W}$, to calculate $\hat{\boldsymbol{\beta}}_{\text{gmm}}$**:

$$\hat{\beta}_{\text{gmm}} = (\boldsymbol{X'Z}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{Z'X})^{-1}\boldsymbol{X'Z}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{Z'Y} \qquad (13.15)$$

This has the advantage of simplifying $\widehat{\boldsymbol{V}}_\beta$:

$$\widehat{\boldsymbol{V}}_\beta = (\boldsymbol{X'Z}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{Z'X})^{-1} \qquad (13.12')$$

In summary, start by using 2SLS to estimate $\hat{\beta}_{\text{2sls}}$. Insert $\hat{\beta}_{\text{2sls}}$ into (13.14) to estimate $\hat{\boldsymbol{e}}_g$, which is inserted in (13.13) to estimate $\widehat{\boldsymbol{S}}$. Hansen suggests using $\widehat{\boldsymbol{S}}$ to estimate $\hat{\beta}_{\text{gmm}}$ in (13.15), and then use (13.12′) to estimate $\widehat{\boldsymbol{V}}_\beta$.

Wald Test (IV model)

To estimate some **(possibly nonlinear) restrictions on $\boldsymbol{\beta}$**, write the q restrictions as $\boldsymbol{r(\beta) = \theta}$. The hypothesis to test is that θ is some value, denoted by θ_0. The GMM estimator of θ is $\hat{\theta} = r(\hat{\beta}_{\text{gmm}})$. By the delta method, $\hat{\theta}$ is asymptotically normal with a covariance matrix $\boldsymbol{V}_\theta = \boldsymbol{R'V}_\beta\boldsymbol{R}$, where $\boldsymbol{R} = \partial r(\beta)/\partial\beta'$. ($\boldsymbol{R}$ is $k\times q$, β has k elements, θ = set of q equations.)

The estimator of the **asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}$** is $\widehat{\boldsymbol{V}}_\theta = \widehat{\boldsymbol{R}}'\widehat{\boldsymbol{V}}_\beta\widehat{\boldsymbol{R}}$, with $\widehat{\boldsymbol{R}} = (\partial r(\hat{\beta}_{\text{gmm}})/\partial\beta)'$. For a single restriction on β, θ and $\widehat{\boldsymbol{V}}_\theta$ are scalars, so the std. error of $\hat{\theta}$ is $\sqrt{(1/n)\widehat{\boldsymbol{V}}_{\boldsymbol{\theta}}}$.

To test the null H_0: $\theta = \theta_0$, versus H_1: $\theta \neq \theta_0$, the **Wald test** is:

$$W = n(\hat{\theta} - \theta_0)'\hat{\boldsymbol{V}}_{\hat{\theta}}(\hat{\theta} - \theta_0)$$

Using $G_q(u)$ to denote the cumulative χ^2_q distribution:

Theorem 13.8. Under Assumptions 12.2 and 7.3 (that $r(\beta)$ is continuously differentiable and rank($\boldsymbol{R}$) = q), if H_0 holds, then as $n \to \infty$, $W \underset{d}{\to} \chi^2_q$. For the critical value c satisfying $\alpha = 1 - G_q(c)$, Prob[$W > c$| H_0] $\to \alpha$, so the test "Reject H_0 if $W > c$" has asymptotic size α.

V. Restricted GMM & Distance Test, IV model (13.15, 13.19)

Many **hypothesis tests** involve estimating restricted models and comparing these estimates to those of unrestricted models. GMM can be used to estimate restricted models. To start simple, consider only linear restrictions: $\boldsymbol{R}'\beta = \boldsymbol{c}$. The **constrained GMM estimator** minimizes the moment equations, denoted by $J(\beta)$, subject to this constraint:

$$\hat{\beta}_{\text{cgmm}} = \arg\min_{\boldsymbol{R}'\beta=\boldsymbol{c}} J(\beta)$$

The constrained estimation methods in Chapter 8 of Hansen can be applied here (**homework?**), which give:

$$\hat{\beta}_{\text{cgmm}} = \hat{\beta}_{\text{gmm}} - (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{R}(\boldsymbol{R}'(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{R})^{-1}(\boldsymbol{R}'\hat{\beta}_{\text{gmm}} - \boldsymbol{c}) \quad (13.16)$$

