

Comparison of the Bootstrapping Method and Generalized Bootstrapping Method

Bin WANG

Mathematics and Statistics Department
University of South Alabama, Mobile, AL

December 6, 2006

Part II

Fitting Data and Distribution with GLD/EGLD

The GLD Moments

The k -th moment of a $GLD(0, \lambda_2, \lambda_3, \lambda_4)$ random variable Z is given by

$$E(Z^k) = \frac{1}{\lambda_2^k} \sum_{i=0}^k \left[\binom{k}{i} (-1)^i \beta(\lambda_3(k-i) + 1, \lambda_4 i + 1) \right] \quad (7)$$

where $\beta(a, b)$ is the beta function defined by

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx.$$

and the k -th moment exists if and only if $\lambda_3 > -1/k$ and $\lambda_4 > -1/k$.

The GLD first four Moments

For a $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ r.v. X , its first four moments, $\alpha_1, \alpha_2, \alpha_3$ and α_4 , are given by

$$\alpha_1 = \mu = E(X) = \lambda_1 + \frac{A}{\lambda_2}, \quad (8)$$

$$\alpha_2 = \sigma^2 = E(X - \mu)^2 = \frac{B - A^2}{\lambda_2^2}, \quad (9)$$

$$\alpha_3 = E(X - \mu)^3 / \sigma^3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}, \quad (10)$$

$$\alpha_4 = E(X - \mu)^4 / \sigma^4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4}, \quad (11)$$

Note that $\lambda^2 \sigma^2$ in (10) and (11) can be replaced by $B - A^2$.

The GLD first four Moments

where

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}, \quad (12)$$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\beta(1 + \lambda_3, 1 + \lambda_4), \quad (13)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\beta(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta(1 + \lambda_3, 1 + 2\lambda_4), \quad (14)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\beta(1 + 3\lambda_3, 1 + \lambda_4) + 6\beta(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\beta(1 + \lambda_3, 1 + 3\lambda_4). \quad (15)$$

Note that A, B, C, D are free of λ_1 and λ_2 .

Multivariate Newton's Method

Rewrite the subsystem and we have

$$f_1(\lambda_3, \lambda_4) = \alpha_3 - \hat{\alpha}_3 = 0 \quad (19)$$

$$f_2(\lambda_3, \lambda_4) = \alpha_4 - \hat{\alpha}_4 = 0. \quad (20)$$

Define the vector-valued functions $F(\lambda_3, \lambda_4) = (f_1, f_2)$. Thus $F(\lambda) = 0$, where $\lambda = (\lambda_3, \lambda_4)$. The **Jacobian matrix** is defined by

$$DF(\lambda) = \begin{bmatrix} \frac{\partial f_1}{\partial \lambda_3} & \frac{\partial f_1}{\partial \lambda_4} \\ \frac{\partial f_2}{\partial \lambda_3} & \frac{\partial f_2}{\partial \lambda_4} \end{bmatrix}$$

Multivariate Newton's Method

The Taylor expansion for vector-valued functions around λ_0 is

$$F(\lambda) = F(\lambda_0) + DF(\lambda_0) \cdot (\lambda - \lambda_0) + O(\lambda - \lambda_0)^2. \quad (21)$$

By ignoring the $O(h^2)$ terms, we can solve (21) for λ ,

$$\lambda \approx \lambda_0 - DF^{-1}(\lambda_0) \cdot F(\lambda_0). \quad (22)$$

Multivariate Newton's Method

Iteration Step:

$$\lambda_3^{k+1} = \lambda_3^k - \frac{f_1 \cdot \frac{\partial f_2}{\partial \lambda_4} - f_2 \cdot \frac{\partial f_1}{\partial \lambda_4}}{\frac{\partial f_1}{\partial \lambda_3} \cdot \frac{\partial f_2}{\partial \lambda_4} - \frac{\partial f_1}{\partial \lambda_4} \cdot \frac{\partial f_2}{\partial \lambda_3}} \quad \lambda_3 = \lambda_3^k, \lambda_4 = \lambda_4^k \quad (23)$$

$$\lambda_4^{k+1} = \lambda_4^k - \frac{f_1 \cdot \frac{\partial f_2}{\partial \lambda_3} - f_2 \cdot \frac{\partial f_1}{\partial \lambda_3}}{\frac{\partial f_1}{\partial \lambda_4} \cdot \frac{\partial f_2}{\partial \lambda_3} - \frac{\partial f_1}{\partial \lambda_3} \cdot \frac{\partial f_2}{\partial \lambda_4}} \quad \lambda_3 = \lambda_3^k, \lambda_4 = \lambda_4^k \quad (24)$$

