Insights for the GMM estimator and the test of overidentifying restrictions

Econometrics (35B206), Lecture 7

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2019

GMM estimator, insights

Consider the SLM with endogenous variables. Suppose that $\boldsymbol{z}_i$ is the vector of instruments. It is $L \times 1$. $\boldsymbol{\beta}$ is vector of unknown coefficients. It is $K \times 1$. The orthogonality assumption of the instrumental variables theory requires that the following population moment conditions hold:

$$\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] = \mathrm{E}\left[\boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}\right)\right] = \mathbf{0}.$$

This is a system of L equations with K unknowns.

GMM estimator, insights

If $L > K$, the system

$$\mathrm{E}\left[\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right] \neq \mathbf{0}$$

usually does not have a unique solution. What can we do?

GMM estimator, insights

Out of the L moment conditions, we could select K of them in an arbitrary way, and discard the remaining $L - K$ moment conditions. The K moments conditions allow to produce a MM estimator.

GMM estimator, insights

Another thing we can do is to linearly combine the L moment conditions to produce K linearly independent moment conditions. That is, use

$$\underbrace{\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}}_{K \times L} \underbrace{\mathrm{E}\left[\boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \beta\right)\right]}_{L \times 1} = \mathbf{0}$$

where

$$\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}$$

is a particular matrix that linearly combines L moment conditions to produce K linearly independent moment conditions. The system becomes exactly identified. Note that these moment conditions are the F.O.C. of the GMM minimisation problem. They are the population versions, however.

GMM estimator, insights

Take a closer look at what these linear combinations are about. This will provide insights into our understating of the GMM principle, and pave the way for a hypothesis test.

GMM estimator, insights

Start by assuming that the decomposition

$$\boldsymbol{W} = \boldsymbol{W}^{1/2'}\boldsymbol{W}^{1/2}.$$

holds, where $\boldsymbol{W}^{1/2}$ is the principal square root of the matrix $\boldsymbol{W}$: remember that $\boldsymbol{W}$ is a symmetric positive definite matrix by assumption. $\boldsymbol{W}^{1/2}$ is $L \times L$ and invertible. Then, we can recast the population version of the F.O.C. above as

$$\underbrace{\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}^{1/2'}}_{K\times L}\underbrace{\boldsymbol{W}^{1/2}\mathrm{E}\left[\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\beta\right)\right]}_{L\times 1} = \mathbf{0}.$$

GMM estimator, insights

$$\underbrace{\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}^{1/2'}}_{\equiv\boldsymbol{A}'}\underbrace{\boldsymbol{W}^{1/2}\mathrm{E}\left[\boldsymbol{z}_i\left(y_i-\boldsymbol{x}_i'\beta\right)\right]}_{\equiv\boldsymbol{B}}=\mathbf{0}.$$

GMM estimator, insights

$$\boldsymbol{A}' = \underbrace{\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right]}_{K \times L} \underbrace{\boldsymbol{W}^{1/2'}}_{L \times L}.$$

$\boldsymbol{A}'$ is $K \times L$. Recall that rank of $\boldsymbol{X}$ is K. Rank of $\boldsymbol{A}'$ is K.

GMM estimator, insights

$$\boldsymbol{B} = \underbrace{\boldsymbol{W}^{1/2}}_{L\times L}\underbrace{\mathrm{E}\left[\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\beta\right)\right]}_{L\times 1}.$$

$\boldsymbol{B}$ is $L \times 1$. The original L moment conditions

$$\mathrm{E}\left[\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\beta\right)\right] = \mathbf{0}$$

imply that

$$\boldsymbol{B} = \boldsymbol{W}^{1/2}\mathrm{E}\left[\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\beta\right)\right] = \mathbf{0}.$$

$\boldsymbol{B}$ is a recast of the original L moment conditions. The structure of the problem remains intact. $\boldsymbol{B}$ is a system of L equations with K unknowns.

