December 19, 2018

Generalized Method of Moments

Definition

Suppose that an econometrician observes the data $\{\boldsymbol{W}_i : i = 1, \ldots, n\}$ where $\boldsymbol{W}_i$ is a random p-vector. Let $\boldsymbol{g} = (g_1, ..., g_l)'$ be a l dimensional function depending on $\boldsymbol{W}_i$ and the k-vector of parameters $\boldsymbol{b}$:

$$\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) = \begin{pmatrix} g_1(\boldsymbol{W}_i, \boldsymbol{b}) \\ \vdots \\ g_l(\boldsymbol{W}_i, \boldsymbol{b}) \end{pmatrix},$$

and $g_j : \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}$ for $j = 1, \ldots, l$. The model is defined by the following *moment condition.*

$$\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\beta}) = \boldsymbol{0} \text{ for some } \boldsymbol{\beta} \in \mathbb{R}^k. \tag{1}$$

Examples:

- **Linear regression.** Let $\boldsymbol{W}_i = (Y_i, \boldsymbol{X}_i')'$ and $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$, where $\boldsymbol{\beta} \in \mathbb{R}^k$, and $\mathbb{E}(\boldsymbol{X}_i e_i) = \boldsymbol{0}$. In this case, $\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) = \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{b})$, $l = k$, and the moment condition is $\mathbb{E}(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})) = \boldsymbol{0}$.

- **IV regression.** Let $\boldsymbol{W}_i = (Y_i, \boldsymbol{X}_i', \boldsymbol{Z}_i')'$, $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$, where $\boldsymbol{\beta} \in \mathbb{R}^k$, and $\mathbb{E}(\boldsymbol{Z}_i e_i) = \boldsymbol{0}$, where $\boldsymbol{Z}_i$ is a l-vector. In this case, $\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) = \boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{b})$ with the moment condition $\mathbb{E}(\boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})) = \boldsymbol{0}$.

- **Lucas' Model.** Suppose that in period t investors receive utility from the consumption C_t be consumption in period t. Let $R_{j,t}$ be the rate of return on the risky asset j. Suppose that there are m assets. Assume that the utility function is of the form $\sum_{t=1}^{\infty} \delta^t C_t^{1-\alpha} / (1-\alpha)$. In the equilibrium, the returns on risky assets are determined by the following Euler equations:

  $$\mathbb{E}\left(\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\alpha}(1 + R_{j,t+1})\right) = 1 \text{ for } j = 1, \ldots, m.$$

  In this case we have $\boldsymbol{W}_t = (C_t, R_{1,t} \ldots, R_{m,t})$, $\boldsymbol{b} = (a, d)$, $g_j(\boldsymbol{W}_t, \boldsymbol{b}) = d\left(\frac{C_{t+1}}{C_t}\right)^{-a}(1 + R_{j,t+1}) - 1$ for $j = 1, \ldots, m$, and the moment conditions given by the above equations. Note that in this case $\boldsymbol{g}$ is nonlinear in the parameters.

We say that the model is *identified* if $\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\beta}) = \boldsymbol{0}$ and $\mathbb{E}\boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}\right) = \boldsymbol{0}$ imply that $\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$, i.e. the solution of (1) is unique. The moment condition gives us l restrictions for k parameters. A necessary condition for the model to be identified is that $l \geq k$, i.e. we must have *at least* k restrictions. The necessary condition is called the *order condition.* We say that the model is not identified or *underidentified* if the order condition fails.

When $k = l$, applying the method of moments (MM) principle, we can estimate $\boldsymbol{\beta}$ by the value of $\boldsymbol{b}$ that solves the sample analogue of (1):

$$n^{-1} \sum_{i=1}^{n} \boldsymbol{g}\left(\boldsymbol{W}_i, \widehat{\boldsymbol{\beta}}_n^{MM}\right) = \mathbf{0}.$$