Multivariate Newton's Method

where

$$\frac{\partial f_1}{\partial \lambda_3} = \frac{(C'_3 - 3A'_3 B - 3AB'_3 + 6A^2 A'_3)(B - A^2)}{(B - A^2)^{5/2}}$$

$$= \frac{-3/2(C - 3AB + 2A^3)(B'_3 - 2AA'_3)}{(B - A^2)^{5/2}}$$

$$\frac{\partial f_1}{\partial \lambda_4} = \frac{(C'_4 - 3A'_4 B - 3AB'_4 + 6A^2 A'_4)(B - A^2)}{(B - A^2)^{5/2}}$$

$$= \frac{-3/2(C - 3AB + 2A^3)(B'_4 - 2AA'_4)}{(B - A^2)^{5/2}}$$

Multivariate Newton's Method

and

$$\frac{\partial f_2}{\partial \lambda_3} = \frac{(D'_3 - 4AC'_3 - 4A'_3C + 12AA'_3B + 12A^2B'_3 - 12A^3A'_3)(B - A^2)}{(B - A^2)^3}$$

$$= \frac{2(D - 4AC + 6A^2B - 3A^4)(B - 2AA'_3)}{(B - A^2)^3}$$

$$\frac{\partial f_2}{\partial \lambda_4} = \frac{(D'_4 - 4AC'_4 - 4A'_4C + 12AA'_4B + 12A^2B'_4 - 12A^3A'_4)(B - A^2)}{(B - A^2)^3}$$

$$= \frac{2(D - 4AC + 6A^2B - 3A^4)(B - 2AA'_4)}{(B - A^2)^3}$$

Multivariate Newton's Method

where

$$A'_3 = \frac{\partial A}{\partial \lambda_3} = -\frac{1}{(1 + \lambda_3)^2}$$

$$A'_4 = \frac{1}{(1 + \lambda_4)^2}$$

$$B'_3 = -\frac{1}{(1 + 2\lambda_3)^2} - 2\beta'_3(1 + \lambda_3, 1 + \lambda_4)$$

$$B'_4 = -\frac{1}{(1 + 2\lambda_4)^2} - 2\beta'_4(1 + \lambda_3, 1 + \lambda_4)$$

$$C'_3 = -\frac{1}{(1 + 3\lambda_3)^2} - 3\beta'_3(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta'_3(1 + \lambda_3, 1 + 2\lambda_4)$$

$$C'_4 = \frac{1}{(1 + 3\lambda_4)^2} - 3\beta'_4(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta'_4(1 + \lambda_3, 1 + 2\lambda_4)$$

Multivariate Newton's Method

$$D'_3 = -\frac{1}{(1+4\lambda_3)^2} - 4\beta'_3(1+3\lambda_3, 1+\lambda_4) + 6\beta'_3(1+2\lambda_3, 1+2\lambda_4) - 4\beta'_3(1+\lambda_3, 1+3\lambda_4)$$

$$D'_4 = -\frac{1}{(1+4\lambda_4)^2} - 4\beta'_4(1+3\lambda_3, 1+\lambda_4) + 6\beta'_4(1+2\lambda_3, 1+2\lambda_4) - 4\beta'_4(1+\lambda_3, 1+3\lambda_4)$$

where

$$\beta'_3(a+b\lambda_3, c+d\lambda_4) = \int_0^1 \log(x) \cdot x^{a-1+b\lambda_3}(1-x)^{c-1+d\lambda_4} dx$$

$$\beta'_4(a+b\lambda_3, c+d\lambda_4) = \int_0^1 \log(1-x) \cdot x^{a-1+b\lambda_3}(1-x)^{c-1+d\lambda_4} dx$$

Integration method: Legendre-Gaussian Quadrature

In R, we can use the built-in beta function to compute $\beta(a, b)$:
 double beta(*double a, double b*)
 $\beta'_3(a+b\lambda_3, c+d\lambda_4)$ and $\beta'_4(a+b\lambda_3, c+d\lambda_4)$ can be computed
 by using the **Legendre-Gaussian Quadrature**:

$$\int_0^1 g(x) = \frac{1}{2} \int_{-1}^1 g\left(\frac{t+1}{2}\right) dt, \quad t = 2x - 1$$

