Generalized method of moments for an extended Gamma process

Zeina Al Masry, Sophie Mercier, Ghislain Verdier

Laboratoire de Mathématiques et de leurs Applications Pau
UMR CNRS 5142
Université de Pau et des Pays de l'Adour, Pau, France

Fourth AMMSI Project Workshop
20 January 2016, Grenoble

Outline

Standard Gamma process(SGP)

Definition

Let $\boldsymbol{Y} = (Y_t)_{t\geq 0}$ be a stochastic process, $A(t)$ an increasing continuous function and let $b_0 > 0$. $\boldsymbol{Y}$ is said to be a SGP ($\boldsymbol{Y} \sim \Gamma_0(A(t), b_0)$) if

- $Y_0 = 0$,
- for $0 \leq t_1 < ... < t_n$, $Y_{t_1}, Y_{t_2} - Y_{t_1}, ..., Y_{t_n} - Y_{t_{n-1}}$ are independents,
- for all $s < t$, $Y_t - Y_s \sim \Gamma_0(A(t) - A(s), b_0)$.

The pdf at time t is given by

$$f_t(x) = \frac{b_0^{A(t)}}{\Gamma(A(t))} x^{A(t)-1} \exp(-b_0 x), \forall x \geq 0.$$

▲ A SGP is not always a proper choice to model the evolution of the cumulative deterioration of a system over time.

Extended Gamma process(EGP)

Definition [Cinlar 1980]

Let $\boldsymbol{X} = (X_t)_{t\geq 0}$ be a stochastic process, $A(t)$ an increasing continuous function and $b(t)$ a measurable positive function such that $\int_0^t \frac{1}{b(s)} a(s)ds < \infty$, for all $t > 0$, with $a(t)$ the derivative of $A(t)$.

$\boldsymbol{X}$ is an EGP ($\boldsymbol{X} \sim \Gamma(A(t), b(t))$) if:

$$X_t = \int_0^t \frac{dY_s}{b(s)} \text{ with } \boldsymbol{Y} \sim \Gamma_0(A(t), 1).$$

- The increments are independent,
- for all $t, \lambda \geq 0, h > 0$,
  $\mathcal{L}_{X_{t+h}-X_t}(\lambda) = \exp\left(-\int_t^{t+h} \ln\left(1 + \frac{\lambda}{b(s)}\right) a(s)ds\right),$
- $\mathbb{E}(X_t) = \int_0^t \frac{a(s)ds}{b(s)}$ and $\mathbb{V}(X_t) = \int_0^t \frac{a(s)ds}{b(s)^2}$.

Technical tools for the use of an EGP

▲ Technical difficulties of an EGP:

- ✦ no exact stochastic simulation
- ✦ no explicit formula for the probability density function(pdf) and cumulative distribution function(cdf)

✔ An approximate EGP with a piecewise constant scale function :

- ✦ simulate approximate paths
- ✦ compute the cdf of a general EGP at a known precision

Z. Al Masry, S. Mercier, G. Verdier,“Approximate simulation techniques and distribution of an extended Gamma process," *Methodology and Computing in Applied Probability* (2015).

Parameter estimation of an EGP

Let $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ a p parameter vector.

$\boldsymbol{X} \sim \Gamma(A(t, \boldsymbol{\theta}), b(t, \boldsymbol{\theta}))$.

Estimate the parameters of an EGP:

- ✖ Standard maximum likelihood estimation is not possible
- ✔ The moments and an explicit form of the Laplace transform are known
- ☞ Generalized method of moments

L. P. Hansen,"Large sample properties of generalized method of moments estimators," *Econometrica* 50(4), 1029-1054(1982)

General approach

Let $\boldsymbol{W}$ be a random vector of dimension d and $\{\boldsymbol{W}_n, n = 1, \ldots, N\}$ a set of i.i.d random vectors having the same distribution as $\boldsymbol{W}$.

Let $\boldsymbol{f} : \mathbb{R}^d \times \Theta \to \mathbb{R}^r$ $(r \geq p)$ be a function such that

$$\boldsymbol{f}(\boldsymbol{w}, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{f}^{(\boldsymbol{1})}(w^{(1)}, \boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{f}^{(\boldsymbol{d})}(w^{(d)}, \boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{w} = \left(w^{(1)}, \ldots, w^{(d)}\right)$ and $\boldsymbol{f}^{(\boldsymbol{i})}(w^{(i)}, \boldsymbol{\theta}), i = 1, \ldots, d$ a vector of dimension k.