However, when $l > k$, in general, there is no $\boldsymbol{b} \in \mathbb{R}^k$ that solves all l equations exactly. In this case, we can choose the value of $\boldsymbol{b}$ that makes the sample moments as close to zero as possible. Let $\boldsymbol{A}_n$ be a (possibly random) $l \times l$ weight matrix such that $\boldsymbol{A}_n \to_p \boldsymbol{A}$, where $\boldsymbol{A}$ is non-random and has full rank (l). The *Generalized Method of Moments (GMM) estimator* of $\boldsymbol{\beta}$ is defined to be the value of $\boldsymbol{b}$ that minimizes the weighted distance of $n^{-1} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b})$ from zero:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_n^{GMM} &= \underset{\boldsymbol{b} \in \Theta}{\operatorname{argmin}} \left\| \boldsymbol{A}_n n^{-1} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) \right\|^2 \\ &= \underset{\boldsymbol{b} \in \Theta}{\operatorname{argmin}} \left(n^{-1} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) \right)' \boldsymbol{A}_n' \boldsymbol{A}_n \left(n^{-1} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{b}) \right). \end{aligned} \tag{2}$$

The set $\Theta \subset \mathbb{R}^k$ is usually assumed to be compact. Note that $\boldsymbol{A}'\boldsymbol{A}$ is positive definite.

Linear case

In this section, we discuss in details the IV regression example. Note that in this case, the function $\boldsymbol{g}$ is linear in parameters. We assume that some or all of the k regressors in $\boldsymbol{X}_i$ are endogenous:

$$\mathbb{E}(\boldsymbol{X}_i e_i) \neq \mathbf{0},$$

and that the l instruments $\boldsymbol{Z}_i$ are weakly exogenous:

$$\mathbb{E}(\boldsymbol{Z}_i e_i) = \mathbf{0}.$$

The model is identified, if the following *rank condition* is satisfied:

$$\operatorname{rank}\left(\mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right)\right) = k.$$

If the rank condition is satisfied and $l = k$ we say that the model is *exactly* or *just identified.* We say that the model is *overidentified* if the rank condition is satisfied and $l > k$ (there are more instruments than the parameters that want to estimate). We allow here the model to be overidentified.

In the linear IV regression case, $\widehat{\boldsymbol{\beta}}_n^{GMM}$ is the minimizer of

$$\left(n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left(Y_i - \boldsymbol{X}_i' \boldsymbol{b}\right) \right)' \boldsymbol{A}_n' \boldsymbol{A}_n \left(n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left(Y_i - \boldsymbol{X}_i' \boldsymbol{b}\right) \right)$$

as a function of $\boldsymbol{b}$ and given by the following expression:

$$\widehat{\boldsymbol{\beta}}_n^{GMM} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n\right) \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{X}_i' \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n\right) \sum_{i=1}^{n} \boldsymbol{Z}_i Y_i.$$

We will show next that the GMM estimator is consistent. We need the following assumptions.

- $\{(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i) : i = 1, \ldots, n\}$ are iid. $\boldsymbol{X}_i = (X_{i,1}, ..., X_{i,k})'$. $\boldsymbol{Z}_i = (Z_{i,1}, ..., Z_{i,l})'$.
- $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$, where $\boldsymbol{\beta} \in \mathbb{R}^k$.
- $\mathbb{E}(\boldsymbol{Z}_i e_i) = \mathbf{0}$.
- $\mathbb{E}(\boldsymbol{Z}_i \boldsymbol{X}_i')$ has rank k.
- $\boldsymbol{A}_n \to_p \boldsymbol{A}$, where $\boldsymbol{A}$ has rank $l \geq k$.
- $\mathbb{E}X_{i,j}^2 < \infty$ for all $j = 1, \ldots, k$.
- $\mathbb{E}Z_{i,j}^2 < \infty$ for all $j = 1, \ldots, l$.

Write

$$\widehat{\boldsymbol{\beta}}_n^{GMM} = \boldsymbol{\beta} + \left(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n \right) n^{-1} \sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{X}_i' \right)^{-1} n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n \right) n^{-1} \sum_{i=1}^n \boldsymbol{Z}_i e_i.$$

The last two of the above assumptions imply that

$$\mathbb{E}\left| X_{i,r} Z_{i,s} \right| < \infty \text{ for all } r = 1, \ldots, k \text{ and } s = 1, \ldots, l.$$

By the WLLN,

$$n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \to_p \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{Z}_i' \right).$$

Since $\boldsymbol{A}_n \to_p \boldsymbol{A}$, we also have that

$$n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n \right) n^{-1} \sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{X}_i' \to_p \mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{Z}_i' \right) \left(\boldsymbol{A}' \boldsymbol{A} \right) \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i' \right).$$