In Section 13.15 (see Theorem 13.19), Hansen works out the asymptotic distribution of $\hat{\beta}_{\text{cgmm}}$, which is very long and messy. Fortunately, **if we use the optimal choice for $\boldsymbol{W}$, that is $\widehat{\boldsymbol{\Omega}}^{-1}$, the expression for $\widehat{\boldsymbol{\beta}}_{\textbf{cgmm}}$ becomes**:

$$\hat{\beta}_{\text{cgmm}} = \hat{\beta}_{\text{gmm}} - \widehat{\boldsymbol{V}}_{\beta}\boldsymbol{R}(\boldsymbol{R}'\widehat{\boldsymbol{V}}_{\beta}\boldsymbol{R})^{-1}(\boldsymbol{R}'\hat{\beta}_{\text{gmm}} - \boldsymbol{c}) \qquad (13.19)$$

where $\widehat{\boldsymbol{V}}_{\beta} = (\widehat{\boldsymbol{Q}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{Q}})^{-1}$.

The asymptotic distribution is much simpler:

Theorem 13.10. Under Assumptions 12.2 and 8.3 ($\boldsymbol{R}'\beta = \boldsymbol{c}$ and rank($\boldsymbol{R}$) = q), the efficient constrained GMM estimator is asymptotically normally distributed as:

$$\sqrt{n}(\hat{\beta}_{\text{cgmm}} - \beta) \underset{d}{\rightarrow} \text{N}(0, \boldsymbol{V}_{\text{cgmm}})$$

as $n \rightarrow \infty$, where

$$\boldsymbol{V}_{\text{cgmm}} = \boldsymbol{V}_{\beta} - \boldsymbol{V}_{\beta}\boldsymbol{R}(\boldsymbol{R}'\boldsymbol{V}_{\beta}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{V}_{\beta} \qquad (13.20)$$

and $\boldsymbol{V}_{\beta} = (\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q})^{-1}$.

The asymptotic covariance matrix (13.20) is estimated by:

$$\widehat{\boldsymbol{V}}_{\text{cgmm}} = \widetilde{\boldsymbol{V}}_{\beta} - \widetilde{\boldsymbol{V}}_{\beta}\boldsymbol{R}(\boldsymbol{R}'\widetilde{\boldsymbol{V}}_{\beta}\boldsymbol{R})^{-1}\boldsymbol{R}'\widetilde{\boldsymbol{V}}_{\beta} \qquad (13.21)$$

$$\widetilde{\boldsymbol{V}}_{\beta} = (\widehat{\boldsymbol{Q}}\widetilde{\boldsymbol{\Omega}}\widehat{\boldsymbol{Q}})^{-1}$$

$$\widetilde{\mathbf{\Omega}} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i{}' \, \tilde{e}_i^2 \qquad (13.22)$$

$$\tilde{e}_i = Y_i - X_i'\hat{\beta}_{\text{cgmm}}$$

It is also **possible to estimate GMM models subject to nonlinear constraints**. In this case the solution usually does not have an explicit solution, but must be found by iteration. This is discussed in Section 13.16 of Hansen.

Section 13.17 of Hansen shows how to apply GMM to constrained OLS regressions when E[Xe] = 0.

Section 13.18 shows how GMM can be applied to systems of equations.

The GMM Distance Test (General case)

The GMM distance test is an **alternative to the Wald test**. **Asymptotically, they are equivalent**, so use the one that is most convenient. The GMM distance test is **convenient for nonlinear constraints**, and for other cases, as seen below.

The GMM distance test is similar to the likelihood ratio test used in maximum likelihood estimation.

The Wald test compares (a function of) the estimated β, $\hat{\beta}_{\text{gmm}}$, to (a function of) a proposed β. Intuitively, it checks whether (a function of) $\hat{\beta}_{\text{gmm}}$ is "close" to (a function of) β.

In contrast, the **GMM distance test checks** whether the **value of** the "criterion function" that one is minimizing, which is ***J(β)***, when estimated **under the constraint**, is "close" to the criterion function **without the constraint**.

