

Teoría Econométrica I - EAE- 250-A

Máxima Verosimilitud

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Noviembre/Diciembre 2021

Introducción

- En esta sección vamos a incluir un nuevo supuesto: la forma funcional de la distribución de los datos
 - Hasta ahora solo habíamos exigido $\mathbb{E}(u|x) = 0$
 - Esto nos va a permitir tener otro tipo de estimadores con mejores propiedades para varios set ups.
- Vamos a seguir las notas del libro de Paul Ruud (Capítulo 14, An Introduction to Classical Econometric Theory) y por momentos las notas de Manuel Arellano (disponibles en su página web)

Modelos de Verosimilitud

- Dada un conjunto de datos $\{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ un modelo de verosimilitud es una familia de densidades $f_n(\omega_1, \dots, \omega_n; \theta)$, con $\theta \in \Theta \subset \mathbb{R}^p$ que esta especificado como un modelo de probabilidad para los datos observados
- Si los datos son una muestra aleatoria de una población con función de densidad $g(\omega_i)$, entonces $f_n(\omega_1, \dots, \omega_n; \theta)$ es un modelo para $\prod_{i=1}^n g(\omega_i)$
- Si el modelo esta correctamente especificado, $f_n(\omega_1, \dots, \omega_n; \theta) = \prod_{i=1}^n f(\omega_i; \theta)$ y $g(\omega_i) = f(\omega_i; \theta_0)$ para algún valor de $\theta_0 \in \Theta$
- Tenemos entonces que en la función de verosimilitud (likelihood) $\mathcal{L}(\theta) = \prod_{i=1}^n f(\omega_i, \theta)$ los datos están dados y θ es el argumento.

Log likelihood y estimador por máxima verosimilitud

- Generalmente trabajaremos con el logaritmo de la verosimilitud

$$L(\theta) = \sum_{i=1}^n \ln f(\omega_i, \theta q)$$

- $L(\theta)$ mide la capacidad de las diferentes densidades (dentro de una clase pre establecida) de generar los datos
- Así, el estimador de máxima verosimilitud es aquél que maximiza la probabilidad de observar los datos que observamos dada la forma (clase) de la densidad

$$\hat{\theta}_{ML} = \underset{\theta \in \Theta}{\operatorname{Argmax}} L(\theta)$$

Likelihood function and conditional likelihood function

- Supongamos ahora que ω_i está compuesto por u_i, v_i . Puede ser que sepamos la densidad conjunta o la condicional de este par de variables aleatorias
- Definimos entonces Likelihood Function

$$l(\theta; u) \equiv f(u; \theta) \quad \text{y la } \underline{\text{log-likelihood}}$$
$$L(\theta; u) = \log l(\theta; u)$$

- y Conditional Likelihood Function:

$$l(\theta; u|v) \equiv f(u|v; \theta) \quad \theta \in \Theta$$
$$L(\theta; u|v) \equiv \log l(\theta; u|v)$$

Dominancia y Log likelihood inequality

- **Supuesto de Dominancia:** $\mathbb{E}[\sup_{\theta \in \Theta} L(\theta; u|v)]$ existe. Este es un requerimiento técnico que garantiza que $\mathbb{E}[L(\theta; u|v)]$ exista para todo $\theta \in \Theta$
- **Lema (Expected Log-likelihood Inequality):** Si $L(\theta; u|v)$ es la conditional log-likelihood function y el supuesto 14.2 se cumple, entonces:

$$\mathbb{E}[L(\theta; u|v)|v] \leq \mathbb{E}[L(\theta_0; u|v)|v]$$

- Este lema es crucial: el valor esperado de la log-likelihood alcanza un valor máximo si $\theta = \theta_0$. Esto da inicio al método de Máxima Verosimilitud.

Estimador de máxima verosimilitud

- El lema 14.1 nos sugiere un estimador natural, dado que θ_0 maximiza el valor esperado de la función l.i. Fisher (1922,1925) propone maximizar el análogo muestral de $E_N(L(\theta; u|v))$

$$E_N(L(\theta; u|v)) \equiv \sum_{n=1}^N L(\theta; u_n|v_n) \cdot \frac{1}{N}$$

donde $E_N(\cdot)$ es la esperanza o media muestral

- Definición (Estimador de Máxima Verosimilitud)**

$$\hat{\theta}_N \equiv \operatorname{argmax}_{\theta \in \Theta} E_N[L(\theta)] \quad \text{simplificando} \quad L(\theta) = L(\theta; u|v)$$