Further, since $\mathbb{E}(\boldsymbol{Z}_i \boldsymbol{X}_i')$ has rank k, A has rank $l \geq k$, it follows that the $k \times k$ matrix $\mathbb{E}(\boldsymbol{X}_i \boldsymbol{Z}_i') (\boldsymbol{A}'\boldsymbol{A}) \mathbb{E}(\boldsymbol{Z}_i \boldsymbol{X}_i')$ has full rank k and, therefore, invertible. Consequently, by the Slutsky's Theorem,

$$\left(n^{-1} \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n \right) n^{-1} \sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{X}_i' \right)^{-1} \to_p \left(\mathbb{E}\left(\boldsymbol{X}_i \boldsymbol{Z}_i' \right) \left(\boldsymbol{A}' \boldsymbol{A} \right) \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i' \right) \right)^{-1}.$$

Next, by the WLLN,

$$n^{-1} \sum_{i=1}^n \boldsymbol{Z}_i e_i \to_p \mathbf{0},$$

and thus $\widehat{\boldsymbol{\beta}}_n^{GMM} \to_p \boldsymbol{\beta}$.

In order to show asymptotic normality, we will need the following two assumptions in addition to the above.

- $\mathbb{E}Z_{i,j}^4 < \infty$ for all $j = 1, \ldots, l$.
- $\mathbb{E}e_i^4 < \infty$.

- $\mathbb{E}\left(e_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'\right)$ is positive definite.

Write

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n^{GMM} - \boldsymbol{\beta}\right) = \left(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n\right) n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{X}_i'\right)^{-1} n^{-1}\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{Z}_i' \left(\boldsymbol{A}_n' \boldsymbol{A}_n\right) n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i e_i.$$

The last two assumptions imply that the variance of $\boldsymbol{Z}_i e_i$, $\mathbb{E}\left(e_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'\right)$ is finite, and, by the CLT, we have that

$$n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i e_i \to_d \mathrm{N}\left(\boldsymbol{0}, \mathbb{E}\left(e_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'\right)\right).$$

Let's define

$$\begin{aligned} \boldsymbol{Q}_{\boldsymbol{ZX}} &= \mathbb{E}\left(\boldsymbol{Z}_i \boldsymbol{X}_i'\right), \\ \boldsymbol{\Omega} &= \mathbb{E}\left(e_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'\right). \end{aligned}$$

Combining the above results, we have

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n^{GMM} - \boldsymbol{\beta}\right) \to_d \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right),$$

where $\boldsymbol{V}_{\boldsymbol{\beta}}$ takes the sandwich form:

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}_{\boldsymbol{ZX}}' \boldsymbol{A}' \boldsymbol{A} \boldsymbol{Q}_{\boldsymbol{ZX}}\right)^{-1} \boldsymbol{Q}_{\boldsymbol{ZX}}' \boldsymbol{A}' \boldsymbol{A} \boldsymbol{\Omega} \boldsymbol{A}' \boldsymbol{A} \boldsymbol{Q}_{\boldsymbol{ZX}} \left(\boldsymbol{Q}_{\boldsymbol{ZX}}' \boldsymbol{A}' \boldsymbol{A} \boldsymbol{Q}_{\boldsymbol{ZX}}\right)^{-1}.$$

The variance-covariance matrix $\boldsymbol{V}_{\boldsymbol{\beta}}$ can be estimated by replacing $\boldsymbol{A}$, $\boldsymbol{Q}_{\boldsymbol{ZX}}$ and $\boldsymbol{\Omega}$ with their consistent estimators $\boldsymbol{A}_n$, $\widehat{\boldsymbol{Q}}_n$ and $\widehat{\boldsymbol{\Omega}}_n$ respectively, where

$$\begin{aligned} \widehat{\boldsymbol{Q}}_n &= n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i \boldsymbol{X}_i', \\ \widehat{\boldsymbol{\Omega}}_n &= n^{-1}\sum_{i=1}^n \widehat{e}_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i', \end{aligned}$$

where $\widehat{e}_i = Y_i - \boldsymbol{X}_i' \widehat{\boldsymbol{\beta}}_n^{GMM}$.

