Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

(Generalized) method of moments

Koen Jochmans François Poinas

2025 — 2026

Motivation

In the classical linear regression model

$$Y = X'\beta + e, \qquad e|X = x \sim N(0, \sigma^2),$$

the best estimator of β is the maximum likelihood estimator.

Can also write down a likelihood for other distributions.

But the likelihood depends on the distribution of $e|X = x$.

The ordinary least-squares estimator is consistent and asymptotically normal when only

$$\mathbb{E}(Xe) = 0,$$

which is a much weaker requirement.

Is there a similar result for more general probems?

The negative exponential-growth model with parameters (α, β) was

$$Y = \alpha(1 - e^{-\beta X}) + e, \qquad \mathbb{E}(e|X = x) = 0.$$

There is no distributional assumption included in the specification of this model.

So, we cannot write down the likelihood function.

However, we do know that

$$\mathbb{E}(Y - \alpha(1 - e^{-\beta X})|X = x) = 0,$$

and can construct method-of-moment estimators from this.

Nonlinear least squares

Consider a problem where

$$Y = m(X;\theta_0) + e, \qquad \mathbb{E}(e|X = x) = 0.$$

Then θ_0 minimises the nonlinear least squares problem

$$\mathbb{E}\left(\{Y - m(X;\theta)\}^2\right)$$

with respect to θ.

Indeed,

$$\mathbb{E}\left(\{Y - m(X;\theta)\}^2\right) = \mathbb{E}(e^2) + \mathbb{E}\left(\{m(X;\theta_0) - m(X;\theta)\}^2\right)$$

is globally minimised at any θ for which the second term is equal to zero. This includes θ_0.

(Global) identification requires sufficient variation in the variable X.

In the linear model

$$\mathbb{E}\left(\{m(X;\theta_0) - m(X;\theta)\}^2\right) = (\theta - \theta_0)' \, \mathbb{E}(XX')(\theta - \theta_0).$$

By linear independence,

$$\alpha' \, \mathbb{E}(XX') \, \alpha > 0$$

for any $\alpha \neq 0$ because the matrix $\mathbb{E}(XX')$ is positive definite in this case.

The implied estimator is

$$\arg\min_{\theta} \sum_{i=1}^{n} (Y_i - m(X_i;\theta))^2$$

which has first-order condition

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(X_i;\theta)}{\partial \theta} (Y_i - m(X_i;\theta)) = 0$$

and Hessian

$$Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 m(X_i;\theta)}{\partial \theta \partial \theta'} (Y_i - m(X_i;\theta)) - \frac{1}{n} \sum_{i=1}^{n} \frac{\partial m(X_i;\theta)}{\partial \theta} \frac{\partial m(X_i;\theta)}{\partial \theta'}$$

We can use an expansion around θ_0 of the first-order condition to write

$$(\hat{\theta} - \theta_0) = -Q_n(\theta_*)^{-1} S_n(\theta_0).$$

Here,

$$Q_n(\theta_*) \underset{p}{\to} Q = -\mathbb{E}\left(\frac{\partial m(X_i;\theta_0)}{\partial \theta} \frac{\partial m(X_i;\theta_0)}{\partial \theta'} \right)$$

while

$$\sqrt{n}\, S_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial m(X_i;\theta_0)}{\partial \theta} e_i \underset{d}{\to} N(0, \varOmega)$$

for

$$\varOmega = \mathbb{E}\left(\frac{\partial m(X_i;\theta_0)}{\partial \theta} \frac{\partial m(X_i;\theta_0)}{\partial \theta'} e^2. \right)$$

Thus,

$$\sqrt{n}(\hat{\theta} - \theta_0) \underset{d}{\to} N(0, Q^{-1} \varOmega\, Q^{-1})$$

as $n \to \infty$.

$Y \geq 0$ with

$$\mathbb{E}(Y|X = x) = \exp(x'\theta_0).$$

Nonlinear least-squares minimises

$$\sum_{i=1}^{n} (Y_i - \exp(X_i'\theta))^2$$

and so solves

$$\sum_{i=1}^{n} X_i \exp(X_i'\theta)(Y_i - \exp(X_i'\theta)) = 0.$$

Here,

$$Q = -\mathbb{E}\left(XX' \exp(X'\theta_0)^2\right), \qquad \Omega = \mathbb{E}\left(XX' \exp(X'\theta_0)^2 e^2\right).$$

When $Y|X = x$ is Poisson distributed with mean $\exp(x'\theta_0)$ the score equation of maximum likelihood is

$$\sum_{i=1}^{n} X_i(Y_i - \exp(X_i'\theta)) = 0.$$

This is not nonlinear least squares. However, in this example, it remains true that

$$\mathbb{E}(X(Y - \exp(X'\theta_0))) = 0,$$

and so the estimator will be consistent and will have a normal limit distribution.

