Proof of consistency of MLE

By uniqueness assumption, $\exists \eta > 0$ s.t. whenever $\|\theta - \theta_0\| \geq \varepsilon$, we have
$M(\theta_0) \geq M(\theta) + \eta \iff M(\theta_0) - M(\theta) \geq \eta$

$$\Rightarrow \Pr(\|\hat{\theta}_n - \theta_0\| \geq \varepsilon) \leq \Pr(M(\theta_0) - M(\hat{\theta}_n) \geq \eta)$$
$$= \Pr(M(\theta_0) - \underbrace{M_n(\hat{\theta}_n)}_{\geq M_n(\theta_0) \text{ since } \hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} M_n(\theta)} + M_n(\hat{\theta}_n) - M(\hat{\theta}_n) \geq \eta)$$
$$\leq \Pr(M(\theta_0) - M_n(\theta_0) + M_n(\hat{\theta}_n) - M(\hat{\theta}_n) \geq \eta)$$
$$\leq \Pr(2 \sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \geq \eta) \longrightarrow 0 \text{ by uniform convergence}$$
$$\Rightarrow \lim_{n \to \infty} \Pr(\|\hat{\theta}_n - \theta_0\| \geq \varepsilon) = 0 \Rightarrow \hat{\theta}_n \xrightarrow{p} \theta_0.$$

GMM

$E_0[\varphi(Y_i, X_i; \theta_0)] = 0$ where $\dim(\varphi) = \ell$, $\dim(\theta_0) = k$, where $k \leq \ell$

eg: OLS

$$E_0[X_i \varepsilon_i] = 0 \iff E_0[\overbrace{X_i(Y_i - X_i'\theta_0)}^{\equiv \varphi(Y_i, X_i; \theta_0)}] = 0$$

eg: IV

$$E_0[Z_i \varepsilon_i] = 0 \iff E_0[\overbrace{Z_i(Y_i - X_i'\theta_0)}^{\equiv \varphi(Y_i, X_i; \theta_0)}] = 0 \text{ where } \dim(Z_i) = \dim(\varphi) = \ell$$

if $\dim(Z_i) > \dim \theta$, then we have overidentification.

Defn: (GMM) $\hat{\theta}_{GMM} \equiv \operatorname{argmin}_{\theta \in \Theta} Q_n(\theta)$ where $Q_n(\theta) \equiv \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, X_i; \theta)\right)'}_{1 \times \ell} \underbrace{V_n^{-1}}_{\ell \times \ell} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, X_i; \theta)\right)}_{\ell \times 1}$

sample moment condition

V_n is some weight matrix. We will want to choose V_n to be the variance of the sample moment conditions. (ie we weight stuff with more variance less.)

Example: IV

$$Q_n = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)' V_n^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)$$

Theorem: (Asymptotic Distribution)

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, \Lambda(\theta_0))$$

where $\Lambda(\theta_0) = \Delta W \Delta'$, $\Delta = (AV^{-1}A')^{-1} A V^{-1}$, $A = E_0\left[\frac{\partial \varphi(Y_i, X_i; \theta_0)}{\partial \theta}\right]$,
$W = E_0[\varphi(Y_i, X_i; \theta_0)\, \varphi(Y_i, X_i; \theta_0)']$, and $V_n \xrightarrow{p} V$.

<u>Lemma:</u> (Optimal V)

The asymptotic variance $\Lambda(\theta_0)$ of $\hat{\theta}_{GMM}$ will be minimized when $V = W$.

<u>Corollary</u> When $V = W$, $\Lambda(\theta_0) = (A W^{-1} A')^{-1}$.

<u>Example</u>: IV

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)' V_n^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right), \quad V_n \to W$$

and $\varphi(Y_i, X_i; \theta) = Z_i \varepsilon_i \Rightarrow W = E[Z_i Z_i' \varepsilon_i^2]$

<u>Simplification</u>: Homoskedasticity: $E[\varepsilon_i^2 | Z_i] = \sigma^2 \Rightarrow W = \sigma^2 E[Z_i Z_i']$

$$\Rightarrow \hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)' \left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\varepsilon_i\right)$$

$$\Rightarrow \hat{\theta}_{GMM} = (X' P_Z X)^{-1} X' P_Z Y = \hat{\theta}_{2SLS}$$

The asymptotic covariance is thus given by

$$\Lambda = (A W^{-1} A)^{-1} \quad \text{where } A = E_0\left[\frac{\partial \varphi(Y_i, X_i; \theta_0)}{\partial \theta}\right] = E_0\left[\frac{\partial Z_i(Y_i - X_i'\theta_0)}{\partial \theta}\right]$$

$$= E_0[X_i Z_i']$$

$$W = \sigma^2 E_0[Z_i Z_i']$$

$$\Rightarrow \Lambda = \sigma^2 \left(E_0[X_i Z_i'] E_0[Z_i Z_i']^{-1} E_0[X_i Z_i']'\right)^{-1}$$

$$\Rightarrow \hat{\Lambda} = \hat{\sigma}^2 (X'Z(Z'Z)^{-1} Z'X)^{-1}$$

$$= \hat{\sigma}^2 (X' P_Z X)^{-1}$$

What if we have heteroskedasticity? Use two-step GMM.