Identificación

- **Identificación** De la misma forma en que en OLS teníamos un par de supuestos fundamentales de identificación ($\mathbb{E}(X'u) = 0$ y X de rango completo) en Máxima Verosimilitud necesitamos supuestos de identificación.
- **Definición (Global Identification):** El vector de parámetro θ_0 es globalmente identificado en Θ si, $\forall \theta_1 \in \Theta, \theta_0 \neq \theta_1$

$$\Rightarrow Pr\{f(u|v; \theta_0) \neq f(u|v; \theta_1)\} > 0$$

- **Supuesto 14.3 (Identificación Global):** Todo vector de parámetros $\theta_0 \in \Theta$ está globalmente identificado.
Con el supuesto anterior tenemos que el lema 14.1 (expected I.I. inequality) se cumple con desigualdad estricta.

Diferenciabilidad

- **Supuesto 14.4 (Diferenciabilidad):** La función de probabilidad $f(u|v; \theta)$ es doblemente diferenciable respecto a $\theta \quad \forall \theta \in \Theta$. Además, el soporte $S(\theta)$ de $f(u|v; \theta)$ no depende de θ y podemos intercambiar derivadas e integrales:

$$\begin{aligned}\frac{\partial}{\partial \theta} \int_S dF(u|v; \theta) &= \int_S \frac{\partial}{\partial \theta} dF(u|v; \theta) \\ \frac{\partial^2}{\partial \theta \partial \theta'} \int_S dF(u|v; \theta) &= \int_S \frac{\partial^2}{\partial \theta \partial \theta'} dF(u|v; \theta) \\ &\quad \text{y}\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbb{E}(L(\theta)|v = \mathbf{v})}{\partial \theta} &= E \left(\frac{\partial L(\theta)}{\partial \theta} | v = \mathbf{v} \right) \\ \frac{\partial^2 \mathbb{E}(L(\theta)|v = \mathbf{v})}{\partial \theta \partial \theta'} &= E \left(\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'} | v = \mathbf{v} \right)\end{aligned}$$

- donde todos los términos existen
- La diferenciabilidad no es necesaria pero la falta de ella pone dificultades técnicas.

Score

- Recuerde que la definición de estimador de máxima verosimilitud está caracterizada por

$$\hat{\theta}_N = \operatorname{argmax}_{\theta \in \Theta} E_N[L(\theta)]$$

luego, la CPO implica

$$E_N[L_\theta(\hat{\theta}_N)] = 0 \iff \hat{\theta}_N \in \operatorname{argzero}_{\theta \in \Theta} E_N[L_\theta(\theta)]$$

- Se llaman “ecuaciones normales” o “ecuaciones de verosimilitud”.
- Definición (Score function):** La función de score se define como el vector de primeras derivadas parciales de la función log-likelihood con respecto al parámetro θ

$$L_\theta(\theta) = \frac{\partial L(\theta)}{\partial \theta}$$

- Dado que θ_0 maximiza $\mathbb{E}(L(\theta))$, esperamos que $\mathbb{E}[L_\theta(\theta_0)] = 0$

Identidad del Score

- **Lema (Identidad del Score):** Bajo los supuestos de Distribución y Diferenciabilidad,

$$\mathbb{E}(L_{\theta}(\theta_0)|\mathbf{v} = \mathbf{v}) = 0$$

- Demostración:

$$1 = \int_S dF(u|\mathbf{v}; \theta)$$

$$1 = \int_S f(u|\mathbf{v}; \theta) du \quad (\text{variable continua})$$

Tomando derivadas a ambos lados (14.1)

$$0 = \int_S \frac{\partial}{\partial \theta} f(u|\mathbf{v}; \theta) du$$

$$0 = \int_S \frac{1}{f(u|\mathbf{v}; \theta)} \cdot f_{\theta}(u|\mathbf{v}; \theta) \cdot dF(u|\mathbf{v}; \theta)$$