General approach

Definition (Population moment condition)

Let $\boldsymbol{\theta}_0$ be the true unknown vector to be estimated. The population moment condition is defined

$$\mathbb{E}[\boldsymbol{f}(\boldsymbol{W},\boldsymbol{\theta}_0)] = 0.$$

Definition (Sample moment condition)

The sample moment condition is derived from the average population moment condition,

$$\hat{\boldsymbol{g}}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{f}(\boldsymbol{W}_n,\boldsymbol{\theta}).$$

General approach

GMM estimator is defined as follows:

Definition

Let $(\boldsymbol{P}_N)$ be a sequence of positive semi–definite weighting matrices that converges in probability to a constant positive definite matrix $\boldsymbol{P}$. Then, the GMM estimator based on these population moments conditions is the value of $\boldsymbol{\theta}$ that minimizes

$$\hat{\boldsymbol{\theta}}_N = arg \min_{\boldsymbol{\theta} \in \Theta} \hat{\boldsymbol{g}}_N(\boldsymbol{\theta})^T \boldsymbol{P}_N \, \hat{\boldsymbol{g}}_N(\boldsymbol{\theta}).$$

Asymptotic properties

W. K. Newey and D. L. McFadden,"Handbook of Econometrics, Large sample estimation and hypothesis testing," *Elsevier Science Publishers, Amsterdam, The Netherlands* 4, 2113-247(1994).

Consistency

Under technical assumptions, $\hat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}_0$.

Asymptotic normality

Under technical assumptions,

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right) \xrightarrow{p} \mathcal{N}(0, \boldsymbol{HSH}^T)$$

where $\boldsymbol{H} = (\boldsymbol{D}_0^T \boldsymbol{P}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0^T\boldsymbol{P}$, $\boldsymbol{D}_0 = \mathbb{E}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{W},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right]$ and $\boldsymbol{S} = \mathbb{E}\left[\boldsymbol{f}(\boldsymbol{W},\boldsymbol{\theta}_0)\boldsymbol{f}(\boldsymbol{W},\boldsymbol{\theta}_0)^T\right]$.

Optimal choice of weighting matrix

A. R. Hall, "Generalized method of moments," *Oxford University Press, Oxford, UK* (2005).

Theorem

If assumptions of the asymptotic normality hold and $\boldsymbol{S}$ is non-singular, then the minimum asymptotic variance of $\hat{\boldsymbol{\theta}}_N$ is

$$\boldsymbol{V} = \left(\boldsymbol{D}_0^T \boldsymbol{S}^{-1} \boldsymbol{D}_0\right)^{-1}$$

and this can be obtained by setting $\boldsymbol{P} = \boldsymbol{S}^{-1}$.

Two-step estimator

Two-step estimator:

1. Set $\boldsymbol{P}_N = \boldsymbol{I}$ and solve

$$\hat{\boldsymbol{\theta}}_N^{(1)} = arg \min_{\boldsymbol{\theta} \in \Theta} \hat{\boldsymbol{g}}_N(\boldsymbol{\theta})^T \hat{\boldsymbol{g}}_N(\boldsymbol{\theta});$$

2. Construct a consistent estimator of $\boldsymbol{S}$ based on $\hat{\boldsymbol{\theta}}_N^{(1)}$

$$\hat{\boldsymbol{S}}_N = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{f}(\boldsymbol{W}_n, \hat{\boldsymbol{\theta}}_N^{(1)}) \boldsymbol{f}(\boldsymbol{W}_n, \hat{\boldsymbol{\theta}}_N^{(1)})^T.$$

The estimator of $\boldsymbol{\theta}_0$ is given by

$$\hat{\boldsymbol{\theta}}_N = arg \min_{\boldsymbol{\theta} \in \Theta} \hat{\boldsymbol{g}}_N(\boldsymbol{\theta})^T \hat{\boldsymbol{S}}_N^{-1} \hat{\boldsymbol{g}}_N(\boldsymbol{\theta}).$$

Two kinds of GMM :

① GMM based on the moments

② GMM based on the Laplace transform

We define the increments of $\boldsymbol{X}$ by

$$W^{(i)} = X_{t_i} - X_{t_{i-1}}, i = 1, 2, .., d$$

where $t_0 = 0 < t_1 < \cdots < t_d = T$.