GMM estimator, insights

The moment conditions $\boldsymbol{B}$ can be decomposed orthogonally as

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{B} + \boldsymbol{M}\boldsymbol{B}$$

since

$$\boldsymbol{I} = \boldsymbol{P} + \boldsymbol{M}.$$

$\boldsymbol{P}$ and $\boldsymbol{M}$ are any pair of orthogonal projection matrices that project $\boldsymbol{B}$ to two subspaces that are orthogonal to each other. The decomposition holds for any pair of projection matrices.

GMM estimator, insights

Choose $\boldsymbol{P}$ as

$$\boldsymbol{P} = \underbrace{\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}}_{L\times K} \underbrace{\boldsymbol{A}'}_{K\times L}$$

where

$$\boldsymbol{A}' = \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}^{1/2'}$$

as defined earlier. Rank of $\boldsymbol{A}'$ is K. Rank of $\boldsymbol{P}$ is K. When multiplied with them, $\boldsymbol{P}$ projects vectors of dimension L onto the span of the K column vectors in $\boldsymbol{A}$.

GMM estimator, insights

We know that

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}.$$

The trace of $\boldsymbol{I}$ is L. The rank of an idempotent matrix is equal to the trace of it. Rank of $\boldsymbol{P}$ is K. Hence, trace of $\boldsymbol{P}$ is K. Hence, trace of $\boldsymbol{M}$ is $L - K$. $\boldsymbol{M}$ is idempotent. Rank of $\boldsymbol{M}$ is $L - K$.

GMM estimator, insights

$\boldsymbol{PB}$ becomes

$$\boldsymbol{PB} = \underbrace{\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'}_{L\times L}\underbrace{\boldsymbol{B}}_{L\times 1}.$$

GMM estimator, insights

What does $\boldsymbol{PB}$ represent? Recalling that

$$\boldsymbol{A}'\boldsymbol{B} = \underbrace{\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}^{1/2'}}_{K \times L} \underbrace{\boldsymbol{W}^{1/2} \mathrm{E}\left[\boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \beta\right)\right]}_{L \times 1},$$

we have

$$\boldsymbol{PB} = \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1} \underbrace{\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}^{1/2'} \boldsymbol{W}^{1/2} \mathrm{E}\left[\boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \beta\right)\right]}_{\text{F.O.C. of the GMM minimisation problem}}.$$

GMM estimator, insights

Recall that $\boldsymbol{B} = \mathbf{0}$. Hence,

$$\boldsymbol{P}\boldsymbol{B} = \underbrace{\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'}_{L\times L}\underbrace{\boldsymbol{B}}_{L\times 1} = \mathbf{0}.$$

$\boldsymbol{P}\boldsymbol{B}$ is $L \times 1$. Rank of $\boldsymbol{P}$ is K. Rank of $\boldsymbol{P}\boldsymbol{B}$ is K. $\boldsymbol{P}\boldsymbol{B}$ is a system of K linearly independent equations with K unknowns.

GMM estimator, insights

Or recall that

$$\boldsymbol{A}'\boldsymbol{B} = \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}^{1/2'} \boldsymbol{W}^{1/2} \mathrm{E}\left[\boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i'\beta\right)\right] = \mathbf{0}$$

by the population version of the F.O.C. of the GMM problem. This is a system of K equations with K unknowns. Then,

$$\boldsymbol{P}\boldsymbol{B} = \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1} \boldsymbol{A}'\boldsymbol{B} = \mathbf{0}$$

since rank of $\boldsymbol{A}'$ is K and rank of $\boldsymbol{P}$ is K.

GMM estimator, insights

$\boldsymbol{MB}$ becomes

$$\boldsymbol{MB} = \underbrace{\left(\boldsymbol{I} - \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\right)}_{L\times L}\underbrace{\boldsymbol{B}}_{L\times 1}.$$

GMM estimator, insights

Recall that $\boldsymbol{B} = \mathbf{0}$. Hence,

$$\boldsymbol{MB} = \underbrace{\left(\boldsymbol{I} - \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\right)}_{L \times L}\underbrace{\boldsymbol{B}}_{L \times 1} = \mathbf{0}.$$

$\boldsymbol{MB}$ is $L \times 1$. Rank of $\boldsymbol{M}$ is $L - K$. Rank of $\boldsymbol{MB}$ is $L - K$. $\boldsymbol{MB}$ is a system of $L - K$ linearly independent equations with K unknowns.

