The MM and the GMM estimators

Econometrics (35B206), Lecture 6

Tunga Kantarcı, TiSEM, Tilburg University, Spring 2019

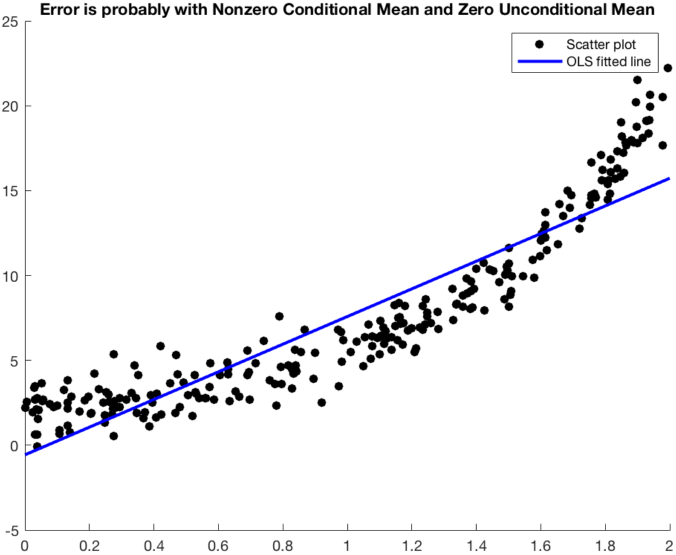
Motivation, violation of the linearity assumption

The SLM assumes that the model is linear in β. The model

$$y_i = 1 + e^{x_i \beta} + \varepsilon_i$$

is non-linear in β. How to estimate β?

Motivation, violation of the linearity assumption



Motivation, violation of the linearity assumption

The fitted line is based on the standard linear model. The vertical difference between an observation and the fitted line is a residual. The overall mean of the residuals is 0. This is true by construction as long as the regression includes a constant. But for specific ranges of x_i the mean is not 0. So, given the sample data, in the population, is

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

likely to hold? Yes. Is

$$\mathrm{E}\left[\varepsilon_i \mid x_i\right] = 0$$

likely to hold? No.

Motivation, violation of the linearity assumption

From the plot, which is based on sample data, we can infer that a linear model is not a good approximation of the true model. We cannot defend the zero conditional mean assumption. In the plot, the sample data for y_i is in fact simulated using the true data generating process

$$y_i = 1 + e^{1.5x_i} + \varepsilon_i$$

where x_i and ε_i take random values from given distributions. This true model is not observed by the researcher. If the true parameter value 1.5 was not observed, how could we estimate it?

Moment

Like the distribution, or the density function, we use the moment to characterise the randomness of a variable.

Moment

The moments of a random variable x are given by

$$\mu^{(i)} := \mathrm{E}\left[x^i\right] = \int_{-\infty}^{\infty} u^i\, f(u)\, \mathrm{d}u$$

if x is continuous, and where f is the probability density function of x, and u is the probable value of x. $\mu^{(i)}$ indicates the mean of x^i. i is the order of the moment. If $i = 1$, $\mu^{(1)}$ is the mean of x. It is typical to write just μ for the mean of x. It is a measure of the location of the center of the distribution of x.

Moment

We interpret the moment as the weighted average of the values that x^i can take, where the weights are given by the probability density function of x: a higher weight is assigned to more probable outcomes.

Moment

The central moments of a random variable x are given by

$$\sigma^{(i)} := \mathrm{E}\left[(x-\mu)^i\right] = \int_{-\infty}^{\infty} (u-\mu)^i \ f(u)\,\mathrm{d}u$$

if x is continuous. If $i = 1$, $\sigma^{(1)} = 0$. If $i = 2$, $\sigma^{(2)}$ is the variance of x. It is typical to write just σ^2 for the variance of x. In this lecture we are not interested in a central moment.

Moment

In this lecture β denotes a generic unknown parameter. β_0 denotes an unknown parameter in the population.

Moment

Let

$$m_i(\beta)$$

denote a moment. It is a general function. It can be nonlinear in β.

Moment

Define

$$\bar{m}_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^{n} m_i(\beta)$$

to be the sample moment.

Moment

Define

$$m(\boldsymbol{\beta}) \equiv \mathrm{E}\left[m_i(\beta)\right]$$

to be the population moment.

Moment

If we impose a restriction on the population moment so that it is equal to some number, like

$$\mathrm{E}\left[m_i\left(\beta\right)\right] = 0,$$

we obtain a population moment condition.

Moment, example

$$\mathrm{E}\left[z_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right]$$

is a population moment where

$$m_i\left(\boldsymbol{\beta}\right) = z_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right).$$

Moment, example

If we require

$$\mathrm{E}\left[z_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0\right)\right] = 0,$$

we obtain a population moment condition. This particular population moment condition is nothing but the exogeneity assumption that the instrument z_i should satisfy in the SLM with an endogenous variable:

$$\mathrm{E}\left[z_i \varepsilon_i\right] = 0.$$

Moment

$$m_i(\beta)$$

represents a moment. If we have L moments, we can stack them together to obtain

$$\boldsymbol{m}_i(\boldsymbol{\beta})$$

which is a $L \times 1$ vector of moments.

Moment, example

$$\boldsymbol{m}_i(\boldsymbol{\beta}) = \boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)$$

If $\boldsymbol{z}_i$ is $L \times 1$, $\boldsymbol{m}_i$ is also $L \times 1$.

The case of $L \geq K$

This lecture is about finding the parameter that solves a system of equations. We consider two cases. In the first case the number of equations is larger than the number of parameters. This case will lead to the GMM estimator. In the second case they are equal. This will lead to the MM estimator. We start with the former, the more difficult case. The case of $L \geq K$.

GMM estimator, assumptions

We will derive the GMM estimator. But first, we make a number of assumptions.

GMM estimator, assumptions

The observations of y_i, $\boldsymbol{x}_i$, and if there is any, the vector of instrumental variables $\boldsymbol{z}_i$, are all random samples from the population.

GMM estimator, assumptions

$$\boldsymbol{\beta}$$

is the vector of unknown coefficients. It is $K \times 1$. It does not satisfy a system of population equations.

$$\boldsymbol{\beta}_0$$

is the vector of unknown coefficients in the population. It satisfies a system of population equations.