Identidad del Score

- Note que $L_\theta(\theta; u|v) = \frac{1}{f(u|v; \theta)} \cdot f_\theta(u|v; \theta)$ (que es la derivada de la log f)
Luego,

$$\mathbb{E}[L_\theta(\theta; u|v)|v = \mathbf{v}] = \int_S \frac{1}{f(u|v; \theta)} \cdot f_\theta(u|v; \theta) \cdot \underbrace{dF(u|v; \theta_0)}_{\text{evaluado en el verdadero parámetro}}$$

evaluando en $\theta = \theta_0$ tenemos

$$0 = \mathbb{E}[L_\theta(\theta_0)|v = \mathbf{v}]$$

donde simplificamos notación $L_\theta(\theta_0; u|v) = L_\theta(\theta)$. ■

Hessiano y las condiciones de segundo orden

- MV necesitamos encontrar una solución a las ecuaciones normales

$$E_N[L_{\theta}(\hat{\theta}_N)] = 0$$

- Pero necesitamos chequear que tenemos un máximo global, sino no es un estimador MV. Una condición suficiente es que el Hessiano sea definido negativo

$$E_N[L_{\theta\theta}(\theta)] = \frac{\partial^2 E_N[L(\theta)]}{\partial\theta\partial\theta'}$$

evaluado en $\hat{\theta}_N$. Es decir

$$c' E_N[L_{\theta\theta}(\hat{\theta}_N)] c < 0$$

- **Supuesto de Información finita:** $Var(L_{\theta}(\theta_0))$ existe.

Matriz de información e identidad de la información

- **Lema de Identidad de la Información:** Bajo los supuestos de Distribución, de Diferenciabilidad y de Información finita,

$$\mathbb{E}[L_{\theta\theta}(\theta_0|\mathbf{v} = \mathbf{v})] = -\text{Var}[L_{\theta}(\theta_0)|\mathbf{v} = \mathbf{v}]$$

y esta matriz es semi definida negativa.

Demostración

- Dados los supuestos de Distribución y Diferenciabilidad teníamos

$$0 = \int_S L_\theta(\theta; u|v) dF(u|v; \theta) \quad / \frac{\partial}{\partial \theta}$$

$$0 = \int_S [L_{\theta\theta}(\theta; u|v) dF(u|v; \theta) + L_\theta(\theta; u|v) f_\theta(u|v; \theta)' du]$$

- Note que $L_\theta(\theta; u|v) \frac{f_\theta(u|v; \theta)'}{f(u|v; \theta)} \cdot f(u|v; \theta) = L_\theta(\theta; u|v) L_\theta(\theta; u|v)' f$.

$$\implies \int_S L_{\theta\theta}(\theta; u|v) dF(u|v; \theta) = - \int_S L_\theta(\theta; u|v) L_\theta(\theta; u|v)' dF(u|v; \theta)$$

Porque $f(u|v; \theta) du = dF(u|v; \theta)$. Fijando $\theta = \theta_0$ tenemos que

$$\mathbb{E}[L_{\theta\theta}(\theta_0; u|v) | v = \mathbf{v}] = -\mathbb{E}[L_\theta(\theta_0; u|v) L_\theta(\theta_0; u|v)' | v = \mathbf{v}]$$

$$\mathbb{E}[L_{\theta\theta}(\theta_0; u|v) | v = \mathbf{v}] = -\text{Var}[L_\theta(\theta_0; u|v) | v = \mathbf{v}]$$

dado que $\mathbb{E}[L_\theta(\theta_0; u|v)] = 0$ (Identidad del Score) ■

Cuota de Cramer Rao

- **Supuesto (Non singular Info.):** $I(\theta_0)$ es no singular $\forall \theta_0$ posible en Θ
- **Teorema: (*Cramèr-Rao Inequality*):** Sea $\tilde{\theta}$ un estimador insesgado de θ_0 con matriz de varianza finita y diferenciación e integración son intercambiables (el soporte no depende de θ_0), se puede demostrar que:

$$\text{Var}(\tilde{\theta} | \nu = \nu) \geq (N E_N[I(\theta_0 | \nu)])^{-1}$$

- Esto implica el cumplimiento de los supuestos Distribución, Diferenciación, Información finita y No singularidad de la matriz de información

Propiedades Asintóticas

- Los estimadores ML son consistentes y asintóticamente normales. Vamos a demostrarlo.
- Vamos a agregar algunos supuestos más:

Supuesto 15.1 (*Compacidad*): Θ cerrado y acotado.

Supuesto 15.2 (*Interior*): $\theta_0 \in \text{int}\{\Theta\}$.

Supuesto 15.3 (*Dominancia II*): $\mathbb{E}[\sup_{\theta \in \Theta} |L_{\theta\theta}|]$ existe.