GMM estimator, insights

These results provide insights into the GMM estimation. When $L > K$, what GMM is doing is that it is decomposing the moment conditions $\boldsymbol{B} = \mathbf{0}$ in two orthogonal parts. $\boldsymbol{PB} = \mathbf{0}$ is one part that is used to exactly identify the relevant parameters. These moment conditions are called the identifying conditions (or restrictions). $\boldsymbol{MB} = \mathbf{0}$ is the other part that is left unused. These moment conditions are called the overidentifying conditions. In this respect, GMM is using K moment conditions where each is a linear combination of the original L moment conditions, and discarding the remaining $L - K$ moment conditions where each is a linear combination of the original L moment conditions. It is not discarding any of the original moment conditions!

The test of overidentifying restrictions

We now use the results above to construct a hypothesis test. The test itself will provide further insights into the GMM principle.

The test of overidentifying restrictions, idea

Consider the GMM objective function, evaluated at $\hat{\beta}$,

$$q\left(\hat{\beta}\right) = \left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right)\right)' \boldsymbol{W}_n^{1/2\prime}\boldsymbol{W}_n^{1/2}\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right)\right).$$

We can write it as

$$q\left(\hat{\beta}\right) = \boldsymbol{B}_n'\boldsymbol{B}_n$$

where

$$\boldsymbol{B}_n = \boldsymbol{W}_n^{1/2}\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right)\right).$$

$\boldsymbol{B}_n$ is the sample counterpart of $\boldsymbol{B}$. Multiply by n to obtain

$$nq\left(\hat{\beta}\right) = n\boldsymbol{B}_n'\boldsymbol{B}_n.$$

The test of overidentifying restrictions, idea

Consider the orthogonal projection introduced earlier,

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{B} + \boldsymbol{M}\boldsymbol{B} = (\boldsymbol{P} + \boldsymbol{M})\,\boldsymbol{B}.$$

The sample counterpart is

$$\boldsymbol{B}_n = \boldsymbol{P}_n\boldsymbol{B}_n + \boldsymbol{M}_n\boldsymbol{B}_n = (\boldsymbol{P}_n + \boldsymbol{M}_n)\,\boldsymbol{B}_n$$

where

$$\boldsymbol{P}_n = \boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'$$

and

$$\boldsymbol{A}_n' = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n^{1/2\prime}.$$

The test of overidentifying restrictions, idea

Using the properties of a projection matrix,

$$\begin{aligned} nq\left(\hat{\beta}\right) &= n\boldsymbol{B}_n'\boldsymbol{B}_n \\ &= n\left[\boldsymbol{B}_n'\left(\boldsymbol{P}_n+\boldsymbol{M}_n\right)\left(\boldsymbol{P}_n+\boldsymbol{M}_n\right)\boldsymbol{B}_n\right] \\ &= n\left[\boldsymbol{B}_n'\left(\boldsymbol{P}_n+\boldsymbol{M}_n\right)\boldsymbol{B}_n\right] \\ &= n\left[\boldsymbol{B}_n'\boldsymbol{P}_n\boldsymbol{B}_n+\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n\right] \\ &= n\boldsymbol{B}_n'\boldsymbol{P}_n\boldsymbol{B}_n+n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n. \end{aligned}$$

The test of overidentifying restrictions, idea

$$nq\left(\hat{\beta}\right) = n\boldsymbol{B}_n'\boldsymbol{P}_n\boldsymbol{B}_n + n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n.$$