GMM estimator, assumptions

$$\boldsymbol{m}\left(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}\right)$$

is a vector of moment functions. It is $L \times 1$. We use the shorthand

$$\boldsymbol{m}_i\left(\boldsymbol{\beta}\right).$$

GMM estimator, assumptions

$$\boldsymbol{m}_i(\boldsymbol{\beta})$$

is $L \times 1$.

$$\boldsymbol{\beta}$$

is $K \times 1$. We consider the case that the number of moments is at least as much as the number of parameters so that $L \geq K$.

GMM estimator, assumptions

$$\boldsymbol{m}_i(\boldsymbol{\beta})$$

is a continuously differentiable function in $\boldsymbol{\beta}$.

$$\frac{\partial \boldsymbol{m}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$$

is the matrix of the derivatives of the moments. It is $L \times K$. It has rank K at each variable and at each parameter.

GMM estimator, assumptions

There exists a sequence of possibly random $L \times L$ symmetric positive definite matrices $\boldsymbol{W}_n$, such that

$$\boldsymbol{W}_n \xrightarrow{p} \boldsymbol{W}$$

where $\boldsymbol{W}$ is also symmetric positive definite. We require $\boldsymbol{W}$ to be positive definite so that we can put a positive and non-zero weight on all moment conditions. Note that we put the weight on the moment conditions, not on the observations of variables as in GLS.

GMM estimator, derivation

$$\mathrm{E}\left[\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\right]$$

is the vector of population moments. It has rank K at each parameter.

GMM estimator, derivation

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}_i\left(\boldsymbol{\beta}\right)$$

is the vector of sample moments based on sample data of size n.

GMM estimator, derivation

The sample counterpart of

$$\mathrm{E}\left[\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\right]$$

is

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)$$

because

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\xrightarrow{p}\mathrm{E}\left[\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\right]$$

by the LLN. This underlies the logic of the GMM estimator. Hence, we will play in the field of a large sample.

GMM estimator, derivation

$$\mathsf{E}\left[\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\right] \neq \mathbf{0}$$

is a vector of population moment conditions. Since $L \geq K$, it represents a system with more equations than unknowns. The inequality because the system usually does not have a unique solution with some $\boldsymbol{\beta}$. However, some econometric theory will require that it does with $\boldsymbol{\beta}_0$. Recall the exogeneity assumption on the instrument, for example.

GMM estimator, derivation

$$\bar{\boldsymbol{m}}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}_i(\boldsymbol{\beta}) \neq \mathbf{0}$$

is the corresponding sample mean if the LLN applies. From here to the previous slide: if there is no solution in the sample mean, it is likely that there is no solution in the population, when LLN applies. The system is over-determined, and there is usually no solution. But we want a solution because, e.g., if we have more instruments than we need, we do not want to discard but use them. The sample mean is what is relevant in practice. Hence, we really have a problem. What to do?

GMM estimator, derivation

If we cannot make $\bar{\boldsymbol{m}}_n(\beta)$ equal to $\mathbf{0}$, we can try to get $\bar{\boldsymbol{m}}_n(\beta)$ close to $\mathbf{0}$. We can do this by minimising the weighted quadratic distance

$$(\bar{\boldsymbol{m}}_n(\beta) - \mathbf{0})' \boldsymbol{W}_n (\bar{\boldsymbol{m}}_n(\beta) - \mathbf{0}).$$

The $\boldsymbol{\beta}$ that minimises this distance is $\hat{\boldsymbol{\beta}}_{GMM}$. The quadratic distance is in fact a norm: $\| \bar{\boldsymbol{m}}_n(\beta) \|^2_{\boldsymbol{W}_n}$. This norm is very general. This means that the estimator we derive here is very general. Different choices of $\boldsymbol{W}_n$ leads to different norms and hence different estimators. The role of the weighting matrix $\boldsymbol{W}_n$, and the reason of considering a quadratic form will be explained later.

GMM estimator, derivation

$$\hat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \bar{\boldsymbol{m}}_n(\beta)' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\beta)\,.$$

GMM estimator, statistical properties

We now study the statistical properties of $\hat{\beta}_{GMM}$. In particular, we study the large sample properties. But we will not require that $\hat{\beta}_{GMM}$ is unbiased. Why would not we want to study the small sample properties? Realise that we motivated $\hat{\beta}_{GMM}$ using the LLN. Hence, by construction $\hat{\beta}_{GMM}$ requires a large sample. Small sample properties of $\hat{\beta}_{GMM}$ depends on the model we are estimating, and hence requires a case-by-case analysis.