- Para demostrar consistencia note que:
 1. $E_N[L(\theta)] \xrightarrow{p} \mathbb{E}[L(\theta)]$
 2. $\hat{\theta}_N = \arg \max_{\theta \in \Theta} E_N(L(\theta)); \theta_0 = \arg \max_{\theta \in \Theta} \mathbb{E}(L(\theta))$ (por Global identification y log likelihood inequality)
 3. Por teorema de continuidad tenemos entonces: $\Rightarrow \hat{\theta}_N \xrightarrow{p} \theta_0$.

Normalidad Asintótica

- Bajo los supuestos adicionales Diferenciabilidad, Información finita, Información no singular, tenemos que aplicando TCL:

$$\left\{ -E_N[L_{\theta\theta}(\hat{\theta}_N)]^{1/2} \right\} \sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, I_k)$$

con lo cual:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

- Estos estimadores son CUAN (Consistent and Uniformly Asymptotically Normal).

Luego

$$\hat{\theta} \sim N\left(\theta_0, [N E_N(-L_{\theta\theta}(\hat{\theta}_N))]^{-1}\right)$$

donde $E_N(-L_{\theta\theta}(\hat{\theta}_N))$ es un estimador consistente de $I(\theta_0)$.

Estimación de la varianza

- Tenemos tres alternativas:

- (i) Media empírica del negativo del Hessiano

$$E_N[-L_{\theta\theta}(\hat{\theta}_N)].$$

- (ii) La varianza empírica del score

$$\text{Var}_N[L_{\theta}(\hat{\theta}_N)] = E_N[L_{\theta}(\hat{\theta}_N)L_{\theta}(\hat{\theta}_N)'].$$

- (iii) La información empírica

$$E_N[I(\hat{\theta}_N|v)]$$

Computación del estimador por Maxima verosimilitud

- Buscamos $\hat{\theta}_N$ tal que $E_N[L_{\theta}(\hat{\theta}_N)] = 0$

- **Búsqueda de Grilla**

Se busca entonces

$$\max_{\theta \in [a, b]} Q(\theta)$$

- Para conseguirlo subdividimos el intervalo $[a, b]$ en un número de intervalos

$$\{[a, \theta_1], [\theta_1, \theta_2], \dots, [\theta_n, b]\}$$

- luego evaluamos $Q(\cdot)$ en las cotas de los intervalos $(\theta_i, \quad i = 1, \dots, n)$
- se escoge el θ_i donde $Q(\cdot)$ toma el mayor valor, se escogen entonces los intervalos $[\theta_{i-1}, \theta_i]$ y $[\theta_i, \theta_{i+1}]$
- Luego se itera la realización con estos dos intervalos y así sucesivamente. Una *stopping rule* estandar puede ser
- $|Q(\theta_i) - Q(\theta_{i+1})| < 10^{-5}$, o tambien $|\theta_i - \theta_{i+1}| < 10^{-5}$.
La búsqueda de grilla funciona para θ escalar.

métodos cuadráticos

- Suponga que $Q(\cdot)$ es exactamente cuadrática

$$Q(\theta) = a + b'\theta + \frac{1}{2}\theta' C \theta$$

Note que

$$Q_{\theta}(\theta) = b + C\theta$$

$$Q_{\theta\theta}(\theta) = C'$$

$Q(\cdot)$ alcanza su máximo en θ^* y notando que $Q_{\theta}(\theta^*) = 0$ se tiene

$$\theta^* = -C^{-1}b$$

Donde asumimos que C es simétrica¹. Note que podemos caracterizar θ^* en función de Q_{θ} y $Q_{\theta\theta}$ evaluado en cualquier parámetro θ_1

$$\theta^* = -C^{-1}b$$

$$\theta^* = \theta_1 - C^{-1}(b + c\theta_1)$$

$$\theta^* = \theta_1 - Q_{\theta\theta}(\theta_1)^{-1}Q_{\theta}(\theta_1)$$

Newton Rapson

- Al setear $\delta = -Q_{\theta\theta}(\theta_1)^{-1} \cdot Q_{\theta}(\theta_1)$ y $\lambda = 1$ y con una line search alcanzamos el óptimo, sin importar el valor de partida (si es cuadrática exacta)
- Métodos Cuadráticos y MLE

Newton-Raphson: El más popular

$\delta_{NR} = \{-E_N[L_{\theta\theta}(\theta_1)]\}^{-1} E_N[L_{\theta}(\theta_1)]$ con $\lambda = 1$. De esta manera tenemos

$$\theta^{k+1} = \theta^k + \delta_{NR}$$

Si $\theta^* = \theta^{k+1}$ y $\theta_1 = \theta^k$. Luego, NR es basado en la expansión de Taylor de segundo orden de $\mathbb{E}_N[L(\theta)]$ en θ_1 .