General case

In the general case, the GMM estimator minimizes the nonlinear function in (2). Usually, we do not have a closed-form expression for $\widehat{\boldsymbol{\beta}}_n^{GMM}$, and the minimization must be done using numerical procedures. Nevertheless, under general regularity conditions, it is possible to show that $\widehat{\boldsymbol{\beta}}_n^{GMM}$ is consistent and asymptotically normal. We will only provide heuristic proofs of consistency and asymptotic normality.

Since the criterion function in (2) involves averages, we should expect that

$$\left\| \boldsymbol{A}_n n^{-1}\sum_{i=1}^n \boldsymbol{g}\left(\boldsymbol{W}_i, \boldsymbol{b}\right)\right\|^2 \to_p \left\|\boldsymbol{A}\mathbb{E}\boldsymbol{g}\left(\boldsymbol{W}_i, \boldsymbol{b}\right)\right\|^2. \tag{3}$$

Assuming that the model is uniquely identified, $\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{b})=\mathbf{0}$ if and only if $\boldsymbol{b}=\boldsymbol{\beta}$. Since $\|\boldsymbol{A}\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{b})\|^2>0$ for all $\boldsymbol{b}\neq\boldsymbol{\beta}$, the true value $\boldsymbol{\beta}$ is the unique minimizer of $\|\boldsymbol{A}\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{b})\|^2$. Intuitively, $\widehat{\boldsymbol{\beta}}_n^{GMM}$ is consistent because

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_n^{GMM} &= \underset{\boldsymbol{b}\in\Theta}{\operatorname{argmin}}\left\|\boldsymbol{A}_n n^{-1}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{b})\right\|^2\\ &\to_p \underset{\boldsymbol{b}\in\Theta}{\operatorname{argmin}}\|\boldsymbol{A}\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{b})\|^2\\ &= \boldsymbol{\beta}.\end{aligned}$$

The formal proof of consistency requires a number of regularity conditions, such as *uniform* in $\boldsymbol{b}$ convergence in (3), compactness of Θ, $\boldsymbol{\beta}$ being the interior point of Θ.

For asymptotic normality, note that $\widehat{\boldsymbol{\beta}}_n^{GMM}$ solves the first-order conditions:

$$\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}\right)'\boldsymbol{A}_n'\boldsymbol{A}_n n^{-1}\sum_{i=1}^n \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)=\mathbf{0}. \tag{4}$$

(In fact, it is sufficient if $\widehat{\boldsymbol{\beta}}_n^{GMM}$ solves the first-order conditions approximately, i.e. on the right-hand side of the above equation, instead of zero, we can have a term that goes to zero in probability at the rate $n^{1/2}$.) Next, using the expansion of $\boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)$ around $\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})$ (the element-by-element mean value theorem), we obtain

$$\boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)=\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})+\frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{*}\right)}{\partial \boldsymbol{b}'}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}-\boldsymbol{\beta}\right), \tag{5}$$

where $\widehat{\boldsymbol{\beta}}_n^{*}$ is between $\widehat{\boldsymbol{\beta}}_n^{GMM}$ and $\boldsymbol{\beta}$. Substitution of (5) into (4) gives

$$\begin{aligned}\mathbf{0} &= \left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}\right)'\boldsymbol{A}_n'\boldsymbol{A}_n n^{-1}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})\\ &\quad+\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}\right)'\boldsymbol{A}_n'\boldsymbol{A}_n\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{*}\right)}{\partial \boldsymbol{b}'}\right)\left(\widehat{\boldsymbol{\beta}}_n^{GMM}-\boldsymbol{\beta}\right),\end{aligned}$$

We can write

$$\begin{aligned}&n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}-\boldsymbol{\beta}\right)\\ &= -\left(\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}\right)'\boldsymbol{A}_n'\boldsymbol{A}_n\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{*}\right)}{\partial \boldsymbol{b}'}\right)\right)^{-1}\\ &\quad\times\left(n^{-1}\sum_{i=1}^n \frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}\right)'\boldsymbol{A}_n'\boldsymbol{A}_n n^{-1/2}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta}).\end{aligned}$$

Since $\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\beta}) = \mathbf{0}$, we should expect that, under some regularity conditions,

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta}) \rightarrow_d \mathrm{N}\left(\mathbf{0}, \mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})\,\mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})'\right).$$

