GMM II

Ethan Ligon

April 18, 2022

The GMM Estimation Problem

We start with a set of ℓ equations implied by a parametric model of behavior which is meant to hold for observations (e.g., people, firms, households) indexed by j, say

$$\mathbb{E}g_j(\beta) = \mathbf{0}_\ell. \tag{1}$$

Here β is a k-vector of parameters, and where $g_j(b)$ can be computed for each observation i (i.e., g_i is a known function) for all $b \in B \subset \mathbb{R}^k$, where B is a compact set with $\beta \in B$. Each function $g_j : B \to \mathbb{R}^\ell$.

Solving the GMM Estimation Problem

Consider the GMM Criterion associated with a weighting matrix $\boldsymbol{A}$ (so not necessarily efficient!):

$$b_N^A = \operatorname*{arg\,min}_{b \in B} N \boldsymbol{g}_N(b)^\top \boldsymbol{A} \boldsymbol{g}_N(b).$$

Provided $g_j(b)$ is continuously differentiable the GMM estimator will satisfy the first order conditions:

$$\frac{1}{N} \sum_{j=1}^{N} \frac{\partial g_j(b)}{\partial b^\top} \boldsymbol{A} \boldsymbol{g}_N(b) = \mathbf{0}_k.$$

Limiting Distribution

Let $\boldsymbol{Q}_N(b) = \frac{\partial g_j(b)}{\partial b^\top}$, an $\ell \times k$ matrix. Then provided β is in the interior of B we'll have

$$\mathbb{E}\boldsymbol{Q}_N(\beta)^\top \boldsymbol{A}\boldsymbol{g}_N(\beta) = \mathbf{0}_k.$$

Let $\mathbb{E}\boldsymbol{Q}_N(\beta) = \boldsymbol{Q}$. What's the limiting distribution of the GMM estimator?

Linear GMM

The linear case is particularly simple to compute, because estimates of the matrix $\boldsymbol{Q}$ don't depend on parameters. We'll focus on the IV problem (but this generalizes straight-forwardly to any linear regression problem):

$$y = X\beta + u \qquad \mathbb{E}(Z^\top u) = 0.$$

Here $g_j(b) = Z_j(y_j - X_j b)$ and $\boldsymbol{g}_N(b) = \boldsymbol{Z}^\top(\boldsymbol{y} - \boldsymbol{X}b)/N$, then sample average of the $g_j(b)$s. And note that

$$\boldsymbol{Q}_N(b) = \partial \boldsymbol{g}_N(b)/\partial b^\top = \boldsymbol{Z}^\top \boldsymbol{X}.$$

Finite Sample Performance of GMM Estimator

See notebook exploring finite sample performance of GMM.

Efficient GMM

We've seen that an "efficient" GMM estimator solves

$$b_N = \operatorname*{arg\,min}_{b \in B} \boldsymbol{g}_N(b)^\top \Omega^{-1} \boldsymbol{g}_N(b)),$$

where

$$\Omega = \mathbb{E} g_j(\beta) g_j(\beta)^\top - \mathbb{E} g_j(\beta) \mathbb{E} g_j(\beta)^\top.$$

Now we want to think about what "efficient" means.

Maximum Likelihood & Efficiency

What's an efficient consistent estimator? The one with the smallest covariance matrix. (Note that this ordering of estimators may be different in finite samples than it is in the limit.)

Maximum Likelihood

Suppose we have a continuous random variable X with density $f(X|\beta)$; we obtain realizations $\boldsymbol{X} = (X_1, \dots, X_N)$. Then the maximum likelihood estimator for β is

$$b_{MLE} = \underset{b \in B}{\arg\max} \frac{1}{N} \sum_{j=1}^{N} \log f(X_j|b).$$

Provided f is continuously differentiable, this implies first order conditions ("Scores"):

$$\frac{1}{N} \sum_{j} \frac{\partial f(X_j|b)/\partial b^\top}{f(X_j|b)} = 0$$

Efficient Scores

If β is in the interior of B and the support of X isn't a function of β, then we have

$$\mathbb{E}\frac{\partial f(X_j|\beta)/\partial b^\top}{f(X_j|\beta)} = 0;$$

In this case we say the "scores are efficient". But note this is a set of moment conditions! So "efficient scores" implies that we can pose maximum likelihood problem as a GMM problem.

Information Matrix

Let

$$g_j(b) = \frac{\partial f(X_j|b)/\partial b^\top}{f(X_j|b)};$$

If scores are efficient

$$\mathbb{E}g_j(\beta) = 0.$$

We've already seen that the optimal weighting matrix for GMM is the inverse of

$$\Omega = \mathbb{E}g_j(\beta)g_j(\beta)^\top - \mathbb{E}g_j(\beta)\mathbb{E}g_j(\beta)^\top.$$

But in this ML context Ω has another interpretation: this is the information matrix $\mathcal{I}$.

Interpretation

We can interpret the information matrix as measuring the information X conveys about the vector of parameters β.

Cramér-Rao Lower Bound

Recall: Under the same conditions for the score to be efficient, then the variance of any unbiased estimator V_b satisfies

$$V_b \geq \frac{1}{N}\mathcal{I}^{-1}$$

So: GMM exploiting efficient scores asymptotically achieves the Cramér-Rao bound, in the sense that

$$\sqrt{N}(b_N - \beta) \xrightarrow{d} Z \qquad \text{with } \mathbb{E}Z = 0 \text{ and } \mathrm{Var}(Z) = \mathcal{I}^{-1}.$$

Maximum Likelihood Efficiency

GMM exploiting scores as moment conditions achieves the same efficiency as ML using the same information. Note that this is a *just-identified* estimator by construction. Some questions:

- We've assumed parametric density functions. How does this affect our interpretation of efficiency?
- Why is it not possible to *improve* efficiency by exploiting additional moment conditions? (Or is it?)

Interpreting Efficiency Bounds

Maximum likelihood asymptotically achieves the Cramér-Rao lower bound; in this sense it makes efficient use of the information provided in the estimation. What is this information?