GMM estimator, statistical properties

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

Instrumental variables (GMM) regression          Number of obs   =      2,220
                                                 Wald chi2(3)    =     515.30
                                                 Prob > chi2     =     0.0000
                                                 R-squared       =     0.1824
GMM weight matrix: Robust                        Root MSE        =     .39748
```

lwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ	.0645545	.008379	7.70	0.000	.048132	.080977
age	.0428922	.0028215	15.20	0.000	.0373622	.0484222
black	-.1774985	.0262029	-6.77	0.000	-.2288554	-.1261417
_cons	4.236309	.1332249	31.80	0.000	3.975193	4.497425

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

GMM estimator, consistency

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0.$$

We do not prove this in the current exposition of the general theory. We will prove it in an example case later.

GMM estimator, asymptotic normality

We sketch the proof of the asymptotic normality of $\hat{\beta}_{GMM}$.

GMM estimator, asymptotic normality

Define

$$\bar{\boldsymbol{G}}_n(\beta) \equiv \frac{\partial \bar{\boldsymbol{m}}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \frac{\partial \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \boldsymbol{m}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$$

using the sum rule in differentiation.

GMM estimator, asymptotic normality

The objective function of the minimisation problem is

$$q_n(\boldsymbol{\beta}) = \bar{\boldsymbol{m}}_n(\boldsymbol{\beta})' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\boldsymbol{\beta}).$$

The first-order conditions from the minimisation problem are

$$\frac{\partial q_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right)}{\partial \hat{\boldsymbol{\beta}}_{GMM}} = 2\bar{\boldsymbol{G}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right)' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right) = \mathbf{0}.$$

We will use this result below and elsewhere. The leading 2 is irrelevant to the solution, so we drop it here.

GMM estimator, asymptotic normality

Employing the first-order Taylor expansion of $\bar{\boldsymbol{m}}_n\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}}\right)$ around $\boldsymbol{\beta}_0$ gives the approximation

$$\bar{\boldsymbol{m}}_n\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}}\right) \approx \bar{\boldsymbol{m}}_n\left(\boldsymbol{\beta}_0\right) + \underbrace{\frac{\partial \bar{\boldsymbol{m}}_n\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}_0'}}_{\bar{\boldsymbol{G}}_n(\boldsymbol{\beta}_0)}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}} - \boldsymbol{\beta}_{\mathbf{0}}\right)$$

GMM estimator, asymptotic normality

Considering the Taylor expansion in the first-order condition,

$$\bar{\boldsymbol{G}}_n\left(\hat{\beta}_{GMM}\right)' \boldsymbol{W}_n \bar{\boldsymbol{G}}_n(\beta_0)\left(\hat{\beta}_{\boldsymbol{GMM}} - \beta_{\boldsymbol{0}}\right) \approx -\bar{\boldsymbol{G}}_n\left(\hat{\beta}_{GMM}\right)' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\beta_0),$$

and solving for the estimation error, and multiplying by $\sqrt{n}$

$$\sqrt{n}\left(\hat{\beta}_{\boldsymbol{GMM}} - \beta_{\boldsymbol{0}}\right) \approx - \left[\bar{\boldsymbol{G}}_n\left(\hat{\beta}_{GMM}\right)' \boldsymbol{W}_n \bar{\boldsymbol{G}}_n(\beta_0)\right]^{-1} \bar{\boldsymbol{G}}_n\left(\hat{\beta}_{GMM}\right)' \boldsymbol{W}_n \sqrt{n}\bar{\boldsymbol{m}}_n(\beta_0).$$

GMM estimator, asymptotic normality

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}}-\boldsymbol{\beta}_{\mathbf{0}}\right)\approx$$
$$-\left[\bar{\boldsymbol{G}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right)'\boldsymbol{W}_n\bar{\boldsymbol{G}}_n\left(\boldsymbol{\beta}_0\right)\right]^{-1}\bar{\boldsymbol{G}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right)'\boldsymbol{W}_n\sqrt{n}\bar{\boldsymbol{m}}_n\left(\boldsymbol{\beta}_0\right).$$

We need to find the limiting distribution of the LHS. Hence, we look at the limiting distribution of the RHS.

GMM estimator, asymptotic normality

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0$$

and, for the sample moment, this implies that

$$\bar{\boldsymbol{G}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right) \xrightarrow{p} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)$$

where

$$\bar{\boldsymbol{G}}_n\left(\hat{\boldsymbol{\beta}}_{GMM}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{m}_i\left(\hat{\boldsymbol{\beta}}_{GMM}\right)}{\partial \hat{\boldsymbol{\beta}}'_{GMM}}$$

and

$$\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right) = \mathrm{E}\left[\frac{\partial \boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}'_0}\right].$$

GMM estimator, asymptotic normality

We assumed that

$$\boldsymbol{W}_n \xrightarrow{p} \boldsymbol{W}.$$

By the CLT,

$$\sqrt{n}\left(\bar{\boldsymbol{m}}_n\left(\boldsymbol{\beta}_0\right) - \mathbf{0}\right) \xrightarrow{d} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\right].$$

GMM estimator, asymptotic normality

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}} - \boldsymbol{\beta}_{\mathbf{0}}\right) \xrightarrow{d}$$
$$-\left[\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right]^{-1} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} N\left[\mathbf{0}, \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\right].$$

Then,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}} \xrightarrow{d}$$
$$N\left[\boldsymbol{\beta}_{\mathbf{0}}, \frac{1}{n} - \left[\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right]^{-1} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\left(-\left[\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right]^{-1} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W}\right)'\right].$$

This implies

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{GMM}} \overset{a}{\sim}$$
$$N\left[\boldsymbol{\beta}_{\mathbf{0}}, \frac{1}{n} - \left[\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right]^{-1} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\left(-\left[\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right]^{-1} \bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)' \boldsymbol{W}\right)'\right].$$

GMM estimator, asymptotic normality

If the weighting matrix is chosen as

$$\boldsymbol{W} = \left(\mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\right)^{-1},$$

the limiting distribution of the estimator simplifies to

$$\hat{\boldsymbol{\beta}}_{GMM}^{O} \xrightarrow{d} N\left[\boldsymbol{\beta_0}, \frac{1}{n}\left(\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)'\left(\mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\right)^{-1}\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)\right)^{-1}\right].$$

Replacing for $\bar{\boldsymbol{G}}\left(\boldsymbol{\beta}_0\right)$, and as convergence in distribution implies,

$$\hat{\boldsymbol{\beta}}_{GMM}^{O} \overset{a}{\sim}$$