(Note that the asymptotic variance depends on the unknown $\boldsymbol{\beta}$). Since $\widehat{\boldsymbol{\beta}}_n^{GMM}$ is consistent, and, as a result $\widehat{\boldsymbol{\beta}}_n^* \rightarrow_p \boldsymbol{\beta}$ as well, we should expect, that under some proper regularity conditions,

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'} &\rightarrow_p \mathbb{E}\left(\frac{\partial g(\boldsymbol{W}_i,\boldsymbol{\beta})}{\partial \boldsymbol{b}'}\right), \\
n^{-1}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{*}\right)}{\partial \boldsymbol{b}'} &\rightarrow_p \mathbb{E}\left(\frac{\partial g(\boldsymbol{W}_i,\boldsymbol{\beta})}{\partial \boldsymbol{b}'}\right),
\end{aligned}$$

and that the matrix

$$\left(\mathbb{E}\left(\frac{\partial g(\boldsymbol{W}_i,\boldsymbol{\beta})}{\partial \boldsymbol{b}'}\right)\right)'\boldsymbol{A}'\boldsymbol{A}\left(\mathbb{E}\left(\frac{\partial g(\boldsymbol{W}_i,\boldsymbol{\beta})}{\partial \boldsymbol{b}'}\right)\right)$$

is invertible. Then,

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n^{GMM} - \boldsymbol{\beta}\right) \rightarrow_d \mathrm{N}(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}),$$

where

$$\begin{aligned}
\boldsymbol{V}_{\boldsymbol{\beta}} &= (\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q})^{-1}, \\
\boldsymbol{Q} &= \mathbb{E}\left(\frac{\partial g(\boldsymbol{W}_i,\boldsymbol{\beta})}{\partial \boldsymbol{b}'}\right), \\
\boldsymbol{\Omega} &= \mathbb{E}\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})\,\boldsymbol{g}(\boldsymbol{W}_i,\boldsymbol{\beta})'.
\end{aligned}$$

The variance-covariance matrix $\boldsymbol{V}_{\boldsymbol{\beta}}$ can be estimated by replacing $\boldsymbol{A}$, $\boldsymbol{Q}$ and $\boldsymbol{\Omega}$ with their consistent estimators $\boldsymbol{A}_n$ and

$$\begin{aligned}
\widehat{\boldsymbol{Q}}_n &= n^{-1}\sum_{i=1}^{n}\frac{\partial \boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'}, \\
\widehat{\boldsymbol{\Omega}}_n &= n^{-1}\sum_{i=1}^{n}\boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)\boldsymbol{g}\left(\boldsymbol{W}_i,\widehat{\boldsymbol{\beta}}_n^{GMM}\right)'.
\end{aligned}$$

Efficient GMM

The GMM estimator depends on the choice of the weight matrix $\boldsymbol{A}_n$. The efficient GMM estimator is the one that has the smallest asymptotic variance among all GMM estimators (defined by different choices of $\boldsymbol{A}_n$). Next, we will show that the efficient GMM corresponds to $\boldsymbol{A}_n$ such that

$$\boldsymbol{A}_n'\boldsymbol{A}_n \rightarrow_p \boldsymbol{\Omega}^{-1}.$$

Theorem 1. *(a) A lower bound for the asymptotic variance of the class of GMM estimators indexed by* $\boldsymbol{A}_n$ *is given by* $\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}$.

(b) The lower bound is achieved if $\boldsymbol{A}_n'\boldsymbol{A}_n \rightarrow_p \boldsymbol{\Omega}^{-1}$.

Proof. In order to prove part (a), we need to show that

$$\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1} - \left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}$$

is negative semi-definite for any $\boldsymbol{A}$ that has rank l. Equivalently, we can show that

$$\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q} - \boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q} \tag{6}$$

is positive semi-definite.

