

# MGE-201: Extremum estimator

Felipe Osorio

[felipe.osorio@uv.cl](mailto:felipe.osorio@uv.cl)

# Extremum estimator

## Definición 1

$\hat{\theta}$  es llamado un **extremum estimator** si existe una función objetivo  $Q_n(\theta)$  tal que:

$$Q_n(\hat{\theta}) = \max_{\theta \in \Theta} Q_n(\theta), \quad \Theta \subset \mathbb{R}^p. \quad (1)$$

### Observación:

- ▶ La notación recalca que  $Q_n(\theta)$  depende de los **datos muestrales**  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ .
- ▶ Una prueba de la existencia así como de la consistencia de  $\hat{\theta}_n$  puede ser hallada en Gourieroux y Monfort (1995).<sup>1</sup>

---

<sup>1</sup> Statistics and Econometrics Models, Volume 2. Cambridge University Press.

## Extremum estimator

### Ejemplo (Mínimos cuadrados):

Considere el siguiente modelo

$$Y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + \epsilon_i, \quad i = 1, \dots, n,$$

donde  $f$  es función conocida,  $\mathbf{x}_i$  denota un vector de regresores,  $\boldsymbol{\beta}$  es vector  $p \times 1$  de parámetros desconocidos, y  $E(\epsilon_i) = 0$ ,  $\text{var}(\epsilon_i) = \sigma^2$ ,  $\text{Cov}(\epsilon_i, \epsilon_j) = 0$  para  $i \neq j$ .

Podemos escribir,

$$\mathbf{Y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\epsilon},$$

con  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$ . De este modo,  $\hat{\boldsymbol{\beta}}_{\text{NLS}}$  es obtenido como la solución de

$$\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{f}(\boldsymbol{\beta})\|^2,$$

es decir,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n (Y_i - f_i(\boldsymbol{\beta}))^2.$$

## Extremum estimator

*Ejemplo (Función de producción):*

Considere la función de producción con elasticidad de sustitución constante (CES), dada por

$$Y_t = A_0 [\delta_0 K_t^{-\rho_0} + (1 - \delta_0) L_t^{-\rho_0}]^{-1/\rho_0} + \epsilon_t,$$

donde  $Y_t$  representa la producción en el periodo  $t$ ,  $K_t$  denota el stock de capital,  $L_t$  es el trabajo y  $\epsilon_t$  es un disturbio aleatorio con media cero.

En este caso,  $\boldsymbol{x}_t = (K_t, L_t)^\top$ ,  $\boldsymbol{\beta} = (A_0, \delta_0, \rho_0)^\top$  y

$$f(\boldsymbol{x}_t; \boldsymbol{\beta}) = A [\delta K_t^{-\rho} + (1 - \delta) L_t^{-\rho_0}]^{-1/\rho}.$$

Además las funciones de producción suelen añadir las siguientes condiciones  $A > 0$ ,  $0 < \delta < 1$  y  $-1 < \rho$ , lo que determina  $\Theta$ .

## Extremum estimator

*Ejemplo (M-estimación):*

Un extremum estimator es un M-estimador<sup>2</sup> si la función objetivo es un promedio muestral del tipo:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n m(Y_i; \boldsymbol{\theta}),$$

donde  $m$  es una función real valorada de  $(Y_i, \boldsymbol{\theta})$ .

Este procedimiento es frecuente para modelos de posición-escala, definidos por la densidad

$$\frac{1}{\sigma} f\left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\sigma}\right),$$

en cuyo caso la función  $m(\cdot)$  es escogida con el objetivo de reducir la influencia de observaciones extremas.