$$\boldsymbol{W}_n = \begin{pmatrix} W_n^{(1)} \\ \vdots \\ W_n^{(d)} \end{pmatrix}.$$

GMM based on the moments

The sample moment condition is given by

$$\hat{\boldsymbol{g}}_N(\boldsymbol{\theta}) = \begin{pmatrix} \hat{m}^{(1)}(\boldsymbol{\theta}) - m^{(1)}(\boldsymbol{\theta}) \\ \hat{v}^{(1)}(\boldsymbol{\theta}) - v^{(1)}(\boldsymbol{\theta}) \\ \vdots \\ \hat{m}^{(d)}(\boldsymbol{\theta}) - m^{(d)}(\boldsymbol{\theta}) \\ \hat{v}^{(d)}(\boldsymbol{\theta}) - v^{(d)}(\boldsymbol{\theta}) \end{pmatrix}$$

where $m^{(i)}(\boldsymbol{\theta}) = \int_{t_{i-1}}^{t_i} \frac{a(s,\boldsymbol{\theta})ds}{b(s,\boldsymbol{\theta})}$, $v^{(i)}(\boldsymbol{\theta}) = \int_{t_{i-1}}^{t_i} \frac{a(s,\boldsymbol{\theta})ds}{b(s,\boldsymbol{\theta})^2}$,

$\hat{m}^{(i)}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N} W_n^{(i)}$, $\hat{v}^{(i)}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\left(W_n^{(i)} - m^{(i)}(\boldsymbol{\theta})\right)^2$.

GMM estimator is

$$\hat{\boldsymbol{\theta}}_N = arg\min_{\boldsymbol{\theta}\in\Theta} \hat{\boldsymbol{g}}_N(\boldsymbol{\theta})^T \boldsymbol{P}_N\ \hat{\boldsymbol{g}}_N(\boldsymbol{\theta}).$$

GMM based on the Laplace transform

The sample moment condition is given by

$$\hat{\boldsymbol{g}}_N(\boldsymbol{\theta}) = \begin{pmatrix} \hat{\mathcal{L}}^{(1)}(\lambda_1, \boldsymbol{\theta}) - \mathcal{L}^{(1)}(\lambda_1, \boldsymbol{\theta}) \\ \hat{\mathcal{L}}^{(1)}(\lambda_2, \boldsymbol{\theta}) - \mathcal{L}^{(1)}(\lambda_2, \boldsymbol{\theta}) \\ \hat{\mathcal{L}}^{(1)}(\lambda_3, \boldsymbol{\theta}) - \mathcal{L}^{(1)}(\lambda_3, \boldsymbol{\theta}) \\ \vdots \\ \hat{\mathcal{L}}^{(d)}(\lambda_1, \boldsymbol{\theta}) - \mathcal{L}^{(d)}(\lambda_1, \boldsymbol{\theta}) \\ \hat{\mathcal{L}}^{(d)}(\lambda_2, \boldsymbol{\theta}) - \mathcal{L}^{(d)}(\lambda_2, \boldsymbol{\theta}) \\ \hat{\mathcal{L}}^{(d)}(\lambda_3, \boldsymbol{\theta}) - \mathcal{L}^{(d)}(\lambda_3, \boldsymbol{\theta}) \end{pmatrix},$$

where $\hat{\mathcal{L}}^{(i)}(\lambda_k, \boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \exp(-\lambda_k W_n^{(i)}),\ k = 1, 2, 3.$

- λ_2 and λ_3 are multiples of λ_1.

Asymptotic properties

Consistency

If the following assumptions hold

M_1. $(\boldsymbol{P}_N)$ converges in probability (almost surely) to a constant positive definite matrix $\boldsymbol{P}$;

M_2. Identification : $\mathbb{E}\left[\boldsymbol{f}(\boldsymbol{W},\boldsymbol{\theta})\right] = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$;

M_3. Compactness : Θ is compact ;

M_4. $\left|\frac{a(s,\boldsymbol{\theta})}{b(s,\boldsymbol{\theta})}\right| \leq J_1(s)$, $\int_{t_{i-1}}^{t_i} J_1(s)ds < \infty$;

M_5. $\left|\frac{a(s,\boldsymbol{\theta})}{b(s,\boldsymbol{\theta})^2}\right| \leq J_2(s)$, $\int_{t_{i-1}}^{t_i} J_2(s)ds < \infty$

then $\hat{\boldsymbol{\theta}}_N \xrightarrow[N\to\infty]{p\ (a.s.)} \boldsymbol{\theta}_0$.