Since the inverse of $\boldsymbol{\Omega}$ exists ($\boldsymbol{\Omega}$ is positive definite), we can write

$$\boldsymbol{\Omega}^{-1} = \boldsymbol{C}'\boldsymbol{C},$$

where $\boldsymbol{C}$ is invertible as well. Write (6) as

$$\begin{aligned}
&\boldsymbol{Q}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{Q} - \boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{C}^{-1}\left(\boldsymbol{C}'\right)^{-1}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q} \\
= \; &\boldsymbol{Q}'\boldsymbol{C}'\left(\boldsymbol{I}_l - \left(\boldsymbol{C}'\right)^{-1}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{C}^{-1}\left(\boldsymbol{C}'\right)^{-1}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{C}^{-1}\right)\boldsymbol{C}\boldsymbol{Q}.
\end{aligned} \tag{7}$$

Define

$$\boldsymbol{H} = \left(\boldsymbol{C}'\right)^{-1}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q},$$

and note that, using this definition, (7) becomes

$$\boldsymbol{Q}'\boldsymbol{C}'\left(\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right)\boldsymbol{C}\boldsymbol{Q}.$$

The above matrix is positive semi-definite if $\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'$ is positive semi-definite. Next,

$$\begin{aligned}
&\left(\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right)\left(\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right) \\
= \; &\boldsymbol{I}_l - 2\boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}' + \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}' \\
= \; &\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'.
\end{aligned}$$

Therefore, $\boldsymbol{I}_l - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'$ is idempotent and, consequently, positive semi-definite. This completes the proof of part (a).

For part (b), if $\boldsymbol{A}_n'\boldsymbol{A}_n \to_p \boldsymbol{A}'\boldsymbol{A} = \boldsymbol{\Omega}^{-1}$, then the asymptotic variance becomes

$$\begin{aligned}
&\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1} \\
= \; &\left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}.
\end{aligned}$$

A natural choice for such $\boldsymbol{A}_n'\boldsymbol{A}_n$ is $\widehat{\boldsymbol{\Omega}}_n^{-1}$. This suggests the following *two-step* procedure:

1. Set $\boldsymbol{A}_n'\boldsymbol{A}_n = \boldsymbol{I}_l$. Obtain the corresponding (inefficient) estimates of $\boldsymbol{\beta}$, say $\widetilde{\boldsymbol{\beta}}_n$. Using the inefficient (but consistent) estimator of $\boldsymbol{\beta}$, obtain $\widehat{\boldsymbol{\Omega}}_n$. For example, in the linear case,

$$\widehat{\boldsymbol{\Omega}}_n \;=\; n^{-1}\sum_{i=1}^{n}\widehat{e}_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i', \text{ where}$$

$$\widehat{e}_i = Y_i - \boldsymbol{X}_i'\widetilde{\boldsymbol{\beta}}_n,$$

and, in the general case,

$$\widehat{\boldsymbol{\Omega}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}_n\right)\boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}_n\right)'.$$

2. Obtain the efficient GMM estimates of $\boldsymbol{\beta}$ by minimizing

$$\left(n^{-1}\sum_{i=1}^n \boldsymbol{g}\left(\boldsymbol{W}_i, \boldsymbol{b}\right)\right)' \widehat{\boldsymbol{\Omega}}_n^{-1}\left(n^{-1}\sum_{i=1}^n \boldsymbol{g}\left(\boldsymbol{W}_i, \boldsymbol{b}\right)\right),$$

   where $\widehat{\boldsymbol{\Omega}}_n$ comes from the first step.

An alternative to $\widehat{\boldsymbol{\Omega}}_n$ in the first step is

$$n^{-1}\sum_{i=1}^n\left(\boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}_n\right) - n^{-1}\sum_{j=1}^n \boldsymbol{g}\left(\boldsymbol{W}_j, \widetilde{\boldsymbol{\beta}}_n\right)\right)\left(\boldsymbol{g}\left(\boldsymbol{W}_i, \widetilde{\boldsymbol{\beta}}_n\right) - n^{-1}\sum_{j=1}^n \boldsymbol{g}\left(\boldsymbol{W}_j, \widetilde{\boldsymbol{\beta}}_n\right)\right)',$$

the centered version of $\widehat{\boldsymbol{\Omega}}_n$. The two versions are asymptotically equivalent, since $\mathbb{E}\boldsymbol{g}\left(\boldsymbol{W}_i, \boldsymbol{\beta}\right) = \boldsymbol{0}$. However, the centered version often performs better.

In the linear case, a better choice for the first stage weight matrix is

$$\begin{aligned} \boldsymbol{A}_n'\boldsymbol{A}_n &= \left(\sum_{i=1}^n \boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1} \\ &= \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}. \end{aligned} \tag{8}$$

The reason for this become clear in the next section.