---

<sup>2</sup>Casos particulares: máxima verosimilitud (ML) y mínimos cuadrados no lineales (NLS)

## Extremum estimator

*Ejemplo (Estimación máximo Lq-verosímil):*

El estimador máximo Lq-verosímil (MLqE) de  $\theta$  es definido como aquél valor  $\hat{\theta}_q$  que maximiza la función (Ferrari y Yang, 2010):<sup>3</sup>

$$L_q(\boldsymbol{\theta}) = \sum_{i=1}^n l_q(f(\mathbf{y}_i; \boldsymbol{\theta})), \quad q > 0,$$

donde  $l_q(u)$  denota el logaritmo deformado de orden  $q$ , cuya definición es dada por:

$$l_q(u) = \begin{cases} (u^{1-q} - 1)/(1 - q), & q \neq 1, \\ \log(u), & q = 1, \end{cases}$$

con  $f(\mathbf{y}_i; \boldsymbol{\theta})$  la función de densidad del modelo asumido para los datos.

*Observación:*

Es fácil notar que, cuando  $q = 1$  se recupera el método de estimación ML.

---

<sup>3</sup>Annals of Statistics 38, 753-783.

## Extremum estimator

*Ejemplo (Modelo mal especificado):*

Asuma que disponemos de  $X_1, \dots, X_n$  variables IID generadas desde  $g(x; \theta)$ . Aunque los datos son modelados considerando

$$\mathcal{P}_f = \{f(x; \theta)^{\otimes n} : \theta \in \Theta\},$$

con funciones de log-verosimilitud y score asociadas

$$\ell_n(\theta) = \sum_{i=1}^n \log f(x_i; \theta), \quad \mathbf{U}_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i; \theta),$$

Sea<sup>4</sup>

$$\mathbf{V}(\theta) = \mathbb{E}_g\{\mathbf{U}_n(\theta)\mathbf{U}_n^\top(\theta)\}, \quad \mathbf{S}(\theta) = \mathbb{E}_g\left\{-\frac{\partial \mathbf{U}_n(\theta)}{\partial \theta^\top}\right\}.$$

---

<sup>4</sup>Note que, cuando  $f = g$ , tenemos  $\mathbf{V}(\theta) = \mathbf{S}(\theta) = \mathcal{F}(\theta)$ .

## Extremum estimator

### Lema 1 (Existencia)

Suponga que

- (i) El espacio paramétrico  $\Theta$  es un subconjunto compacto de  $\mathbb{R}^p$ .
- (ii)  $Q_n(\theta)$  es continua en  $\theta$  para todo  $Y_1, \dots, Y_n$ .
- (iii)  $Q_n(\theta)$  es una función medible de los datos para todo  $\theta \in \Theta$ .

Entonces existe una función medible  $\hat{\theta}_n$  que es solución del problema

$$\max_{\theta \in \Theta} Q_n(\theta), \quad \Theta \subset \mathbb{R}^p.$$

*Demostración:*

ver Gourieroux y Monfort (1995), Sección 24.1.

## Extremum estimator

Considere la transformación  $\gamma = \tau(\theta)$  definida sobre  $\Theta$  y sea

$$\Gamma \equiv \tau(\Theta) \equiv \{\gamma = \tau(\theta) \text{ para algún } \theta \in \Theta\}.$$

$\tau : \Theta \rightarrow \Gamma$  es llamada una **reparametrización** si es 1-a-1.

### Propiedad 1

Sea  $\tilde{Q}_n(\gamma)$  la función objetivo asociada con el **modelo reparametrizado**. Un extremum estimator es **invariante** si y solo si

$$\tilde{Q}_n(\gamma) = Q_n(\tau^{-1}(\gamma)),$$

para todo  $\gamma \in \Gamma$ .

*Observación:*

Existen estimadores que **no** son invariantes. (p.ej. GMM)

## Extremum estimator

### Propiedad 2 (Consistencia)

Suponga que

- (i) El parámetro verdadero  $\theta_0$  es un elemento del interior del espacio paramétrico  $\Theta$ .
- (ii)  $Q_n(\theta)$  es cóncava sobre el espacio paramétrico (para cualquier  $Y_1, \dots, Y_n$ ).
- (iii)  $Q_n(\theta)$  es una función medible para todo  $\theta \in \Theta$ .

Sea  $\hat{\theta}_n$  el extremum estimator definido por (1). Si existe una función  $Q_0(\theta)$  tal que

- (a)  $Q_0(\theta)$  es únicamente maximizada sobre  $\Theta$  en  $\theta_0 \in \Theta$ .
- (b)  $Q_n(\theta)$  converge en probabilidad a  $Q_0(\theta)$  para todo  $\theta \in \Theta$ .

