GENERALIZED METHOD OF MOMENTS

ETHAN LIGON

1. Generalized Method of Moments

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 $g_j : B \to \mathbb{R}^\ell$.

- **$k = \ell$:** Just identified;
- **$k > \ell$:** Under identified;
- **$k < \ell$:** Over (Über) identified.

(1) Well-specified models A well-specified model will have a set (1) which uniquely determine β; i.e., for which there's a unique solution.

2. Production Example

We talked earlier about the problem facing a price-taking firm; that problem led to a collection of first order conditions of the form

$$\mathbb{E}(pF_i(\boldsymbol{x})|\boldsymbol{Z}) - w_i = 0 \qquad i = 1, \dots, m,$$

where F_i is the partial derivative of the production function w.r.t. the ith input. If we choose some parametric form for F (with say k parameters) and some moments implied by the CEF then we'd obtain a problem which takes the method of moments form.

(1) Causality? These moment conditions aren't obviously interpretable as regressions, may lack some of the implicit causal structure we saw in structural equations, Cowles' commission style.

. *Date*: May 8, 2020.

3. Analogy Principle

If our model has $\mathbb{E}g_j(\beta) = 0$ and we have a independent sample of observations $j = 1, \dots, N$ we construct the obvious sample analog:

$$\boldsymbol{g}_N(b) = \frac{1}{N}\sum_{j=1}^{N} g_j(b)$$

which has the property that $\mathbb{E}\boldsymbol{g}_N(\beta) = \mathbb{E}g_j(\beta)$.

(1) Identification & the Analogy Principle One way of thinking about the identification of a model with k unknown parameters is that each parameter can be written as a function of moments of observable variables. Once conceived of this way the analogy principle makes passage to actual estimation rather immediate.

4. Examples

- Producer
- OLS
- Linear IV
- Central moments
- Non-linear Least Squares

5. Overidentification

If we have ℓ moment restrictions and $k < \ell$ parameters, then even though we may have $\mathbb{E}g_j(\beta) = 0$ in theory, we'll basically *never* have a b which solves the overidentified $\boldsymbol{g}_N(b) = 0$, if only because of sampling variation.

(1) What to do?

6. Criterion

It may be better to consider formulating the problem of solving a system of moment equations as a minimization problem.

(1) Least Squares Criterion? One Idea: Treat each sample moment as though it was an observation. And in a sense it is! But even though each sample moment is a random variable, if we can apply a Central Limit Theorem then we know something about its limiting distribution that we don't know in the usual regression setting.

$$\min_{b\in B} \frac{1}{\ell}\sum_{i=1}^{\ell} \boldsymbol{g}_{iN}(b)^2 = \frac{1}{\ell}\boldsymbol{g}_N(b)^\top \boldsymbol{g}_N(b).$$

Under standard conditions this defines an estimator which is consistent and asymptotically normal.

7. Weighting

In general we can do better! The least squares criterion above is like OLS, in that it assigns the same weight to each moment restriction. But maybe different weights should be assigned to different moment conditions?

(1) Different Weight Choose $\boldsymbol{A}$ positive definite, and solve:

$$\min_{b \in B} \boldsymbol{g}_N(b)^\top \boldsymbol{A} \boldsymbol{g}_N(b).$$

8. Extremum Estimators

Posed in the form above, we can regard GMM as an *extremum* estimator (or M-estimator). We have some excellent general results on the asymptotic behavior of such estimators, including standard results on consistency and central limit theorems (Newey and McFadden 1994).

(1) Extremum Estimators If there exists a Q such that

$$\beta = \arg\max_{b \in B} Q(b) \qquad \beta \text{ unique},$$

then an extremum estimator solves sample analog problem

$$b_N = \arg\max_{b \in B} Q_N(b)$$

9. Consistency

We need:

(1) $Q(b)$ continuous on B;
(2) B compact;
(3) $\beta \in B$ unique maximizer of $Q(b)$.
(4) $Q_N(b) \to Q(b)$ *uniformly*

(1) Uniform Law of Large Numbers If the above conditions hold, then $b_N \xrightarrow{p} \beta$.

10. Efficient GMM

Back to our weighting problem. What is the right weighting matrix? Standard answer draws on GLS logic: the matrix that minimizes the variance of a consistent estimator b_N.

(1) Efficient GMM as a GLS Estimator We can use a Gauss-Markov sort of argument to derive the optimal weighting matrix. Let

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

This is positive definite; let $D\Sigma^2 D^\top = \Omega$ be the spectral decomposition of Ω, with Σ^2 a diagonal matrix with the ℓ eigenvalues σ_i^2 of Ω on the diagonal.

Let $\boldsymbol{S} = (\boldsymbol{D}\Sigma^{-1})$. Then $\sqrt{N}\boldsymbol{S}^\top \boldsymbol{g}_N(\beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_\ell)$.

11. Overidentification Tests

We're familiar with Wald tests for testing restrictions in a linear regression. A similar logic allows us to construct tests of the moment restrictions in GMM. In particular, for the efficient GMM estimator our criterion function is

$$J_N = \min_{b\in B} N(\boldsymbol{g}_N(b)^\top \boldsymbol{S})(\boldsymbol{S}^\top \boldsymbol{g}_N(b)),$$

which is asymptotically the sum of ℓ products of standard normal random variables, which any statistics textbook will remind us has a $\chi^2_{\ell-k}$ distribution.