Portfolios robustos basados en máxima Lq-verosimilitud penalizada

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Laboratorio de Modelación I

Agosto 26, 2022

Estimación máximo Lq-verosímil

El estimador máximo Lq-verosímil (MLqE) de $\boldsymbol{\theta}$ es definido como aquél valor $\widehat{\boldsymbol{\theta}}_q$ que maximiza la función (Ferrari y Yang, 2010)[1]:

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

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

Observación:

Si $q \to 1$, entonces $l_q(u) \to \log(u)$ y recuperamos el método de estimación ML.

[1] The Annals of Statistics **38**, 753-783.

Estimación máximo Lq-verosímil

El MLqE $\widehat{\boldsymbol{\theta}}_q$ puede ser calculado resolviendo las ecuaciones de Lq-verosimilitud:

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} l_q(f(\boldsymbol{y}_i;\boldsymbol{\theta})) = \mathbf{0}.$$

Es decir, $\boldsymbol{\Psi}(\boldsymbol{\theta})$ puede ser escrito como una función de estimación asociada a la solución de una verosimilitud ponderada

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} U_i(\boldsymbol{\theta})\boldsymbol{s}_i(\boldsymbol{\theta}),$$

donde $U_i(\boldsymbol{\theta}) = \{f(\boldsymbol{y}_i;\boldsymbol{\theta})\}^{1-q}$ y $\boldsymbol{s}_i(\boldsymbol{\theta}) = \partial \log f(\boldsymbol{y}_i;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, corresponde a la función score para la i-ésima observación.

Observación:

Cuando q es una constante fijada, $\widehat{\boldsymbol{\theta}}_q$ pertenece a la clase de los M-estimadores.

EX UMBRA IN SOLEM

Portfolios basados en estimación MLq penalizada (Giuzio et al., 2016)[2]

Considere $\boldsymbol{X} = (X_1, \dots, X_p)^\top$ vector denotando las rentabilidades de p activos medidos para un periodo de tiempo fijado, y suponga que siguen una distribución multivariada.

Podemos definir la rentabilidad de un portfolio mediante la combinación lineal $\boldsymbol{X}^\top \boldsymbol{\beta}$, siendo $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$ un vector que representa como se realiza la distribución de los fondos entre los p activos.

Objetivo:

Deseamos escoger $\boldsymbol{\beta}$ tal que $\boldsymbol{X}^\top \boldsymbol{\beta}$ sea lo más cercano posible a un índice de rentabilidad de mercado, Y.

[2] European Journal of Operational Research **250**, 251-261.

Portfolios basados en estimación MLq penalizada

Suponga $(Y_1, \boldsymbol{x}_1), \dots, (Y_n, \boldsymbol{x}_n)$ pares de observaciones medidas en los tiempos $i = 1, \dots, n$. Se desea calcular $\widehat{\boldsymbol{\beta}}_{q,\lambda}$ maximizando la función objetivo:

$$D_{q,\lambda}(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_q \Big\{ f\Big(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\sigma} \Big) \Big\} + \sum_{j=1}^{p} l_q \{ h(\beta_j, \lambda) \},$$

donde $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma)^\top$, con $\lambda > 0$ y $q \leq 1$ constantes de regularización y distorsión fijas, mientras que $f(\cdot)$ es función de densidad en la familia de posición y escala y $h(\cdot,\cdot)$ representa una densidad simétrica con posición cero y escala λ.

Ejemplos de $h(\beta, \lambda)$:

Podemos considerar las siguientes densidades para la función de penalización,

- Normal:
$$h(\beta, \lambda) = (2\pi/\lambda)^{-1/2} \exp(-\lambda \beta^2).$$
- Laplace:
$$h(\beta, \lambda) = \frac{\lambda}{2} \exp(-\lambda |\beta|).$$