$\hat{\boldsymbol{\beta}}$ satisfies strictly the identifying restrictions $\boldsymbol{P}_n\boldsymbol{B}_n$ by the F.O.C. of the GMM problem. That is,

$$\boldsymbol{P}_n\boldsymbol{B}_n = \boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{B}_n = \boldsymbol{0}.$$

since

$$\boldsymbol{A}_n'\boldsymbol{B}_n = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n^{1/2'}\boldsymbol{W}_n^{1/2}\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right)\right) = \boldsymbol{0}$$

by the F.O.C. $\hat{\boldsymbol{\beta}}$ is trying to make the overidentifying restrictions $\boldsymbol{M}_n\boldsymbol{B}_n$ as small as possible. That is, $\hat{\boldsymbol{\beta}}$ is trying to minimise

$$nq\left(\hat{\beta}\right) = n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n.$$

The test of overidentifying restrictions, idea

$\hat{\boldsymbol{\beta}}$ tries to minimise

$$nq\left(\hat{\boldsymbol{\beta}}\right) = n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n.$$

$n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n$ measures how far is the sample from satisfying the overidentifying restrictions. This fact can be exploited to design a formal specification test. We want to test for whether $n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n$ is too large. If it is, some of the conditions that guarantee the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{GMM}$ are likely to be false.

The test of overidentifying restrictions, derivation

We derive the test as follows. Since

$$\boldsymbol{P}_n \boldsymbol{B}_n = \mathbf{0}$$

by the F.O.C. of the GMM problem, it follows that

$$\begin{aligned} nq\left(\hat{\beta}_{GMM}\right) &= n\boldsymbol{B}_n' \boldsymbol{M}_n \boldsymbol{B}_n \\ &= n\left(\boldsymbol{W}_n{}^{1/2}\frac{1}{n}\boldsymbol{Z}'\hat{\varepsilon}\right)' \boldsymbol{M}_n \left(\boldsymbol{W}_n{}^{1/2}\frac{1}{n}\boldsymbol{Z}'\hat{\varepsilon}\right)\frac{n}{\sqrt{n}\sqrt{n}} \\ &= \left(\boldsymbol{W}_n{}^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon}\right)' \boldsymbol{M}_n \left(\boldsymbol{W}_n{}^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon}\right). \end{aligned}$$

We now consider the term

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon}.$$

The test of overidentifying restrictions, derivation

$$\begin{aligned}
\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon} &= \frac{1}{\sqrt{n}}\boldsymbol{Z}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{GMM}\right) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{Z}'\left(\varepsilon + \boldsymbol{X}\beta - \boldsymbol{X}\hat{\beta}_{GMM}\right) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{Z}'\left(\varepsilon - \boldsymbol{X}\left(\hat{\beta}_{GMM} - \beta\right)\right) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon - \frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{X}\left(\hat{\beta}_{GMM} - \beta\right)\frac{\sqrt{n}}{\sqrt{n}} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon - \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\left(\hat{\beta}_{GMM} - \beta\right)\sqrt{n}.
\end{aligned}$$

The test of overidentifying restrictions, derivation

Using

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}\right) = \boldsymbol{H}_n \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon,$$

from our derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}_{GMM}$,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \boldsymbol{Z}'\hat{\varepsilon} &= \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon - \frac{1}{n} \boldsymbol{Z}'\boldsymbol{X} \left(\hat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}\right) \sqrt{n} \\
&= \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon - \frac{1}{n} \boldsymbol{Z}'\boldsymbol{X} \boldsymbol{H}_n \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon \\
&= \left(\boldsymbol{I} - \frac{1}{n} \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}_n\right) \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon.
\end{aligned}$$

We now consider the term

$$\boldsymbol{I} - \frac{1}{n} \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}_n.$$

The test of overidentifying restrictions, derivation

Consider our earlier definition that

$$\boldsymbol{H}_n = \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n.$$

Then,

$$\begin{aligned}
&\boldsymbol{I} - \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}_n \\
&= \boldsymbol{I} - \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n \\
&= \boldsymbol{I} - \boldsymbol{W}_n^{-1/2}\boldsymbol{W}_n^{1/2}\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n^{1/2}\boldsymbol{W}_n^{1/2}.
\end{aligned}$$