$$N\left[\boldsymbol{\beta}_0, \frac{1}{n}\left(\left(\mathrm{E}\left[\frac{\partial \boldsymbol{m}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0'}\right]\right)'\left(\mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)'\right]\right)^{-1}\mathrm{E}\left[\frac{\partial \boldsymbol{m}_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_0'}\right]\right)^{-1}\right].$$

GMM estimator, the optimal one

The weighting matrix

$$\boldsymbol{W} = \left(\mathsf{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta})\boldsymbol{m}_i(\boldsymbol{\beta})'\right]\right)^{-1}$$

is the optimal weighting matrix. $\hat{\boldsymbol{\beta}}_{GMM}$ that uses this weighting matrix is the optimal $\hat{\boldsymbol{\beta}}_{GMM}$ which we could denote by $\hat{\boldsymbol{\beta}}^{O}_{GMM}$ if we wanted to. Optimal in the sense that it has the smallest variance compared to another GMM estimator using a different $\boldsymbol{W}$. More on this later.

GMM estimator, the optimal one

In the asymptotic distribution, the population term

$$\mathrm{E}\left[\frac{\partial \boldsymbol{m}_i\left(\boldsymbol{\beta}_0\right)}{\partial \boldsymbol{\beta}_0'}\right]$$

can be estimated with

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{m}_i\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}'},$$

and

$$\boldsymbol{W} = \left(\mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta})\boldsymbol{m}_i(\boldsymbol{\beta})'\right]\right)^{-1}$$

with

$$\boldsymbol{W}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}_i(\boldsymbol{\beta})\boldsymbol{m}_i(\boldsymbol{\beta})'\right)^{-1}.$$

GMM estimator, the optimal one, calculation in two steps

Step one. Choose a preliminary weighting matrix

$$\boldsymbol{W}_n^1$$

to obtain a preliminary GMM estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}^1_{GMM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \bar{\boldsymbol{m}}_n(\boldsymbol{\beta})' \, \boldsymbol{W}_n^1 \bar{\boldsymbol{m}}_n(\boldsymbol{\beta}) \,.$$

For example,

$$\boldsymbol{W}_n^1 = \boldsymbol{I}.$$

This preliminary estimator is consistent because GMM estimator is so, but it is inefficient because $\boldsymbol{W}_n$ not optimal yet.

GMM estimator, the optimal one, calculation in two steps

Step two. Use the consistent $\hat{\boldsymbol{\beta}}^1_{GMM}$ to calculate

$$\boldsymbol{W}_n^2 = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}_i(\hat{\boldsymbol{\beta}}^1_{GMM}) \boldsymbol{m}_i(\hat{\boldsymbol{\beta}}^1_{GMM})' \right)^{-1}$$

which is a consistent estimate of the optimal weighting matrix

$$\boldsymbol{W} = \left(\mathrm{E} \left[\boldsymbol{m}_i(\boldsymbol{\beta}) \boldsymbol{m}_i(\boldsymbol{\beta})' \right] \right)^{-1}.$$

Use the estimated optimal weighting matrix to calculate the optimal GMM estimator

$$\hat{\boldsymbol{\beta}}^2_{GMM} = \hat{\boldsymbol{\beta}}^O_{GMM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\ \bar{\boldsymbol{m}}_n(\boldsymbol{\beta})' \boldsymbol{W}_n^2 \bar{\boldsymbol{m}}_n(\boldsymbol{\beta}).$$

GMM estimator, end. lin. model example

Consider the SLM with endogenous variables as before where

- A1.IV. Linearity
- A2.IV. Relevance of $\boldsymbol{z}_i$
- A3.IV. Exogeneity of $\boldsymbol{z}_i$
- A5.IV., Random sampling

hold. But allow that

- A4.IV. Homoskedasticity

is violated so that $\operatorname{Var}\left[\boldsymbol{z}_i \varepsilon_i\right]$ is $\operatorname{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i' \varepsilon_i \varepsilon_i\right]$ which is a finite positive definite matrix, and not $\operatorname{E}\left[\varepsilon_i \varepsilon_i \mid \boldsymbol{z}_i\right] = \sigma^2$. So heteroskedasticity is allowed!

GMM estimator, end. lin. model example

Our case is still for $L > K$.

GMM estimator, end. lin. model example

In this model, the moment is

$$\boldsymbol{m}_i(\boldsymbol{\beta}) = \boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)$$

Then,

$$\frac{\partial \boldsymbol{m}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = -\boldsymbol{x}_i \boldsymbol{z}_i'.$$

The sample moment condition is

$$\bar{\boldsymbol{m}}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right) = 0$$

The population moment condition is

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

We will now replace these terms in the general theory to obtain the results derived above for the SLM with endogenous variables example!

GMM estimator, end. lin. model example

The objective function is

$$
\begin{aligned}
q(\beta) &\equiv \bar{\boldsymbol{m}}_n(\beta)' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\beta) \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \beta \right) \right)' \boldsymbol{W}_n \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \left(y_i - \boldsymbol{x}_i' \beta \right) \right) \\
&= \left(\frac{1}{n} \boldsymbol{Z}' (\boldsymbol{y} - \boldsymbol{X}\beta) \right)' \boldsymbol{W}_n \left(\frac{1}{n} \boldsymbol{Z}' (\boldsymbol{y} - \boldsymbol{X}\beta) \right).
\end{aligned}
$$

GMM estimator, end. lin. model example

The minimisation problem is

$$\hat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, q(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \bar{\boldsymbol{m}}_n(\boldsymbol{\beta})' \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\boldsymbol{\beta}).$$

GMM estimator, end. lin. model example

The F.O.C. are

$$\frac{\partial q\left(\hat{\boldsymbol{\beta}}\right)}{\partial\hat{\boldsymbol{\beta}}} = -2\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}\right)\right) = \mathbf{0}.$$

This derivation involves matrix differentiation. Checkpoint.

GMM estimator, end. lin. model example

The F.O.C. are

$$\frac{\partial q\left(\hat{\boldsymbol{\beta}}_{GMM}\right)}{\partial \hat{\boldsymbol{\beta}}_{GMM}} = -2\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_{GMM}\right)\right) = \mathbf{0}.$$

Pay attention to the matrix dimension of the F.O.C. equation.

$$\underbrace{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n}_{\color{red}{K\times L}}\underbrace{\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_{GMM}\right)\right)}_{\color{blue}{L\times 1}} = \mathbf{0}.$$

These are K moment conditions and K unknown parameters. The system is exactly identified! Using a $K \times L$ matrix, we are choosing a particular way of linearly combining L moment conditions, so that the system becomes exactly identified!

GMM estimator, end. lin. model example

Solving for $\hat{\boldsymbol{\beta}}_{GMM}$,

$$\begin{aligned}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GMM}\right)\right) &= \mathbf{0}\\ \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} &= \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_{GMM}\\ \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} &= \hat{\boldsymbol{\beta}}_{GMM}.\end{aligned}$$

This solution assumes that $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}$ is invertible. We will show that our assumptions guarantee this asymptotically.

GMM estimator, end. lin. model example, consistency

$$\hat{\boldsymbol{\beta}}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}.$$

Using $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$,

$$\hat{\boldsymbol{\beta}}_{GMM} = \boldsymbol{\beta}_0 + \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{\varepsilon}.$$

Dividing and multiplying by n,