The variance-covariane matrix of the efficient GMM estimator can be estimated consistently by

$$\left(\widehat{\boldsymbol{Q}}_n'\widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{Q}}_n\right)^{-1}.$$

Two-stage Least Squares (2SLS)

Consider the linear IV regression model, and assume that

$$\mathbb{E}\left(e_i^2|\boldsymbol{Z}_i\right) = \sigma^2. \tag{9}$$

In this case,

$$\begin{aligned} \boldsymbol{\Omega} &= \mathbb{E}\left(e_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(e_i^2|\boldsymbol{Z}_i\right)\boldsymbol{Z}_i\boldsymbol{Z}_i'\right) \\ &= \sigma^2\mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right). \end{aligned}$$

A natural estimator of $\mathbb{E}\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)$ is

$$n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i',$$

which gives the optimal weight matrix as in (8). Note that, in this case, the efficient GMM estimator can be obtained without the first step, since the weight matrix in (8) does not depend on $\widehat{e}_i$'s. The efficient GMM is given by

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_n^{2SLS} &= \left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{X}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_iY_i \\ &= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y}.\end{aligned}$$

We have that

$$n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n^{2SLS}-\boldsymbol{\beta}\right)\to_d \mathrm{N}\left(0,\sigma^2\left(\mathbb{E}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\mathbb{E}\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\mathbb{E}\boldsymbol{Z}_i\boldsymbol{X}_i'\right)^{-1}\right).$$

The 2SLS estimator is not efficient when the conditional homoskedasticity assumption (9) fails. In this case, the efficient GMM estimator is

$$\widehat{\boldsymbol{\beta}}_n^{GMM} = \left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\sum_{i=1}^{n}\widehat{e}_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\boldsymbol{X}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\sum_{i=1}^{n}\widehat{e}_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_iY_i$$

Exactly identified case

When the number of instruments is equal to the number of regressors ($l=k$), and the $k\times k$ matrix $\boldsymbol{Z}'\boldsymbol{X}$ is of full rank, the 2SLS estimator reduces to the IV estimator

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_n^{2SLS} &= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\ &= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)\left(\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\ &= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\ &= \widehat{\boldsymbol{\beta}}_n^{IV}.\end{aligned}$$

The IV estimator is an example (linear) of the exactly identified case. In this case, the weight matrix $\boldsymbol{A}_n$ plays no role. If the model is exactly identified, the we have k equations in k unknowns. Therefore, it is possible to solve $n^{-1}\sum_{i=1}^{n}\boldsymbol{g}\left(\boldsymbol{W}_i,\boldsymbol{b}\right)=\boldsymbol{0}$ exactly. As a result, the solution to the GMM minimization problem

$$\min_{\boldsymbol{b}}\left\|\boldsymbol{A}_n n^{-1}\sum_{i=1}^{n}\boldsymbol{g}\left(\boldsymbol{W}_i,\boldsymbol{b}\right)\right\|^2$$

does not depend on $\boldsymbol{A}_n$.

Since, in the exactly identified case, $\boldsymbol{Q}$ is $k\times k$ and invertible, the asymptotic variance-covariance matrix takes the following form

$$\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\right)^{-1}$$

$$
\begin{aligned}
&= \boldsymbol{Q}^{-1}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\left(\boldsymbol{Q}'\right)^{-1}\boldsymbol{Q}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}\boldsymbol{Q}^{-1}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\left(\boldsymbol{Q}'\right)^{-1} \\
&= \boldsymbol{Q}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Q}^{-1}\right)' \\
&= \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}
\end{aligned}
$$

independent of $\boldsymbol{A}$ and, naturally, efficient.

Confidence intervals and hypothesis testing in the GMM framework

In this section, we discuss constructing of confidence intervals and hypothesis testing. Let $\widehat{\boldsymbol{\beta}}_n^{GMM}$ be the efficient GMM estimator with the asymptotic variance-covariance matrix $\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\boldsymbol{Q}'\boldsymbol{\Omega}^{-1}\boldsymbol{Q}\right)^{-1}$. Let $\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}$ denote a consistent estimator of $\boldsymbol{V}_{\boldsymbol{\beta}}$.