$$\mathbb{E}_N[L(\theta)] \simeq \mathbb{E}_N[L(\theta_1)] + (\theta - \theta_1)' E_N[L_{\theta}(\theta_1)] + \frac{1}{2}(\theta - \theta_1)' E_N[L_{\theta\theta}(\theta_1)](\theta - \theta_1)$$

NLine search

- **Búsqueda de Línea**

Dado un valor inicial θ_1 y una dirección de búsqueda “ δ ” una iteración de este método busca resolver

$$\lambda^* = \arg \max_{\lambda} Q(\theta_1 + \lambda \delta)$$

- Luego $\theta_2 = \theta_1 + \lambda^* \delta$.

Hay distintos métodos de Line Search, los que difieren en cómo elegir λ y δ , donde λ es el largo del paso (*size step*) y por convención $\lambda \geq 0$.

- Dado que la derivada direccional de Q es

$$\frac{\partial Q(\theta_1 + \lambda \delta)}{\partial \lambda} = Q_{\theta}(\theta_1 + \lambda \delta)' \delta > 0$$

lo que condiciona nuestra elección de δ .

Newton Rapson

- Notando lo siguiente:

$$\begin{aligned}\frac{\partial}{\partial \theta} = 0 &\implies 0 = \mathbb{E}_N[L_{\theta}(\theta_1)] + \frac{2}{2} \mathbb{E}_N[L_{\theta\theta}(\theta_1)](\theta^* - \theta_1) \\ &\implies \theta^* = \theta_1 + \{-E_N[L_{\theta\theta}(\theta_1)]\}^{-1} E_N[L_{\theta}(\theta_1)]\end{aligned}$$

Si $\theta^* = \theta_2$ tenemos

$$\theta_2 = \theta_1 + \{-E_N[L_{\theta\theta}(\theta_1)]\}^{-1} E_N[L_{\theta}(\theta_1)]$$

Modified scoring

- **Modified Scoring:** En lugar del Hessiano ocupa la matriz de información empírica:

$$\delta_S = E_N[I(\theta_1|\nu)] \cdot E_N[L_\theta(\theta_1)]$$

- Con $\lambda = 1$. A veces el Hessiano no es semidefinido negativo y se corrige con este método porque la información es semidefinida positiva. Este hecho hace que δ_S siempre apunta en una dirección de aumento.
- Rao (1973) llamó a la iteración

$$\theta_i = \theta_{i-1} + \delta_S(\theta_{i-1}) \quad i \in \mathbb{N}$$

el **Método de Scoring**. Si tomamos lo anterior más line search tenemos el **Modified Scoring**.

Criterio de convergencia

- Un criterio muy común en estadística es un criterio sobre la variación de los parámetros, por ejemplo

$$\|\theta_i - \theta_{i-1}\| < 10^{-5}$$

- Convergencia debiese ser definida por cuán cerca de cero es el score y si el Hessiano es negativo definido
- Para asegurar convergencia a un máximo global debemos probar distintos valores iniciales
- Si la log-likelihood es globalmente cóncava, cualquier criterio basado en parámetros o función objetivo debiera cuadrar porque solo existe un máximo global.

Inferencia en Máxima Verosimilitud

- Veremos primero los tests de hipótesis clásicos
- Considere el vector $\theta = [\theta'_1, \theta'_2]$ y la nula $H_0 : \theta_2 = \theta_{02} = 0$, $\dim(\theta_2) = K - M$, $\dim(\theta_1) = M$.
- Por simplicidad usaremos $I(\theta_0)^{-1}$ como matriz de varianza asintótica (asumimos datos iid y por ende la varianza es la misma para todas las observaciones)

Test de Wald

1. Compute estimador MV no restringido $\hat{\theta} \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_N[L(\theta)]$.
2. Compute un estimador de la matriz de varianza de $\sqrt{n}(\hat{\theta} - \theta_0)$ asintótica $I(\hat{\theta}_N)^{-1}$.
3. Compute la forma cuadrática $\mathcal{W} = n \cdot \hat{\theta}'_2 \hat{V}_{\mathcal{W}}^{-1} \hat{\theta}_2$.
Donde $\hat{V}_{\mathcal{W}}$ es el elemento (2,2) o bloque de la matriz de varianzas que toma lo referente a θ_2^2 . Así,

$$\hat{V}_w = \{I_{22}(\hat{\theta}) - I_{21}(\hat{\theta})[I_{11}(\hat{\theta})]^{-1}I_{12}(\hat{\theta})\}^{-1}$$

4. Compare $\mathcal{W} \sim \chi^2_{k-m}$.

Test del Score o LM

- El test del score examina cuanto $\mathbb{E}_N[L_2(\hat{\theta}_R)]$ se desvía de vector 0.
- Bajo la nula $H_0 = \theta_{02} = 0$, $\mathbb{E}_N[L_2(\hat{\theta}_R)]$ no debiera desviarse mucho de cero.