$$\hat{\boldsymbol{\beta}}_{GMM} = \boldsymbol{\beta}_0 + \left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{\varepsilon}}{n}.$$

GMM estimator, end. lin. model example, consistency

By assumption,

$$\boldsymbol{W}_n \xrightarrow{p} \boldsymbol{W}.$$

By the weak LLN moment by moment,

$$\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i' \xrightarrow{p} \mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right],$$

$$\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i' \xrightarrow{p} \mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right],$$

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

GMM estimator, end. lin. model example, consistency

Using the sum and product rules of plim (Greene, Theorem D.14),

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0 + \left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right]\right)^{-1} \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right].$$

Since we assume that

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

we have

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0.$$

GMM estimator, end. lin. model example, consistency

In

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0 + \left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right],$$

the matrix

$$\underbrace{\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]}_{K\times L}\underbrace{\boldsymbol{W}}_{L\times L}\underbrace{\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]}_{L\times K}$$

is $K \times K$. It has full column rank. This is because the last element of the term has full column rank by A2.IV. That is why the matrix has an inverse.

GMM estimator, end. lin. model example, asy. normality

For notational convenience define $\boldsymbol{H}_n$ and $\boldsymbol{H}$ as

$$\hat{\boldsymbol{\beta}}_{GMM} = \boldsymbol{\beta}_0 + \underbrace{\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n}_{\equiv \boldsymbol{H}_n}\frac{\boldsymbol{Z}'\boldsymbol{\varepsilon}}{n},$$

and

$$\hat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}_0 + \underbrace{\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}}_{\equiv \boldsymbol{H}}\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right].$$

GMM estimator, end. lin. model example, asy. normality

$$\hat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}_0 = \boldsymbol{H}_n \frac{\boldsymbol{Z}'\varepsilon}{n}.$$

Multiplying by $\sqrt{n}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}_0\right) = \boldsymbol{H}_n \frac{1}{\sqrt{n}} \boldsymbol{Z}'\varepsilon.$$

Consider the probability limits of

$$\boldsymbol{H}_n$$

and

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

GMM estimator, end. lin. model example, asy. normality

We already showed that

$$\boldsymbol{H}_n \xrightarrow{p} \boldsymbol{H}.$$

By A3.IV, $\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right] = 0$. By A5.IV, $\boldsymbol{z}_i\varepsilon_i$ is i.i.d.. Using these and applying the CLT (Greene, Theorem D.19A),

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{\varepsilon} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i \xrightarrow{d} N\big(0, \underbrace{\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right]}_{\equiv \boldsymbol{S}}\big).$$

GMM estimator, end. lin. model example, asy. normality

Using the product rule of limiting distributions (Greene, Theorem D.16)

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}_0\right) = \boldsymbol{H}_n \frac{1}{\sqrt{n}} \boldsymbol{Z}'\boldsymbol{\varepsilon} \xrightarrow{d} N\left(0, \boldsymbol{HSH}'\right).$$

Convergence in distribution implies (Greene, Section D.3) that

$$\hat{\boldsymbol{\beta}}_{GMM} \overset{a}{\sim} N\left[\boldsymbol{\beta}_0, \frac{1}{n}\boldsymbol{HSH}'\right].$$

$\hat{\boldsymbol{\beta}}_{GMM}$ is normal, and it does not require that ε_i is normal.

GMM estimator, end. lin. model example, asy. normality

$$
\begin{aligned}
\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] &= \frac{1}{n}\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}' \\
&= \frac{1}{n}\left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\right) \\
&\qquad \mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\varepsilon_i\varepsilon_i\right] \\
&\qquad \left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\right)'.
\end{aligned}
$$

We can estimate

$$
\begin{aligned}
\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right] \;&\text{by}\; \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i', \\
\boldsymbol{W} \;&\text{by}\; \boldsymbol{W}_n, \\
\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right] \;&\text{by}\; \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i
\end{aligned}
$$

where $\hat{\varepsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$.

GMM estimator, end. lin. model example, asy. normality

The consistent estimator of Asy. Var $\left[\hat{\boldsymbol{\beta}}_{GMM}\right]$ is then given by

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)$$
$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i$$
$$\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)'.$$

GMM estimator, end. lin. model example, asy. normality

Note the close relationship with the consistent

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)$$
$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i$$
$$\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)'.$$

and the consistent HCE, which we derived as

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{OLS}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\hat{\varepsilon}_i\boldsymbol{x}_i'\hat{\varepsilon}_i\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

$\hat{\boldsymbol{\beta}}_{GMM}$ does not require that ε_i to be homoskedastic!

GMM estimator, end. lin. model example, asy. efficiency

Note that

$$\hat{\boldsymbol{\beta}}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}$$

depends on $\boldsymbol{W}_n$! Note also that

$$\begin{aligned}\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}&\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)\\ &\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\\ &\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)'\end{aligned}$$

depends on $\boldsymbol{W}_n$!

GMM estimator, end. lin. model example, asy. efficiency

$\hat{\boldsymbol{\beta}}_{GMM}$ depends on $\boldsymbol{W}_n$. Different choices of $\boldsymbol{W}_n$ will give different point estimates. More on this later.

Est. Asy. Var $\left[\hat{\boldsymbol{\beta}}_{GMM}\right]$ depends on $\boldsymbol{W}_n$. Different choices of $\boldsymbol{W}_n$ will give different estimates for the asymptotic variance of $\hat{\boldsymbol{\beta}}_{GMM}$.

The consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{GMM}$ do not depend on the choice of $\boldsymbol{W}_n$, however. For any $\boldsymbol{W}_n$ these results hold.

If two researchers come up with a same model, and a same dataset, but with different choices of $\boldsymbol{W}_n$, they will report different point estimates with different variances, and both researchers can correctly claim that they are using a consistent estimator. This is a problem. We need a criterion for choosing $\boldsymbol{W}_n$.