Optimal estimation in conditional mean problems

When

$$\mathbb{E}(Y - m(X;\theta_0)|X = x) = 0,$$

we have, by the law of iterated expectations, that

$$\mathbb{E}(\varphi(X)\,(Y - m(X;\theta_0))) = 0$$

for any function φ.

So we can in principle construct many estimators, by solving

$$\sum_{i=1}^{n} \varphi(X_i)\,(Y_i - m(X_i;\theta)) = 0$$

for θ, or even by minimizing

$$\left(\sum_{i=1}^{n}(\varphi(X_i)\,(Y_i - m(X_i;\theta)))\right)' W \left(\sum_{i=1}^{n}(\varphi(X_i)\,(Y_i - m(X_i;\theta)))\right)$$

with respect to θ.

Generalized least squares

The optimal choice of φ turns out to be

$$\varphi(x;\theta) = -\frac{\partial m(x;\theta)}{\partial \theta}\,\frac{1}{\sigma^2(x)},$$

where $\sigma^2(x) = \mathbb{E}(e^2|X=x)$.

The optimal moment condition becomes

$$\sum_{i=1}^{n} \frac{\partial m(X_i;\theta)}{\partial \theta}\,\frac{(Y_i - m(X_i;\theta))}{\sigma^2(X_i)} = 0.$$

Under homoskedasticity, this is least squares.

In nonlinear models heteroskedasticity is standard.

In the Poisson model, for example, $\sigma^2(x) = \exp(x'\theta_0)$.

Choose (infinite) consumption stream $\{c_i\}$ as to maximize expected discounted utility stream

$$\mathbb{E}\left(\sum_{h=0}^{\infty} \alpha_0^h\, u(c_{i+h}, \beta_0) \middle| I_i\right),$$

where I_i is the information set available to the agent at time period i.

Here, u is a chosen utility function parametrized by β_0 and α_0 is the discount rate

Aim to learn $\theta_0 = (\alpha_0, \beta_0)$.

Must have that

$$\alpha_0^i \frac{\partial u(c_i, \beta_0)}{\partial c} = \alpha_0^{i+1} \frac{\partial u(c_{i+1}, \beta_0)}{\partial c} r,$$

where r is the return on a risk-free asset (taken to be constant over time here).

This indifference requirement is a characteristic of optimal behavior.

It implies that

$$\mathbb{E}\left(r\, \alpha_0 \frac{\partial u(c_{i+1}, \beta_0) \big/ \partial c}{\partial u(c_i, \beta_0) \big/ \partial c} - 1 \middle| I_i \right) = 0$$

holds.

This yields moment conditions that can be used to construct a GMM estimator.

For random variable $V = (Y, X, Z)$, our model is summarised by the moment equation

$$\mathbb{E}(g(V;\theta_0)) = 0,$$

where g is a known function.

If $\dim g \geq \dim \theta$ we construct a GMM estimator as the minimiser of

$$\left(n^{-1}\sum_{i=1}^{n} g(V_i;\theta)\right)' W \left(n^{-1}\sum_{i=1}^{n} g(V_i;\theta)\right)$$

for weight matrix W.

At the optimum,

$$\left(n^{-1}\sum_{i=1}^{n}\frac{\partial g(V_i;\hat{\theta})}{\partial\theta'}\right)' W \left(n^{-1}\sum_{i=1}^{n} g(V_i;\hat{\theta})\right) = 0.$$

Here,

$$n^{-1}\sum_{i=1}^{n} g(V_i;\hat{\theta}) = n^{-1}\sum_{i=1}^{n} g(V_i;\theta_0) + \left(n^{-1}\sum_{i=1}^{n}\frac{\partial g(V_i;\theta_*)}{\partial\theta'}\right)(\hat{\theta}-\theta_0)$$

and, also,

$$n^{-1}\sum_{i=1}^{n}\frac{\partial g(V_i;\hat{\theta})}{\partial\theta'} \underset{p}{\rightarrow} \mathbb{E}\left(\frac{\partial g(V_i;\theta_0)}{\partial\theta'}\right) =: G'.$$

Therefore,

$$GW\left(n^{-1}\sum_{i=1}^{n} g(V_i;\theta_0) + G'(\hat{\theta}-\theta_0)\right) = o_p(1)$$

But then

$$\hat{\theta} - \theta_0 = (GWG')^{-1}GW \left(n^{-1} \sum_{i=1}^{n} g(V_i; \theta_0) \right) + o_p(1)$$

and so, with

$$Q := (GWG')^{-1}GW, \qquad \Omega = \mathbb{E}(g(V;\theta_0)g(V;\theta_0)')$$

we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) \underset{d}{\rightarrow} N(Q\Omega Q').$$

The optimal choice for W is again Ω^{-1}.

All estimators we have seen are special cases of this.