$$= \sum_{i=1}^n c_i g(x_i),$$

where c_i and roots x_i are listed below,

Legendre-Gaussian Quadrature Coefficients

n	roots x_i	coefficients c_i
2	$\pm\sqrt{1/3}$	1
3	$\pm\sqrt{3/5}$ 0	5/9 8/9
4	$\pm\sqrt{\frac{15+2\sqrt{30}}{35}}$ $\pm\sqrt{\frac{15-2\sqrt{30}}{35}}$	$\frac{18-\sqrt{30}}{36}$ $\frac{18+\sqrt{30}}{36}$
5	0 $\pm\frac{1}{21}\sqrt{245-14\sqrt{70}}$ $\pm\frac{1}{21}\sqrt{245+14\sqrt{70}}$	128/225 $\frac{1}{900}(322+13\sqrt{70})$ $\frac{1}{900}(322-13\sqrt{70})$

An Alternative Way

The algorithm can be simplified by approximate the partial derivatives in the Jacobian matrix by three-point center-difference approximation

$$\frac{\partial f_j}{\partial \lambda_3} = \frac{f_j(\lambda_3 + h, \lambda_4) - f_j(\lambda_3 - h, \lambda_4)}{2h} - \frac{h^2}{6} f_j'''(c_3)$$

$$\frac{\partial f_j}{\partial \lambda_4} = \frac{f_j(\lambda_3, \lambda_4 + h) - f_j(\lambda_3, \lambda_4 - h)}{2h} - \frac{h^2}{6} f_j'''(c_4)$$

where $c_k \in (\lambda_k - h, \lambda_k + h)$, $k = 3, 4$.

Limitations of the MoM

- Analyses of actual data shows that it is most common for data to produce (α_3^2, α_4) with

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15,$$

for $-4 \leq \hat{\alpha}_3 \leq 4$ and $1 < \hat{\alpha}_4 \leq 50$.

- The (α_3^2, α_4) -space confined by

$$1.8(\hat{\alpha}_3^2 + 1) \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15$$

is attainable by the GLD.

- The (α_3^2, α_4) -space not attainable by the GLD:

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8(\hat{\alpha}_3^2 + 1).$$

- Impossible region for all distributions:

$$\hat{\alpha}_4 < \hat{\alpha}_3^2 + 1.$$

Limitations of the MoM

- Analyses of actual data shows that it is most common for data to produce (α_3^2, α_4) with

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15,$$

for $-4 \leq \hat{\alpha}_3 \leq 4$ and $1 < \hat{\alpha}_4 \leq 50$.

- The (α_3^2, α_4) -space confined by

$$1.8(\hat{\alpha}_3^2 + 1) \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15$$

is attainable by the GLD.

- The (α_3^2, α_4) -space not attainable by the GLD:

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8(\hat{\alpha}_3^2 + 1).$$

- Impossible region for all distributions:

$$\hat{\alpha}_4 < \hat{\alpha}_3^2 + 1.$$

Limitations of the MoM

- Analyses of actual data shows that it is most common for data to produce (α_3^2, α_4) with

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15,$$

for $-4 \leq \hat{\alpha}_3 \leq 4$ and $1 < \hat{\alpha}_4 \leq 50$.

- The (α_3^2, α_4) -space confined by
$$1.8(\hat{\alpha}_3^2 + 1) \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15$$
is attainable by the GLD.
- The (α_3^2, α_4) -space not attainable by the GLD:
$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8(\hat{\alpha}_3^2 + 1).$$

- Impossible region for all distributions:

$$\hat{\alpha}_4 < \hat{\alpha}_3^2 + 1.$$

Limitations of the MoM

- Analyses of actual data shows that it is most common for data to produce (α_3^2, α_4) with