GMM estimator, end. lin. model example, asy. efficiency

Among all possible $\boldsymbol{W}$, we could choose the one that leads to the smallest asymptotic variance for $\hat{\beta}_{GMM}$. It can be shown that

$$\boldsymbol{S}^{-1} = \left(\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \boldsymbol{z}_i' \varepsilon_i\right]\right)^{-1}$$

is the $\boldsymbol{W}$ that minimises Asy. Var $\left[\hat{\beta}_{GMM}\right]$. Hence, $\boldsymbol{S}^{-1}$ is the optimal choice for $\boldsymbol{W}$ because it makes $\hat{\beta}_{GMM}$ asymptotically efficient.

GMM estimator, end. lin. model example, asy. efficiency

$$
\begin{aligned}
\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n} & \Big(\left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] \right)^{-1} \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \Big) \\
& \mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \boldsymbol{z}_i' \varepsilon_i\right] \\
& \Big(\left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] \right)^{-1} \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \boldsymbol{W} \Big)'.
\end{aligned}
$$

Using $\left(\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \boldsymbol{z}_i' \varepsilon_i\right]\right)^{-1}$ for $\boldsymbol{W}$, the variance simplifies to

$$
\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n} \left(\mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{z}_i'\right] \left(\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i \boldsymbol{z}_i' \varepsilon_i\right] \right)^{-1} \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{x}_i'\right] \right)^{-1}.
$$

This is the lower bound for Asy. Var $\left[\hat{\boldsymbol{\beta}}_{GMM}\right]$.

GMM estimator, end. lin. model example, asy. efficiency

$$\text{Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}\left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\right)$$
$$\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\varepsilon_i\varepsilon_i\right]$$
$$\left(\left(\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\mathrm{E}\left[\boldsymbol{z}_i\boldsymbol{x}_i'\right]\right)^{-1}\mathrm{E}\left[\boldsymbol{x}_i\boldsymbol{z}_i'\right]\boldsymbol{W}\right)'.$$

Consider two asymptotic variances, one using the optimal weighting matrix

$$\boldsymbol{W} = \left(\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right]\right)^{-1},$$

and another using a different one

$$\boldsymbol{W} = \boldsymbol{W}^*.$$

It can be shown that the difference between the two asymptotic variances is positive semidefinite. That is, $\hat{\boldsymbol{\beta}}_{GMM}$ using the optimal weighting matrix has an asymptotic variance that is smaller than or equal to the asymptotic variance of any other $\hat{\boldsymbol{\beta}}_{GMM}$ using a different weighting matrix.

GMM estimator, end. lin. model example, best moments

We already argued that a consistent estimator of

$$\underbrace{\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right]}_{\equiv\boldsymbol{S}}$$

is

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i}_{\equiv\boldsymbol{S}_n}$$

with $\hat{\varepsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$. Hence, we can estimate $\boldsymbol{S}^{-1}$ by $\boldsymbol{S}_n^{-1}$. We can then use $\boldsymbol{S}_n^{-1}$ as the optimal weighting matrix.

GMM estimator, end. lin. model example, best moments

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)$$
$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\hat{\varepsilon}_i$$
$$\left(\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\right)'.$$

Using $\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}$ for $\boldsymbol{W}_n$, the variance simplifies to

$$\text{Est. Asy. Var}\left[\hat{\boldsymbol{\beta}}_{GMM}\right] = \frac{1}{n}\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}.$$

GMM estimator, end. lin. model example, best moments

Why do we choose

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}$$

as $\boldsymbol{W}_n$? What is the intuition?

GMM estimator, end. lin. model example, best moments

Our derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}_{GMM}$ implied that

$$\left(\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i\right]\right)^{-1} \xrightarrow{p} \left(\mathrm{Var}\left[\boldsymbol{z}_i\varepsilon_i\right]\right)^{-1}$$

$$= \left(\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right] - \mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right]\underbrace{\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\right]}_{\mathbf{0}}\right)^{-1}.$$

We can estimate

$$\left(\mathrm{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right]\right)^{-1}$$

by

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}.$$

The term includes L variances for L moments, $z_{ik}\varepsilon_i$, in $\boldsymbol{z}_i\varepsilon_i$. The value of the term is larger if the variances of the moments are smaller. Why do we want these moments to have a small variance?

GMM estimator, end. lin. model example, best moments

We favour a lower

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \hat{\varepsilon}_i \boldsymbol{z}_i' \hat{\varepsilon}_i,$$

which requires us to favour a lower

$$\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{z}_i \varepsilon_i\right].$$

That is, we favour a lower variance for

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{z}_i \varepsilon_i.$$

Why? If a moment has a small variance, the distribution of it is more concentrated around its mean. But we require this mean to be as close to zero as possible according to A2.IV, which is $\mathrm{E}\left[\boldsymbol{z}_i \varepsilon_i\right] = \mathbf{0}$. Hence, best of the stated moments are those with a small variance!

GMM estimator, end. lin. model example, best moments

Recall the criterion function used to derive $\hat{\boldsymbol{\beta}}_{GMM}$. It is given by

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \varepsilon_i \right)' \boldsymbol{W}_n \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \varepsilon_i \right).$$

We have L variances for L moments in

$$\boldsymbol{W}_n = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \hat{\varepsilon}_i \boldsymbol{z}_i' \hat{\varepsilon}_i \right)^{-1}.$$

Since a moment with a small variance is desirable, a moment with a small variance is given a higher weight in the criterion function. If the term inside the brackets is smaller, the inverse is larger!

GMM estimator, end. lin. model example, best moments

The estimated asymptotic variance of $\hat{\boldsymbol{\beta}}_{GMM}$ is

$$\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\right)^{-1}.$$

Moments with small variances lead to a lower variance for $\hat{\boldsymbol{\beta}}_{GMM}$!

GMM estimator, end. lin. model example, best moments

Use the estimated asymptotic variance of $\hat{\boldsymbol{\beta}}_{GMM}$,

$$\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\hat{\varepsilon}_i\boldsymbol{z}_i'\hat{\varepsilon}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\right)^{-1},$$

still for another implication. Larger values of the sample moments

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'$$