The test of overidentifying restrictions, derivation

Recall that

$$\boldsymbol{A}' = \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W}^{1/2\prime}.$$

Its sample counterpart is

$$\boldsymbol{A}_n' = \frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n^{1/2\prime}.$$

Then,

$$\begin{aligned}
&\boldsymbol{I} - \boldsymbol{W}_n^{-1/2} \boldsymbol{W}_n^{1/2} \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n^{1/2}\boldsymbol{W}_n^{1/2} \\
&= \boldsymbol{I} - \boldsymbol{W}_n^{-1/2}\boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{W}_n^{1/2} \\
&= \boldsymbol{W}_n^{-1/2}\boldsymbol{I}\boldsymbol{W}_n^{1/2} - \boldsymbol{W}_n^{-1/2}\boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{W}_n^{1/2} \\
&= \boldsymbol{W}_n^{-1/2}\left(\boldsymbol{I}\boldsymbol{W}_n^{1/2} - \boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{W}_n^{1/2}\right) \\
&= \boldsymbol{W}_n^{-1/2}\left(\boldsymbol{I} - \boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n\right)^{-1}\boldsymbol{A}_n'\right)\boldsymbol{W}_n^{1/2} \\
&= \boldsymbol{W}_n^{-1/2}\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}.
\end{aligned}$$

The test of overidentifying restrictions, derivation

We have shown that

$$\boldsymbol{I} - \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}_n = \boldsymbol{W}_n^{-1/2}\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}.$$

Earlier in the derivation we had

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon} = \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}_n\right)\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon$$

which becomes, with the above stated equality,

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon} = \boldsymbol{W}_n^{-1/2}\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon$$

The test of overidentifying restrictions, derivation

$$
\begin{aligned}
nq\left(\hat{\beta}_{GMM}\right) &= n\boldsymbol{B}_n'\boldsymbol{M}_n\boldsymbol{B}_n \\
&= \left(\boldsymbol{W}_n{}^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon}\right)'\boldsymbol{M}_n\left(\boldsymbol{W}_n{}^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\hat{\varepsilon}\right) \\
&= \left(\boldsymbol{W}_n{}^{1/2}\boldsymbol{W}_n^{-1/2}\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n \\
&\quad \left(\boldsymbol{W}_n{}^{1/2}\boldsymbol{W}_n^{-1/2}\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right) \\
&= \left(\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n\left(\boldsymbol{M}_n\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right) \\
&= \left(\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n'\boldsymbol{M}_n\boldsymbol{M}_n\left(\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right) \\
&= \left(\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n\left(\boldsymbol{W}_n^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)
\end{aligned}
$$

since $\boldsymbol{M}_n'\boldsymbol{M}_n = \boldsymbol{M}_n$ and $\boldsymbol{M}_n' = \boldsymbol{M}_n$.

The test of overidentifying restrictions, derivation

Recall from our derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}_{GMM}$ that

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon \xrightarrow{d} N\big(0, \boldsymbol{S}_n\big).$$

Using the optimal weighting matrix, so that $\boldsymbol{W}_n = \boldsymbol{S}_n^{-1}$, and that $\boldsymbol{W}_n^{1/2} = \boldsymbol{S}_n^{-1/2}$,

$$\boldsymbol{W}_n{}^{1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon = \boldsymbol{S}_n^{-1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon \xrightarrow{d} N\big(0, \boldsymbol{I}_L\big).$$