$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15,$$

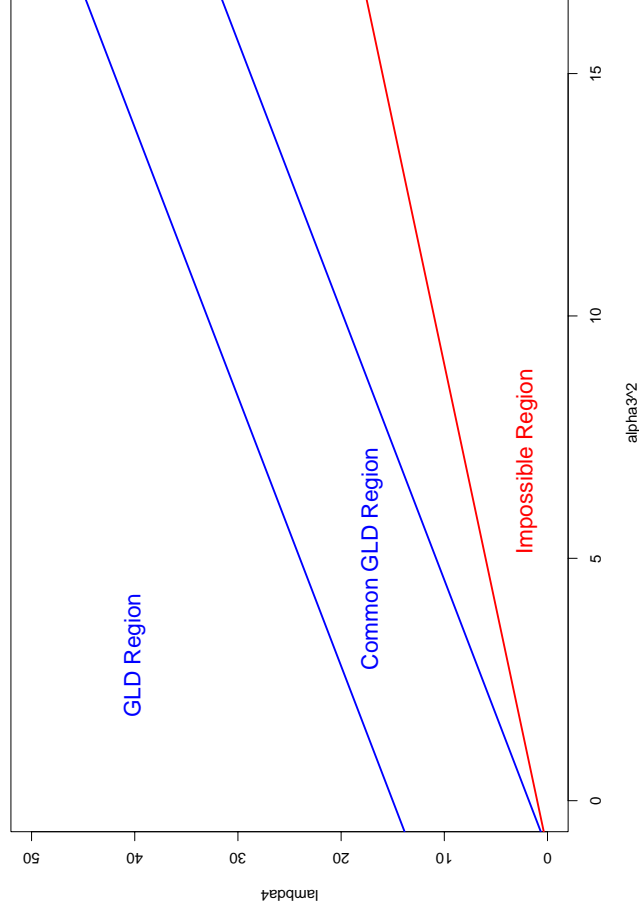
for $-4 \leq \hat{\alpha}_3 \leq 4$ and $1 < \hat{\alpha}_4 \leq 50$.

- The (α_3^2, α_4) -space confined by
$$1.8(\hat{\alpha}_3^2 + 1) \leq \hat{\alpha}_4 \leq 1.8\hat{\alpha}_3^2 + 15$$
is attainable by the GLD.
- The (α_3^2, α_4) -space not attainable by the GLD:
$$\hat{\alpha}_3^2 + 1 \leq \hat{\alpha}_4 \leq 1.8(\hat{\alpha}_3^2 + 1).$$

- Impossible region for all distributions:

$$\hat{\alpha}_4 < \hat{\alpha}_3^2 + 1.$$

(α_3^2, α_4) -space



The Generalized β -distribution/EGLD

The EGLD region is not attainable by the method of moments and thus we consider fitting a **Generalized Beta Distribution (GBD)** or **Extended Generalized Lambda Distribution (EGLD)** in that region.

The beta distribution with parameters $\beta_3 > -1$ and $\beta_4 > -1$ is defined through its p.d.f. by

$$f(x) = \begin{cases} \frac{x^{\beta_3}(1-x)^{\beta_4}}{\beta(\beta_3+1, \beta_4+1)}, & \text{for } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

where $\beta(a, b)$ is the beta function.

The Generalized β -distribution/EGLD

If $X \sim \text{Beta}()$ and $Y = \beta_1 + \beta_2 X$,

$$Y \sim \text{GBD}(\beta_1, \beta_2, \beta_3, \beta_4).$$

The p.d.f. of a $\text{GBD}(\beta_1, \beta_2, \beta_3, \beta_4)$ r.v. is

$$f(x) = \begin{cases} \frac{(x-\beta_1)^{\beta_3}(\beta_1+\beta_2-x)^{\beta_4}}{\beta(\beta_3+1, \beta_4+1)\beta_2^{\beta_3+\beta_4+1}}, & \text{for } \beta_1 \leq x \leq \beta_1 + \beta_2 \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

The EGLD Moments

The first four moments of a $\text{GBD}(\beta_1, \beta_2, \beta_3, \beta_4)$ r.v. are

$$\alpha_1 = \mu'_Y = \beta_1 + \frac{\beta_2(\beta_3 + 1)}{B_2}, \quad (27)$$

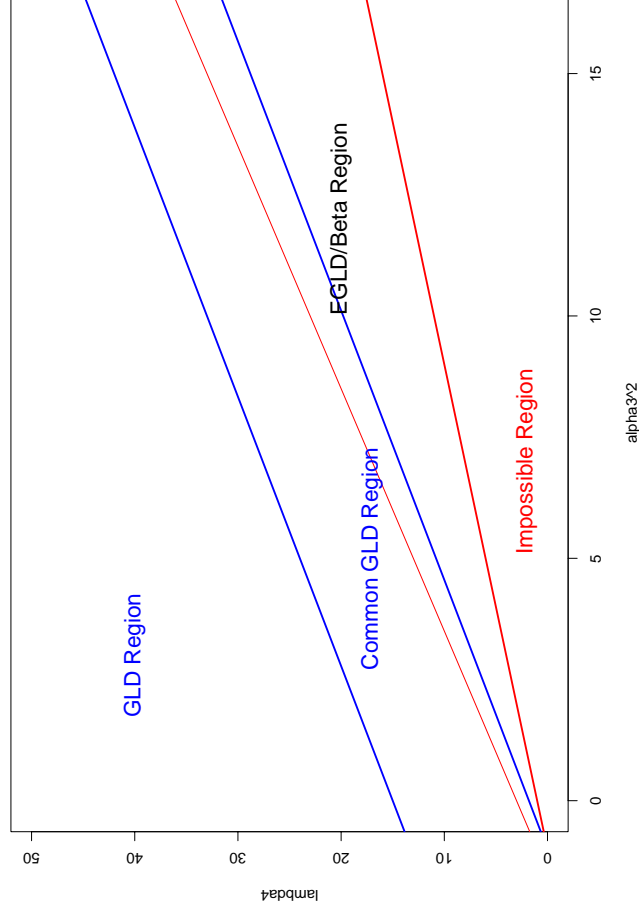
$$\alpha_2 = \mu_2 = \sigma_Y^2 = \frac{\beta_2^2(\beta_3 + 1)(\beta_4 + 1)}{B_2^2 B_3}, \quad (28)$$