Entonces,  $\hat{\theta}_n \xrightarrow{P} \theta_0$ .

### Propiedad 3

Sea  $\{\hat{\theta}_n\}$  una secuencia de extremum estimators de  $\theta$  obtenidos mediante maximizar la función objetivo  $Q_n(\theta)$ . Suponga que  $\hat{\theta}_n \xrightarrow{P} \theta_0$ , con  $\theta_0$  en el interior del espacio paramétrico  $\Theta$ . Además, asuma que  $Q_n(\cdot)$  es diferenciable con relación a  $\theta$ .

Entonces el estimador  $\hat{\theta}_n$  satisface (asintóticamente) las [condiciones de primer orden](#)

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = \mathbf{0}.$$

# Extremum estimator

## Supuestos

**B1:**  $Q_n(\theta)$  es dos veces diferenciable en  $\theta$ .

**B2:** Sea  $A_n(\theta) = \partial^2 Q_n(\theta) / \partial \theta \partial \theta^\top$  y asuma que

$$A_n(\theta) \xrightarrow{\text{a.s.}} A,$$

uniformemente.

**B3:** La matriz  $A$  es no singular.

**B4:** La secuencia  $\{\sqrt{n}(\partial Q_n(\theta)/\partial \theta)\}$  converge en distribución a una normal estándar con vector de medias cero y matriz de covarianza  $B$ . Es decir,

$$\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{\text{D}} N_p(\mathbf{0}, B).$$

## Extremum estimator

### Resultado 1 (Normalidad asintótica)

Bajo los supuestos de las Propiedades 2 y 3, y los Supuestos B1 a B4, tenemos

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N_p(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}).$$

*Esbozo de la demostración:*

Considere una función objetivo aditiva:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n Q_i(\boldsymbol{\theta}),$$

de este modo,

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial Q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Q_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.$$

## Extremum estimator

Asuma que  $\hat{\theta}_n$  satisface las condiciones de primer orden, es decir  $\partial Q_n(\hat{\theta})/\partial\theta = \mathbf{0}$ . Por el Teorema del valor medio tenemos la siguiente expansión,

$$\begin{aligned}\frac{\partial Q_n(\hat{\theta})}{\partial\theta} &= \frac{\partial Q_n(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta^\top}(\hat{\theta} - \theta_0) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial Q_i(\theta_0)}{\partial\theta} + \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Q_i(\bar{\theta})}{\partial\theta\partial\theta^\top} \right] (\hat{\theta}_n - \theta_0),\end{aligned}$$

donde  $\bar{\theta}$  es un valor medio entre  $\hat{\theta}_n$  y  $\theta_0$ . Usando la condición de primer orden

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \frac{\partial Q_i(\theta_0)}{\partial\theta} + \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Q_i(\bar{\theta})}{\partial\theta\partial\theta^\top} \right] (\hat{\theta}_n - \theta_0).$$

Asumiendo que  $\frac{1}{n} \sum_{i=1}^n \partial^2 Q_i(\bar{\theta})/\partial\theta\partial\theta^\top$  es no singular, podemos escribir

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = - \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Q_i(\bar{\theta})}{\partial\theta\partial\theta^\top} \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial Q_i(\theta_0)}{\partial\theta}.$$

## Extremum estimator

Como  $\bar{\theta}$  está entre  $\hat{\theta}_n$  y  $\theta_0$ , y siempre que  $\hat{\theta}_n$  sea un estimador consistente, entonces  $\bar{\theta} \xrightarrow{P} \theta_0$ . Además

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Q_i(\bar{\theta})}{\partial \theta \partial \theta^\top} \xrightarrow{\text{a.s.}} \mathbb{E} \left[ \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta^\top} \right] = \mathbf{A}.$$

Por otro lado,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial Q_i(\theta_0)}{\partial \theta} \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \mathbf{B}).$$

Finalmente por el Teorema de Slutsky, sigue que

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}_p(0, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}).$$