Section 11: GMM

ARE 210

November 14, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section11" folder.

1 Definitions

- Math tricks
  - ULLN (defined below) used repeatedly for convergence of functions, when we want to apply LLN but don't satisfy requirements
  - **MVT**: $h : \mathbf{R}^p \to \mathbf{R}^q$ continuously differentiable, then $h(b) = h(\theta) + \frac{\partial h(\tilde{b})}{\partial b}(b - \theta)$, where $\tilde{b} = \alpha b + (1-\alpha)\theta$ for $\alpha \in (0, 1)$
- **GMM**: $\hat{\theta}_n = \arg\max_{b \in \Theta} -\frac{1}{2}[\frac{1}{n}\sum_{i=1}^n m(W_i, b)]^{\mathrm{T}} S_n(W)[\frac{1}{n}\sum_{i=1}^n m(W_i, b)]$
  - **Identification**: Let $Q_0(b) = -\frac{1}{2}\mathbf{E}[m(W,b)]^{\mathrm{T}} S \mathbf{E}[m(W,b)]$. Suppose $\mathbf{E}[m(W,b)] = 0 \Leftrightarrow b = \theta$. Then if S is strictly positive definite, $\theta = \arg\max_b Q_0(b)$.
  - **Consistency 1**: Suppose (i) Θ is a compact subset of $\mathbf{R}^d$, (ii) $m(W,b)$ is continuous in b, (iii) $m(W,b)$ is measurable in W, and (iv) $S_n(W) \xrightarrow{p} S$, where S is symmetric positive definite. Suppose further (1) assumptions made above for identification, and (2) $\mathbf{E}[\sup_{b\in\Theta} |m(W,b)|] < \infty$. Then $\hat{\theta}_n \xrightarrow{p} \theta$.
  - **Consistency 2**: Replace (i), (ii), and (2) with $Q_n(b)$ is concave and $\mathbf{E}[m(W,b)]$ exists and is finite
  - **Asymptotic normality**: Let $m_n(b) = \sum_{i=1}^n m(W_i, b)$, $M_n(b) = \frac{\partial m_n(b)}{\partial b}$, and $M(b) = \mathbf{E}[\frac{\partial m(W,b)}{\partial b}]$. Then assuming 1) $M_n(\hat{\theta}_n) W_n M_n(\tilde{b}_n)'$ is invertible, for $\tilde{b}_n = \alpha\hat{\theta}_n + (1-\alpha)\theta$ for $\alpha \in [0,1]$, 2) $M_n(\hat{\theta}_n) \xrightarrow{p} M(\theta)$, and 3) assumptions made above for consistency, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$ where $V = (M(\theta) S M(\theta)^{\mathrm{T}})^{-1} M(\theta) S V[m(W,\theta)] S M(\theta)^{\mathrm{T}} (M(\theta) S M(\theta)^{\mathrm{T}})^{-1}$
  - **Efficient GMM**: $W = V[m(W,\theta)]^{-1}$ minimizes the variance of the GMM estimator.
    - Two step estimation : 1) Estimate $\hat{\theta}$ using simple weighting matrix W_0, e.g. identity matrix. 2) Estimate $W_e = V[m(W,\hat{\theta})]^{-1}$. 3) Estimate $\hat{\theta}_e$ using W_e.

* For inference, use $\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} M(X_i,\hat{\theta})W_0 m(X_i,\hat{\theta}) \\ m_1(x_i,\hat{\theta})m_1(x_i,\hat{\theta}) - W_{e,11} \\ m_2(x_i,\hat{\theta})m_1(x_i,\hat{\theta}) - W_{e,21} \\ \vdots \\ m_d(x_i,\hat{\theta})m_d(x_i,\hat{\theta}) - W_{e,dd} \\ M(X_i,\hat{\theta}_e)W_e m(X_i,\hat{\theta}_e) \end{bmatrix} = 0$
* ...or bootstrap (which we haven't defined yet)

2 Some useful bits

1. Estimation strategies (fancier than OLS/2SLS) we saw in class that you'll see most frequently "in the wild" are GMM, NLLS, and MLE (in no particular order), so we'll spend a little more time on GMM.

3 Practice questions

1)

a) Suppose we draw a random sample of individuals from a large population to estimate the average income. We have two unbiased observations of individual i's income μ_i - $X_i^T = (x_{1i}, x_{2i})$. Construct the efficient GMM estimator of the population average income μ. Interpret the optimal weighting matrix and the estimator.

b) (Random effects) Now assume for simplicity $V[X_i] = \mathbf{I}$. Also assume $\mu_i \sim N(\mu, \sigma^2)$, and assume for simplicity μ and σ^2 are known. What is $\hat{\mu}_{i,RE} \equiv \mathbf{E}[\mu_i|X_i]$? Compare this estimator to $\hat{\mu} = \overline{X}$.

2) (Overidentified linear IV) Consider the linear instrumental variable model from class. $\{y_i, x_i, z_i\}$ are independently and identically distributed, where y_i and x_i are scalars and z_i is a 2x1 vector, and we impose the moment restriction $\mathbf{E}[z_i(y_i - x_i\beta)] = 0$, and assume that y_i, x_i, and z_i are all mean 0.

a) Show that the GMM estimator using the weighting matrix $W = \hat{V}[z_i]^{-1}$ is just $\hat{\beta}_{2SLS}$. Justify this choice of weights.
Hint: See your answer to 1).
Note: This is called "Mahalanobis distance", which may seem weird but makes sense when we think of GMM as a minimum distance estimator, with the choice

of weights allowing us to calculate the distance between the moments and 0.

b) Show how to construct the efficient GMM estimator. Interpret the weighting matrix used here. Compare it to the weighting matrix in part a).