lead to a smaller variance. A larger value is desirable because it implies a stronger correlation between $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$, as required by the relevance assumption A2.IV. Hence, best of the stated sample moments are those with larger values!

GMM estimator, end. lin. model example

The GMM estimator is

$$\hat{\beta}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}.$$

The weighting matrix

$$\boldsymbol{W}_n$$

is optimal if it is equal to

$$\boldsymbol{S}_n^{-1}.$$

The optimal GMM estimator is

$$\hat{\beta}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{S}_n^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{S}_n^{-1}\boldsymbol{Z}'\boldsymbol{y}.$$

GMM estimator, end. lin. model example

The optimal GMM estimator is

$$\hat{\beta}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{S}_n^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{S}_n^{-1}\boldsymbol{Z}'\boldsymbol{y}.$$

How do we actually construct this? We do this in two steps.

GMM estimator, end. lin. mod. exa., two-step estimator

$$\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \hat{\varepsilon}_i \boldsymbol{z}_i' \hat{\varepsilon}_i$$

with $\hat{\varepsilon}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$.

GMM estimator, end. lin. mod. exa., two-step estimator

Step one. Choose a preliminary weighting matrix

$$\boldsymbol{W}_n^1$$

to obtain a preliminary GMM estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}_{GMM}^1 = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n^1\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n^1\boldsymbol{Z}'\boldsymbol{y}.$$

For example,

$$\boldsymbol{W}_n^1 = \boldsymbol{I}.$$

This preliminary estimator is consistent because GMM estimator is so, but it is inefficient because $\boldsymbol{W}_n$ not optimal yet.

GMM estimator, end. lin. mod. exa., two-step estimator

Step two. Use the consistent $\hat{\boldsymbol{\beta}}^1_{GMM}$ to calculate the consistent residual

$$\hat{\varepsilon}^1_i = y_i - \boldsymbol{x}'_i \hat{\boldsymbol{\beta}}^1_{GMM}.$$

Use $\hat{\varepsilon}^1_i$ to calculate the optimal weighting matrix

$$\boldsymbol{S}^2_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \hat{\varepsilon}^1_i \boldsymbol{z}'_i \hat{\varepsilon}^1_i.$$

Use $\boldsymbol{S}^{2^{-1}}_n$ as $\boldsymbol{W}^2_n$ to calculate the consistent and efficient (optimal) GMM estimator

$$\hat{\boldsymbol{\beta}}^2_{GMM} = \left(\boldsymbol{X}' \boldsymbol{Z} \boldsymbol{S}^{2^{-1}}_n \boldsymbol{Z}' \boldsymbol{X} \right)^{-1} \boldsymbol{X}' \boldsymbol{Z} \boldsymbol{S}^{2^{-1}}_n \boldsymbol{Z}' \boldsymbol{y}.$$

$\hat{\boldsymbol{\beta}}^2_{GMM}$ is not unique because it depends on the initial weighting matrix $\boldsymbol{W}^1_n$. What can we do?

GMM estimator, end. lin. mod. exa., iterated estimator

Iterated steps. Use $\hat{\boldsymbol{\beta}}^2_{GMM}$ obtained in the second step to update $\hat{\varepsilon}^1_i$ considered in the first step. Using the updated $\hat{\varepsilon}^2_i$, update $\boldsymbol{S}^2_n$ considered in the first step. Using the updated ${\boldsymbol{S}^3_n}^{-1}$, update $\hat{\boldsymbol{\beta}}^2_{GMM}$ considered in the second step. This gives $\hat{\boldsymbol{\beta}}^3_{GMM}$. Repeat this procedure for a large number of times. The estimates of $\hat{\boldsymbol{\beta}}_{GMM}$ will get closer to each other as the number of iterations increase. The approximation of the weighting matrix will stabilise with large number of iterations. Therefore, the iterated GMM estimator does not depend, or depends weakly, on the initial weighting matrix $\boldsymbol{W}^1_n$. Hence, we would not care too much about if we have chosen, e.g., in an ad-hoc manner the identity matrix as the initial weighting matrix.

GMM estimator, end. lin. mod. exa., iterated estimator

The two-step and the iterated GMM estimators are asymptotically equivalent. This is because $\hat{\beta}_{GMM}$ is a consistent estimator and therefore it will converge to the true parameter it is estimating, leading to a certain set of residuals and a weighting matrix.

GMM estimator, end. lin. mod. exa., special case: IV

We have considered that $\text{Var}\left[\boldsymbol{z}_i\varepsilon_i\right] = \text{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\varepsilon_i\varepsilon_i\right]$. Hence, heteroskedasticity is allowed. Now take the restrictive case that homoskedasticity holds, so that $\text{E}\left[\varepsilon_i\varepsilon_i \mid \boldsymbol{z}_i\right] = \sigma^2$. Using the LIE,

$$
\begin{aligned}
\text{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i\right] &= \text{E}_{\boldsymbol{z}_i}\left[\text{E}\left[\boldsymbol{z}_i\varepsilon_i\boldsymbol{z}_i'\varepsilon_i \mid \boldsymbol{z}_i\right]\right] \\
&= \text{E}_{\boldsymbol{z}_i}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\text{E}\left[\varepsilon_i\varepsilon_i \mid \boldsymbol{z}_i\right]\right] \\
&= \text{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\sigma^2\right] \\
&= \sigma^2\text{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i'\right].
\end{aligned}
$$

Estimate this by

$$
\hat{\sigma}^2\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i' = \hat{\sigma}^2\frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}.
$$

assuming that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

GMM estimator, end. lin. mod. exa., special case: IV

Note that

$$\left(\hat{\sigma}^2 \frac{1}{n} \boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} = \boldsymbol{S}_n{}^{-1} = \boldsymbol{W}_n.$$

Replacing in the formula for the optimal GMM estimator,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GMM} &= \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \hat{\boldsymbol{\beta}}_{2SLS}
\end{aligned}$$

since $\hat{\sigma}^2$ and $\frac{1}{n}$ cancel out! This shows that $\hat{\boldsymbol{\beta}}_{2SLS}$ is a particular case of $\hat{\boldsymbol{\beta}}_{GMM}$ with a particular choice of $\boldsymbol{W}_n$ if conditional homoskedasticity holds. It also shows how different choices of $\boldsymbol{W}_n$ produce different estimators.

GMM estimator, end. lin. mod. exa., special case: IV

$$\hat{\boldsymbol{\beta}}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}.$$

Suppose $L = K$. Hence, $\boldsymbol{X}'\boldsymbol{Z}$ is invertible. Then,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{GMM} &= \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{W}_n^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
&= \hat{\boldsymbol{\beta}}_{MM} \\
&= \hat{\boldsymbol{\beta}}_{IV}.
\end{aligned}$$

This shows that if $L = K$, $\boldsymbol{W}_n$ plays no role! Then $\hat{\boldsymbol{\beta}}_{MM} = \hat{\boldsymbol{\beta}}_{GMM}$.

GMM estimator, end. lin. mod. exa., special case: OLS

$$\hat{\beta}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}.$$

Suppose $\boldsymbol{Z} = \boldsymbol{X}$. Then, $L = K$. $\boldsymbol{X}'\boldsymbol{X}$ is invertible if A2 holds. $\boldsymbol{W}_n$ plays no role. We obtain

$$\begin{aligned}\hat{\beta}_{GMM} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \hat{\beta}_{MM} \\ &= \hat{\beta}_{OLS}.\end{aligned}$$

The case of $L = K$

The case of $L = K$.

MM estimator, derivation

$$\mathrm{E}\left[\boldsymbol{m}_i\left(\boldsymbol{\beta}\right)\right] = \mathbf{0}$$