Since $\widehat{\boldsymbol{\beta}}_n^{GMM}$ is approximately normal in large samples, a confidence interval with the coverage probability $1-\alpha$ for element j of $\boldsymbol{\beta}$ is given by

$$
\left[\widehat{\beta}_{n,j}^{GMM} - z_{1-\alpha/2}\sqrt{\left[\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\right]_{jj}/n}, \widehat{\beta}_{n,j}^{GMM} + z_{1-\alpha/2}\sqrt{\left[\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\right]_{jj}/n}\right],
$$

for $j = 1, \ldots, k$.

For example, in the linear and homoskedastic case, the asymptotic variance of $\widehat{\boldsymbol{\beta}}_n^{2SLS}$ is

$$
\boldsymbol{V}_{\boldsymbol{\beta}} = \sigma^2\left(\mathbb{E}\boldsymbol{X}_i\boldsymbol{Z}_i'\left(\mathbb{E}\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\mathbb{E}\boldsymbol{Z}_i\boldsymbol{X}_i'\right)^{-1},
$$

and its consistent estimator is

$$
\begin{aligned}
\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} &= \widehat{\sigma}_n^2\left(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{Z}_i'\left(n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1} n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i\boldsymbol{X}_i'\right)^{-1} \\
&= n\widehat{\sigma}_n^2\left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1},
\end{aligned}
$$

where $\widehat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n\left(Y_i - \boldsymbol{X}_i'\widehat{\boldsymbol{\beta}}_n^{2SLS}\right)^2$. Therefore, the $1-\alpha$ asymptotic confidence interval for β_j is given by

$$
\widehat{\beta}_{n,j}^{2SLS} \pm z_{1-\alpha/2}\sqrt{\widehat{\sigma}_n^2\left[\left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\right]_{jj}}.
$$

One can construct a test of the null hypothesis $H_0 : \beta_j = \beta_{0,j}$ against $H_1 : \beta_j \neq \beta_{0,j}$ by using the following test statistic:

$$
T_{n,j} = \frac{\widehat{\beta}_{n,j}^{GMM} - \beta_{0,j}}{\sqrt{\left[\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}\right]_{jj}/n}}.
$$

Since under the null hypothesis $T_{n,j} \to_d \mathrm{N}(0,1)$, the asymptotic α-size test is given by

$$
\text{Reject } H_0 \text{ if } |T_{n,j}| > z_{1-\alpha/2}.
$$

One can use a Wald statistic in order to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$:

$$
W_n = n\left(\widehat{\boldsymbol{\beta}}_n^{GMM} - \boldsymbol{\beta}_0\right)'\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{-1}\left(\widehat{\boldsymbol{\beta}}_n^{GMM} - \boldsymbol{\beta}_0\right).
$$

More generally, suppose that the null and alternative are given by $H_0 : \boldsymbol{h}(\boldsymbol{\beta}) = \mathbf{0}$ and $H_1 : \boldsymbol{h}(\boldsymbol{\beta}) \neq \mathbf{0}$ where $\boldsymbol{h} : \mathbb{R}^k \to \mathbb{R}^q$. By the delta method, under H_0,

$$n^{1/2}\boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}\right) \to_d \mathrm{N}\left(0, \left.\frac{\partial \boldsymbol{h}(\boldsymbol{b})}{\partial \boldsymbol{b}'}\right|_{\boldsymbol{b}=\boldsymbol{\beta}} \boldsymbol{V}_{\boldsymbol{\beta}} \left.\frac{\partial \boldsymbol{h}(\boldsymbol{b})'}{\partial \boldsymbol{b}}\right|_{\boldsymbol{b}=\boldsymbol{\beta}}\right).$$

Therefore, the Wald statistic is given by

$$W_n = n \cdot \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}\right)' \left(\frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}\right)}{\partial \boldsymbol{b}'} \widehat{\boldsymbol{V}}_{\boldsymbol{\beta}} \frac{\partial \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}\right)'}{\partial \boldsymbol{b}}\right)^{-1} \boldsymbol{h}\left(\widehat{\boldsymbol{\beta}}_n^{GMM}\right).$$

The asymptotic α-size test is given by

$$\text{Reject } H_0 \text{ if } W_n > \chi^2_q.$$