Recall that the unrestricted GMM estimator is defined as:

$$\hat{\beta}_{\text{gmm}} = \underset{\beta}{\operatorname{argmin}}\, J(\beta)$$

Where (compare to page 4)

$$\hat{J}(\beta) = n\bar{g}_n(\beta)'\widehat{\boldsymbol{\Omega}}^{-1}\bar{g}_n(\beta)$$

Consider minimizing $J(\beta)$ subject to the (possibly nonlinear) constraint $r(\beta) = \theta_0$.

$$\hat{\beta}_{\text{cgmm}} = \arg \min_{r(\beta)=\theta_0} \tilde{J}(\beta)$$

where $\tilde{J}(\beta) = n\bar{g}_n(\beta)'\widetilde{\boldsymbol{\Omega}}^{-1}\bar{g}_n(\beta)$, where $\widetilde{\boldsymbol{\Omega}}^{-1}$ could be $\widehat{\boldsymbol{\Omega}}^{-1}$ or something else (see below).

The **GMM distance statistic**, D, is defined $\boldsymbol{D = \tilde{J}(\beta) - \hat{J}(\beta)}$.

Theorem 13.12. Under Assumptions 12.2 (so IV case) and 7.3, and H_0 ($r(\beta) = \theta_0$), as $n \to \infty$ then $D \underset{d}{\to} \chi^2_q$. For critical value c satisfying $\alpha = 1 - G_q(c)$,

$$\text{Prob}[D > c |\, H_0] \underset{p}{\to} \alpha$$

so the test “Reject H_0 if $D > c$ has asymptotic size α.

Note that this is the same asymptotic distribution as the **Wald test**, so they are **asymptotically equivalent**.

So what matrix should we use for $\widetilde{\boldsymbol{\Omega}}$ in $\tilde{J}(\beta)$? One choice is $\widehat{\boldsymbol{\Omega}}$ (from unconstrained estimation), which leads to:

Theorem 13.13. If $\widetilde{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}$, then $D \geq 0$, and if r is a linear function of β then D is exactly equal to the Wald statistic.

$D \geq 0$ follows from the fact that $\widetilde{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Omega}}$ then the two criterion function, $\tilde{J}(\beta)$ and $\hat{J}(\beta)$, are the same functions, and the constrained minimum $\tilde{J}(\beta)$ is likely larger than the unconstrained minimum $\hat{J}(\beta)$, and it cannot be smaller.

Hansen argues that a “natural” choice for $\widetilde{\boldsymbol{\Omega}}$ is to use the version in equation (13.22), but he gives no “proof” that this is better. He also notes that the Wald test “works quite poorly” for nonlinear hypotheses, so use the D test.

VI. OverID Test (Hansen’s *J*-test) for IV model (13.21, 13.22)

Recall (Apec 8212) that the **Sargan overID test** for 2SLS estimation requires the **assumption of homoscedasticity**. **GMM provides an overID test** that is **robust to heteroscedasticity**. We want to test the null hypothesis H_0: $E[Ze] = 0$.

Recall also that $\bar{g}_n(\beta) \underset{p}{\rightarrow} \mathrm{E}[Ze]$, and the criterion function is:

$$J(\hat{\beta}_{\mathrm{gmm}}) = n\bar{g}_n(\hat{\beta}_{\mathrm{gmm}})'\widehat{\mathbf{\Omega}}^{-1}\bar{g}_n(\hat{\beta}_{\mathrm{gmm}})$$

Thus $\hat{J}(\beta)$ will, in general, $= 0$ if $\mathrm{E}[Ze] = 0$ and $\neq 0$ if $\mathrm{E}[Ze] \neq 0$. This leads to:

Theorem 13.14. Under Assumption 12.2, and assuming H_0: $\mathrm{E}[Ze] = 0$, then as $n \rightarrow \infty$, $J(\hat{\beta}_{\mathrm{gmm}}) \underset{d}{\rightarrow} \chi^2_{\ell-k}$. For critical value c satisfying $\alpha = 1 - G_{\ell-k}(c)$, $\mathrm{Prob}[J(\hat{\beta}_{\mathrm{gmm}}) > c] \rightarrow \alpha$, so the test "Reject H_0 if $J(\hat{\beta}_{\mathrm{gmm}}) > c$" has asymptotic size α.