is a vector of population moment conditions. Since $L = K$, it represents a system with the same number of equations as unknowns. The equality because the system can have a unique solution with some $\boldsymbol{\beta}$. Indeed, some econometric theory will require that it does with $\boldsymbol{\beta}_0$. Recall the exogeneity assumption on the instrument, when $L = K$, for example.

MM estimator, derivation

$$\bar{\boldsymbol{m}}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}_i(\boldsymbol{\beta}) = \mathbf{0}$$

is the corresponding sample mean if LLN applies. From here to the previous slide: if there is a solution in the sample mean, it is likely that there is a solution in the population, when LLN applies. The sample mean is what is relevant in practice. The system is exactly determined, and there is usually a solution. The solution is the MM estimator. That is, the $\boldsymbol{\beta}$ that solves the system is $\hat{\boldsymbol{\beta}}_{MM}$.

MM estimator, derivation

We can also obtain the MM estimator as

$$\hat{\boldsymbol{\beta}}_{MM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \bar{\boldsymbol{m}}_n(\beta)' \, \bar{\boldsymbol{m}}_n(\beta) \, .$$

That is, the GMM estimator is obtained by minimising the weighted quadratic distance

$$\hat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \bar{\boldsymbol{m}}_n(\beta)' \, \boldsymbol{W}_n \bar{\boldsymbol{m}}_n(\beta) \, .$$

In the case of $L = K$, we do not need to minimise the weighted quadratic distance. Minimising the quadratic distance is enough. This means that the MM estimator is a special case of the GMM estimator where $\boldsymbol{W}_n$ does not play a role.

MM estimator, asymptotic normality

The asymptotic distribution of the GMM estimator is

$$\hat{\beta}_{GMM} \overset{a}{\sim}$$

$$N\left[\boldsymbol{\beta_0}, \frac{1}{n} - \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \boldsymbol{W} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \boldsymbol{W} \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i(\boldsymbol{\beta}_0)'\right] \left(- \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \boldsymbol{W} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \boldsymbol{W}\right)'\right].$$

Since the MM estimator is a special case of the GMM estimator where the weighting matrix does not play a role, we can derive the asymptotic distribution of the MM estimator as a special case of the asymptotic distribution of the GMM estimator derived above:

$$\hat{\beta}_{MM} \overset{a}{\sim}$$

$$N\left[\boldsymbol{\beta_0}, \frac{1}{n} - \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i(\boldsymbol{\beta}_0)'\right] \left(- \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)'\right)'\right] =$$

$$N\left[\boldsymbol{\beta_0}, \frac{1}{n} - \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)'\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i(\boldsymbol{\beta}_0)'\right] \left(- \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)'\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)'\right)'\right] =$$

$$N\left[\boldsymbol{\beta_0}, \frac{1}{n} \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1} \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i(\boldsymbol{\beta}_0)'\right] \left(\left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1}\right)'\right] =$$

$$N\left[\boldsymbol{\beta_0}, \frac{1}{n} \left[\bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)' \mathrm{E}\left[\boldsymbol{m}_i(\boldsymbol{\beta}_0)\boldsymbol{m}_i(\boldsymbol{\beta}_0)'\right]^{-1} \bar{\boldsymbol{G}}(\boldsymbol{\beta}_0)\right]^{-1}\right].$$

MM estimator, end. lin. model example

In the SLM with endogenous variables, the population moment condition is

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

Hence, the moment is

$$\boldsymbol{m}_i\left(\boldsymbol{\beta}\right) = \boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right).$$

The sample moment condition is

$$\bar{\boldsymbol{m}}_n\left(\boldsymbol{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right) = 0.$$

MM estimator, end. lin. model example

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)=\boldsymbol{0}$$

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i y_i-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\boldsymbol{\beta}=\boldsymbol{0}$$

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i y_i=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\boldsymbol{\beta}$$

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i y_i=\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'\right]\boldsymbol{\beta}.$$

MM estimator, end. lin. model example

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i y_i = \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right] \boldsymbol{\beta}.$$

$\boldsymbol{z}_i$ is $L \times 1$. $\boldsymbol{x}_i$ is $K \times 1$. $\boldsymbol{\beta}$ is $1 \times K$. Suppose that $L = K$. Then, $\boldsymbol{z}_i$ is $K \times 1$. Hence, the number of equations (instruments) is the same as the number of unknowns (parameters). The system is exactly identified. Then,

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'$$

is a $K \times K$ square matrix. It is invertible! Then,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i y_i. \\ &= \hat{\boldsymbol{\beta}}_{MM}. \end{aligned}$$

MM estimator, end. lin. model example

Invertibility implies that the solution is unique. The unique $\hat{\boldsymbol{\beta}}$ that solves an exactly identified system of equations is the method of moments estimator, $\hat{\boldsymbol{\beta}}_{MM}$.

MM estimator, end. lin. model example, special case

Notice that

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i'\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i y_i \\
&= \left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1} \frac{1}{n}\boldsymbol{Z}'\boldsymbol{y} \\
&= \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1} \boldsymbol{Z}'\boldsymbol{y} \\
&= \hat{\boldsymbol{\beta}}_{IV} \\
&= \hat{\boldsymbol{\beta}}_{MM}.
\end{aligned}
$$

That is, the $\hat{\boldsymbol{\beta}}_{IV}$ is a special case of $\hat{\boldsymbol{\beta}}_{MM}$ when the number of instruments (equations) is the same as the number of endogenous variables (parameters), that is when $\boldsymbol{Z}'\boldsymbol{X}$ is invertible.

MM estimator, lin. model example, special case

If $\boldsymbol{x}_i$ is exogenous, or if $\boldsymbol{x}_i$ is an instrument for itself,

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

Solving the sample counterpart

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\left(y_i - \boldsymbol{x}_i'\hat{\beta}\right) = \mathbf{0},$$

we have

$$\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{X}\hat{\beta}.$$

These are just the normal equations of the least squares problem! Since $\boldsymbol{X}\boldsymbol{X}'$ is invertible by A2, we can solve for $\hat{\boldsymbol{\beta}}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \hat{\boldsymbol{\beta}}_{OLS} \\ &= \hat{\boldsymbol{\beta}}_{MM}.\end{aligned}$$

That is, the $\hat{\boldsymbol{\beta}}_{OLS}$ is a special case of $\hat{\boldsymbol{\beta}}_{MM}$ when $\boldsymbol{x}_i$ is exogenous.