The test of overidentifying restrictions, derivation

Using the optimal weighting matrix so that $\boldsymbol{W}_n^{1/2} = \boldsymbol{S}_n^{-1/2}$,

$$nq\left(\hat{\boldsymbol{\beta}}_{GMM}^{O}\right) = \left(\boldsymbol{S}_n^{-1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n\left(\boldsymbol{S}_n^{-1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right),$$

where $\hat{\boldsymbol{\beta}}_{GMM}^{O}$ denotes the optimal $\hat{\boldsymbol{\beta}}_{GMM}$. The first and the third terms converge in distribution to a standard normal distribution. Recalling that rank of $\boldsymbol{M}_n$ is $L-K$, and using Theorem B.8 in Greene,

$$nq\left(\hat{\boldsymbol{\beta}}_{GMM}^{O}\right) \xrightarrow{d} \chi^2\left[L-K\right].$$

The test of overidentifying restrictions, derivation

$$nq\left(\hat{\beta}^{O}_{GMM}\right) = \left(\boldsymbol{S}_n^{-1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right)'\boldsymbol{M}_n\left(\boldsymbol{S}_n^{-1/2}\frac{1}{\sqrt{n}}\boldsymbol{Z}'\varepsilon\right).$$

ε is not observed. However, using the earlier result $\boldsymbol{P}_n\boldsymbol{B}_n = \boldsymbol{0}$, the definition of $\boldsymbol{B}_n$, and the fact that $\boldsymbol{M}_n = \boldsymbol{I} - \boldsymbol{P}_n$, it can be shown that

$$J \equiv nq\left(\hat{\beta}^{O}_{GMM}\right) = n\left(\frac{1}{n}\boldsymbol{Z}'\hat{\varepsilon}\right)'\boldsymbol{S}_n{}^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\hat{\varepsilon}\right).$$

The test of overidentifying restrictions, derivation

The J statistic is used to test the overidentifying restrictions. The null and the alternative hypotheses are

$$H_0 : \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] = 0$$
$$H_1 : \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] \neq 0.$$

The test statistic is given by

$$J = n \left(\frac{1}{n}\boldsymbol{Z}'\hat{\boldsymbol{\varepsilon}}\right)' \boldsymbol{S}_n{}^{-1} \left(\frac{1}{n}\boldsymbol{Z}'\hat{\boldsymbol{\varepsilon}}\right)$$

and, under the null,

$$J \xrightarrow{d} \chi^2 \left[L - K\right].$$

The test of overidentifying restrictions, intuition

The test statistic is

$$J = n \left(\frac{1}{n} \boldsymbol{Z}' \hat{\boldsymbol{\varepsilon}} \right)' \boldsymbol{S}_n{}^{-1} \left(\frac{1}{n} \boldsymbol{Z}' \hat{\boldsymbol{\varepsilon}} \right).$$

By the LLN,

$$\frac{1}{n} \boldsymbol{Z}' \hat{\boldsymbol{\varepsilon}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \hat{\varepsilon}_i \xrightarrow{p} \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \right].$$

Hence, J will be close to 0, when n is large, if the null

$$\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \right] = 0$$

is true. J will explode, when n is large, if the alternative

$$\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \right] \neq 0$$

is true.

The test of overidentifying restrictions, intutition

Under the null, meaning that the model is correctly specified so that $\hat{\beta}_{GMM}$ is consistent, the overidentifying restrictions should be close to zero. The J test checks if the overidentifying restrictions are small. Therefore, we seek a small J to fail to reject the test.

The test of overidentifying restrictions, example

```
. ivregress gmm lwage age black (educ = motheduc fatheduc)

Instrumental variables (GMM) regression           Number of obs   =      2,220
                                                  Wald chi2(3)    =     523.27
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.1892
GMM weight matrix: Robust                         Root MSE        =     .39583
```

lwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ	.0602296	.0071722	8.40	0.000	.0461723	.0742869
age	.0429854	.0028103	15.30	0.000	.0374772	.0484935
black	-.185577	.0249487	-7.44	0.000	-.2344756	-.1366785
_cons	4.294079	.1200834	35.76	0.000	4.05872	4.529438

```
Instrumented:  educ
Instruments:   age black motheduc fatheduc

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = 1.02668 (p = 0.3109)
```