$$\alpha_3 = \frac{\mu_3}{\sigma_Y^3} = \frac{2(\beta_4 - \beta_3)\sqrt{B_3}}{B_4\sqrt{(\beta_3 + 1)(\beta_4 + 1)}}, \quad (29)$$

$$\alpha_4 = \frac{\mu_4}{\sigma_Y^4} = \frac{3B_3(\beta_3\beta_4 B_2 + 3\beta_3^2 + 5\beta_3 + 3\beta_4^2 + 5\beta_4 + 4)}{B_4 B_5(\beta_3 + 1)(\beta_4 + 1)} \quad (30)$$

where $B_i = \beta_3 + \beta_4 + i$ for $i = 1, 2, \dots$.

(α_3^2, α_4) -space



Estimation of GBD Parameters via the MoM

By equating the first four moments above to the first four sample moments, we can estimate the parameters. We can find that the 3rd and 4th moments are free of β_1 and β_2 . So, we can solve a subsystem

$$\alpha_i = \hat{\alpha}_i, \quad i = 3, 4,$$

for β_3 and β_4 first. Then use the first two moments to estimate β_2 and β_1 .

Numerical solution will be found by using the Newton's method due to the complexity of the subsystem.

Estimation via the Percentiles

Certain moments of some distributions do not exist. The $(100p)$ th percentile of a sample X_1, \dots, X_n , $\hat{\pi}_p$, can be obtained from

$$\hat{\pi}_p = X_{(r)} + \frac{a}{b}(X_{(r+1)} - X_{(r)}), \quad (31)$$

where r is the integer part of $(n + 1)p$ and a/b is the proper fraction, possibly zero. For $u \in [0, 1/4]$, the sample statistics that we will use are defined by

$$\hat{\rho}_1 = \hat{\pi}_{0.5}, \quad (32)$$

$$\hat{\rho}_2 = \hat{\pi}_{1-u} - \hat{\pi}_u, \quad (33)$$

$$\hat{\rho}_3 = \frac{\hat{\pi}_{0.5} - \hat{\pi}_u}{\hat{\pi}_{1-u} - \hat{\pi}_{0.5}}, \quad (34)$$

$$\hat{\rho}_4 = \frac{\hat{\pi}_{0.75} - \hat{\pi}_{0.25}}{\hat{\rho}_2}, \quad (35)$$



Estimation via the Percentiles

The $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ counterparts are defined as

$$\rho_1 = \lambda_1 + \frac{(1/2)^{\lambda_3} - (1/2)^{\lambda_4}}{\lambda_2}, \quad (36)$$

$$\rho_2 = \frac{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}{\lambda_2}, \quad (37)$$

$$\rho_3 = \frac{(1-u)^{\lambda_4} - u^{\lambda_3} + (1/2)^{\lambda_3} - (1/2)^{\lambda_4}}{(1-u)^{\lambda_3} - u^{\lambda_4} + (1/2)^{\lambda_4} - (1/2)^{\lambda_3}}, \quad (38)$$

$$\rho_4 = \frac{(3/4)^{\lambda_3} - (1/4)^{\lambda_4} + (3/4)^{\lambda_4} - (1/4)^{\lambda_3}}{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}. \quad (39)$$



Fitting the GLD via the L-moments

It is easy to show that Λ_3/Λ_2 and λ_4/Λ_2 are free of λ_1 and λ_2 . So we can solve a subsystem

$$\frac{\lambda_3}{\Lambda_2} = \frac{\ell_3}{\ell_2} \quad \text{and} \quad \frac{\lambda_4}{\Lambda_2} = \frac{\ell_4}{\ell_2}$$

for λ_3 and λ_4 . Then estimate λ_1 and λ_2 through the first two L-moments.