The degrees of freedom, $\ell - k$, is the number of over-identifying restrictions. Hansen recommends (and I agree): **always report this test statistic any time you use GMM and $\boldsymbol{\ell > k}$.**

Subset OverID Tests

The above overID test is a joint test of *all* the moment restrictions. Sometimes you may want to test only *some* of them. For example, for 2SLS you may be sure that some of your IVs are uncorrelated with e but you are not sure about other IVs and you want to test only the latter.

For 2SLS, divide the IVs, Z, into Z_a (ℓ_a IVs) that you are sure are exogenous and Z_b (ℓ_b IVs) that you want to test.

The "maintained hypothesis" is E[$Z_a e$] = 0, while H_0 is E[$Z_b e$] = 0 and H_1 is E[$Z_b e$] $\neq$ 0.

This test can be **done as follows**. **First**, estimate the model using efficient GMM **using only the Z_a IVs**. Save the estimated criterion function, which can be denoted as $\tilde{J}$. **Second**, estimate the model using efficient GMM **using all IVs** (Z_a and Z_b). The estimated criterion function can be denoted as $\hat{J}$. The test statistic is $C = \hat{J} - \tilde{J} \underset{d}{\rightarrow} \chi^2_{l_b}$. This is for the case where $\ell_a > k$; if $\ell_a = k$ then use the overall *J*-test ($\tilde{J} = 0$). Hansen summarizes this in Theorem 13.15.

VII. Endogeneity Test for IV models (13.23, 13.24)

In GMM estimation, endogeneity tests are a type of over-ID test. **Consider 2SLS estimation**, where the model is $Y = Z_1'\beta_1 + Y_2'\beta_2 + e$. We want to test whether Y_2 is endogenous. We "know" that Z_1 is exogenous, so the "maintained hypothesis" is E[$Z_1 e$] = 0. The null hypothesis is H_0: E[$Y_2 e$] = 0. The alternative hypothesis is H_1: E[$Y_2 e$] $\neq$ 0.

To estimate the model under H_1 the only valid IVs are Z_1 and Z_2 (Z_2 = excluded instruments), so the only moment conditions are that Z_1 and Z_2 are uncorrelated with e. **Under H_0**, Y_2 are also uncorrelated with e and so they are added as IVs in the sense that we **add the additional moment condition E[$Y_2 e$] = 0** for estimation. Under H_0, the J statistics from these two estimates should be equal, but **under H_1 adding the moment condition $E[Y_2 e] = 0$ will increase** the value of **the J statistic**.

To implement this test, first estimate the model with only Z_1 and Z_2 as IVs (moment conditions $E[Z_1 e] = E[Z_2 e] = 0$). Save the value of J, denoted by $\tilde{J}$. Second, estimate the model again with Z_1, Z_2 and Y_2 as IVs (moment conditions $E[Z_1 e] = E[Z_2 e] = E[Y_2 e] = 0$). This value of J is denoted by $\hat{J}$. The test statistic is $C = \hat{J} - \tilde{J}$.

Theorem 13.16. Under Assumption 12.2 and rank($E[Z_2 Y_2']$) $= k_2$, then as $n \to \infty$, $C = \hat{J} - \tilde{J} \underset{d}{\to} \chi^2_{k_2}$. For critical value c satisfying $\alpha = 1 - G_{k_2}(c)$,

$$\text{Prob}[C = \hat{J} - \tilde{J} > c] \to \alpha$$

so the test "Reject H_0 if $C = \hat{J} - \tilde{J} > c$" has asymptotic size α.

This **can be extended to testing a subset of the endogenous variables**. This is Theorem 13.17 in Section 13.24.

Section 13.25 in Hansen provides an introduction to GMM estimate for nonlinear models. See Chapter 14 of Wooldridge (2010), *Econometric Analysis of Cross Section and Panel Data* for a much more detailed treatment.

Section 13.26 discusses bootstrapping for GMM. Do not bootstrap standard errors; bootstrap confidence intervals.