La ecuación de estimación para $\boldsymbol{\beta}$, digamos $\partial D_{q,\lambda}(\boldsymbol{\theta})/\partial\boldsymbol{\beta} = \mathbf{0}$, adopta la forma

$$\sum_{i=1}^{n} U_i(\boldsymbol{\theta}) \frac{\partial}{\partial\boldsymbol{\beta}} \log f\Big(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{\sigma}\Big) + \dot{\boldsymbol{h}}_\lambda(\boldsymbol{\beta}) = \mathbf{0},$$

donde $\dot{\boldsymbol{h}}_\lambda(\boldsymbol{\beta})$ es un vector de derivadas cuyo j-ésimo elemento es dado por

$$W_j(\beta_j) \frac{\partial}{\partial\beta_j} \log h(\beta_j, \lambda), \qquad j = 1, \dots, p,$$

y

$$U_i(\boldsymbol{\theta}) = \Big\{f\Big(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}}{\sigma}\Big)\Big\}^{1-q}, \qquad W_i(\beta_j) = \{h(\beta_j, \lambda)\}^{1-q}.$$

Observación:

Note que los "pesos" $U_i(\boldsymbol{\theta})$ son dependientes de los datos y serán pequeños para observaciones atípicas, mientras que $W_j(\beta_j)$ será pequeño si $|\beta_j|$ es "grande".

Algoritmo 1: Estimación doblemente re-ponderada (2RE)

Entrada : Datos $\boldsymbol{X}$ y $\boldsymbol{y}$
Parámetros: Constantes $q \leq 1$ y $\lambda > 0$.

begin
  Hacer $s = 0$, y calcular estimaciones iniciales $\boldsymbol{\beta}^{(s)}$, $\sigma^{(s)}$
  repeat
    Calcular $\widehat{U}_i(\boldsymbol{\theta}^{(s)})$ y $\widehat{W}_j(\beta_j^{(s)})$ para $i = 1, \dots, n$ y $j = 1, \dots, p$
    Obtener $\widetilde{\boldsymbol{\beta}}$ y $\widetilde{\sigma}$, maximizando:

$$\sum_{i=1}^{n} \widehat{U}_i \log f\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) + \sum_{j=1}^{p} \widehat{W}_i \log h(\beta_j, \lambda).$$

    Actualizar $\boldsymbol{\beta}^{(s)}$, $\sigma^{(s)}$ resolviendo la ecuación:

$$f_q(z) = f_q(\widehat{z}^{(s)}), \qquad z = (Y - \boldsymbol{x}^\top \boldsymbol{\beta})/\sigma,$$

    donde $f_q(z) = \{f(z)\}^q / \int \{f(z)\}^q \,\mathrm{d}\, z$.
  until *alcanzar convergencia*
  return $\widehat{\boldsymbol{\beta}}_{q,\lambda} = \widehat{\boldsymbol{\beta}}^{(s)}$, $\widehat{\sigma}_{q,\lambda} = \sigma^{(s)}$.
end

Estimación máximo Lq-verosímil

Considere la transformación

$$f_q(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\{f(\boldsymbol{x};\boldsymbol{\theta})\}^q}{\int_{\mathcal{X}} \{f(\boldsymbol{x};\boldsymbol{\theta})\}^q \, \mathsf{d}\boldsymbol{x}}.$$

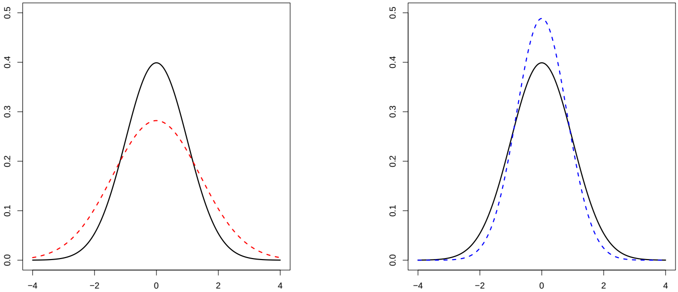
La función $f_q(\boldsymbol{x};\boldsymbol{\theta})$ es conocida como densidad distorsionada (escort). El párametro q provee un mecanismo para acentuar diferentes regiones de la densidad verdadera no transformada $f(\boldsymbol{x};\boldsymbol{\theta})$.

Observación:

En efecto, para:

- $q > 1$, regiones con valores cuya densidad son cercanos a 0 son acentuados.
- $q < 1$, regiones con valores lejanos desde 0 son enfatizados, es decir, el rol de valores atípicos es reducido.