Los pasos a seguir son:

1. Compute el estimador **restringido**

$$\hat{\theta}_R = \underset{\theta \in \Theta: \theta_2=0}{\operatorname{argmax}} \mathbb{E}_N[L(\theta)] = \begin{bmatrix} \underset{\theta_1}{\operatorname{argmax}} \mathbb{E}_N[L(\theta_1, 0)] \\ 0 \end{bmatrix}$$

y el score para los parámetros restringidos es $\mathbb{E}_N[L_2(\hat{\theta}_R)] = \mathbb{E}_N \left[\left. \frac{\partial L(\theta)}{\partial \theta_2} \right|_{\hat{\theta}_R} \right]$ (puede depender de $\hat{\theta}_1$ también).

2. Compute un estimador consistente de la varianza asintótica de $\sqrt{n}\mathbb{E}_N[L_\theta(\theta_0)]$, por ejemplo $I(\hat{\theta}_R)$.
3. Compute $S = n \cdot \mathbb{E}_N[L_2(\hat{\theta}_R)]' \hat{V}_S^{-1} \mathbb{E}_N[L_2(\hat{\theta}_R)]$ donde $V_S = I_{22}(\hat{\theta}_R) - I_{21}(\hat{\theta}_R)[I_{11}(\hat{\theta}_R)]^{-1}I_{12}(\hat{\theta}_R)$ que es la varianza condicional de $\sqrt{n}\mathbb{E}_N[L_2(\theta_0)]$ condicional en $\sqrt{n}\mathbb{E}_N[L(\theta_0)]$.
4. Compare $S \sim \chi_{k-m}^2$.

Como computar el LM test

- Hay dos maneras convencionales de escribir este test
 - i) $S = N \cdot \mathbb{E}_N[L_\theta(\hat{\theta}_R)]' [I(\hat{\theta}_R)^{-1}] \mathbb{E}_N[L_\theta(\hat{\theta}_R)]$
 - Es idéntica a la anterior porque $\mathbb{E}_N[L_1(\hat{\theta}_R)] \equiv 0$ y \hat{V}_S^{-1} es el bloque (2,2) de la matriz particionada $I(\hat{\theta}_R)^{-1}$.
 - De esta manera el test del score es el criterio de convergencia evaluado en $\hat{\theta}_R$.
 - ii) $S_{OLS} = N \cdot \mathbb{E}_N[L_\theta(\hat{\theta}_R)]' \{ \text{Var}_N[L_\theta(\hat{\theta}_R)] \}^{-1} \mathbb{E}_N[L_\theta(\hat{\theta}_R)]$ porque es la suma de los residuos al cuadrado de la regresión de un vector de 1's en las columnas de la matriz $L_\theta(\hat{\theta}_R; U_n)$.

Como computar el LM test

- De acuerdo a Ruud, si denotamos la matriz de $N \times K$
- de derivadas (evaluadas en todas las observaciones, por eso no me cuadraba...) con

$$\hat{G} = [L_{\theta}(\hat{\theta}_R; U_n)]'$$

- tenemos que

$$\mathbb{E}_N[L_{\theta}(\hat{\theta}_R)] = N^{-1} L_{\theta}(\hat{\theta}_R; u_n)' \cdot 1$$

y

$$\text{Var}_N[L_{\theta}(\hat{\theta}_R; U_n)] = N^{-1} \hat{G}' \hat{G}$$

luego

$$S_{OLS} = 1' \hat{G} [\hat{G}' \hat{G}]^{-1} \hat{G}' 1$$

Test de Razón de Verosimilitud

- **Test de Razón de Verosimilitud (LR)**

$$LR = 2 \cdot N\{\mathbb{E}_N[L(\hat{\theta})] - \mathbb{E}_N[L(\hat{\theta}_R)]\} \sim \chi^2_{k-m}$$

- Es importante notar que hay que calcular los dos modelos.
- En modelos lineales se puede probar que :

$$W \geq LR \geq LM$$