Ejemplo: Distribución normal estándar N$(0, 1)$



(a) $q < 1$ (b) $q > 1$

Ejemplo: Distribución normal estándar $\mathsf{N}(0,1)$

Para la distribución $\mathsf{N}(0,1)$, tenemos[3]

$$\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2), \qquad \phi_q(z) = (2\pi/q)^{-1/2}\exp(-qz^2/2).$$

El paso 6 del Algoritmo 1 para la distribución $\mathsf{N}(0,1)$, lleva a la actualización

$$\boldsymbol{\beta}^{(s)} = \widetilde{\boldsymbol{\beta}}, \qquad \sigma^{(s)} = \widetilde{\sigma}/\sqrt{q}.$$

Adicionalmente, para asegurar la convergencia del Algoritmo 1 es requerido que $f_q(z)$ pertenezca a la misma familia que $f(z)$.

La motivación del paso 6 del Algoritmo 1 proviene de que, en general, el método de máxima Lq-verosimilitud, no satisface[4]

$$\mathsf{E}_\theta\{\boldsymbol{\Psi}_n(\boldsymbol{\theta})\} = \mathbf{0}.$$

La estrategia en el paso 6 es sólo una manera de resolver este problema.

[3] $\phi_q(z)$ es la densidad de una variable aleatoria $\mathsf{N}(0, q^{-1})$.

[4] Es decir, resolver $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \mathbf{0}$ no lleva a estimadores Fisher-consistentes.

Portfolios basados en estimación MLq penalizada

En Giuzio et al. (2016) se realiza una presentación detallada de:

- Algoritmo de estimación basado en portfolios normales.

- Algoritmo EM (Qin y Priebe, 2013)[5] para portfolios t-Student basados en la propiedad
$$Y_i|\tau_i \sim \mathsf{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2/\tau_i), \qquad \tau_i \sim \mathsf{Gamma}\Big(\frac{\nu}{2}, \frac{\nu}{2}\Big),$$
para $i = 1, \dots, n$.

- Para la selección del parámetro de regularización, se considera
$$\mathrm{SIC}_q = -2 \sum_{i=1}^{n} l_q \Big\{ f\Big(\frac{y_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}_{q,\lambda}}{\widehat{\sigma}_{q,\lambda}} \Big) \Big\} + k \log n,$$
donde $k \leq n$ denota el número de portfolios activos (es decir, el número de β's distintos de cero). Además $\mathrm{SIC}_q \to \mathrm{SIC}$ conforme $q \to 1$.

[5] Journal of the American Statistical Association **108**, 914-928

Propuesta de trabajo

- Se desea implementar:
  - **a)** Estimación MLq-verosímil para portfolios normales y t-Student.
  - **b)** Selección del parámetro de regulación usando criterios de información robustos (Ronchetti, 1997)[6] o validación cruzada generalizada ponderada (O'Sullivan et al., 1986)[7].
- Aplicar la métodología de Giuzio et al. (2016) a datos del mercado de AFP's chileno.
- Datos disponibles en GitHub: `https://github.com/faosorios/AFP`

[6] Statistica Sinica **7**, 327-338

[7] Journal of the American Statistical Association **81**, 96-103

Referencias bibliográficas

Giuzio, M., Ferrari, D., Paterlini, S. (2016).
Sparse and robust normal and t- portfolios by penalized Lq-likelihood.
European Journal of Operational Research **250**, 251-261.

Ferrari, D., Paterlini, S. (2009).
The maximum Lq-likelihood method: An application to extreme quantile estimation in finance.
Methodology and Computing in Applied Probability **11**, 3-19.

Ferrari, D., Yang, Y. (2010).
Maximum Lq-likelihood estimation.
The Annals of Statistics **38**, 753-783.

Qin, Y., Priebe, C.E. (2013).
Maximum Lq-likelihood estimation via the Expectation-Maximization algorithm: A robust estimation of mixture models.
Journal of the American Statistical Association **108**, 914-928.

Ronchetti, E. (1997).
Robustness aspects of model choice.
Statistica Sinica **7**, 327-338.