Asymptotic properties

Asymptotic normality

Assumptions $(M_1 - M_5)$ and

M_6. $\boldsymbol{\theta}_0$ is an interior point in Θ ;

M_7. $\boldsymbol{D}_0^T \boldsymbol{P} \boldsymbol{D}_0$ is non-singular ;

M_8. $\left| \frac{\partial}{\partial \boldsymbol{\theta}^T} \left(\frac{a(s,\boldsymbol{\theta})}{b(s,\boldsymbol{\theta})} \right) \right| \leq l_1(s)$ where $\int_{t_{i-1}}^{t_i} l_1(s) ds < \infty$;

M_9. $\left| \frac{\partial}{\partial \boldsymbol{\theta}^T} \left(\frac{a(s,\boldsymbol{\theta})}{b(s,\boldsymbol{\theta})^2} \right) \right| \leq l_2(s)$ where $\int_{t_{i-1}}^{t_i} l_2(s) ds < \infty$;

M_{10}. $\int_{t_{i-1}}^{t_i} \frac{a(s,\boldsymbol{\theta}) ds}{b(s,\boldsymbol{\theta})^3} < \infty, \int_{t_{i-1}}^{t_i} \frac{a(s,\boldsymbol{\theta}) ds}{b(s,\boldsymbol{\theta})^4} < \infty$

imply $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \to \mathcal{N}(0, \boldsymbol{HSH}^T)$ where
$\boldsymbol{H} = (\boldsymbol{D}_0^T \boldsymbol{P} \boldsymbol{D}_0)^{-1} \boldsymbol{D}_0^T \boldsymbol{P}$ and $\boldsymbol{D}_0 = \mathbb{E}\left[\frac{\partial \boldsymbol{f}(\boldsymbol{W}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} \right]$.

Asymptotic properties

Optimal choice of weighting matrix

If the previous assumptions hold then the minimum asymptotic variance of $\hat{\theta}_N$ is $\boldsymbol{V} = \left(\boldsymbol{D}_0^T \boldsymbol{S}^{-1} \boldsymbol{D}_0\right)^{-1}$.

▲ Difficulty: $\boldsymbol{S}$ is non-singular

Choosing parametric forms

$\boldsymbol{\theta} = (a, \alpha, b, \beta, c)$

① $a(t, \boldsymbol{\theta}) = at^{\alpha}dt, b(t, \theta) = b(t + c)^{\beta}$

$$\Theta_1 = \left(\mathbb{R}_+^{*\,3} \times \mathbb{R} \times \mathbb{R}_+\right) \cup \{(a, \alpha, b, \beta, c) \in \mathbb{R}_+^{*\,3} \times \mathbb{R} \times \{0\} / \alpha > 2\beta - 1\}.$$

② $a(t, \boldsymbol{\theta}) = a\alpha \exp(-\alpha s)dt, b(t, \theta) = b(1 - \exp(-\beta(s + c))$

$$\Theta_2 = \{(a, \alpha, b, \beta, c) \in \mathbb{R}_+^{*\,2} \times \mathbb{R}_-^{*\,2} \times \mathbb{R}_+^{*} / b \times \beta > 0\}$$
$$\cup \{(a, \alpha, b, \beta, c) \in \mathbb{R}_+^{*\,2} \times \mathbb{R}_+^{*\,2} \times \mathbb{R}_+^{*}\}.$$

✔ Asymptotic properties

▲ Difficulty: Identification

TABLE: $A(t) = t^\alpha$, $b(t) = t^\beta$, *Number of increments* $= 10$, $T = 10$, $N = 400$, *Number of replications* $= 500$

	$\hat{\alpha}$	$\hat{\beta}$
True value	2	0.5
	Mean (std)	Mean (std)
GMM_{MM}	1.9954 (0.0363)	0.4942 (0.0457)
GMM_{Lap}	1.9994 (0.0320)	0.4991 (0.0413)
	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$
GMM_{MM}	$[1.9235, 2.0655]$	$[0.4027, 0.5839]$
GMM_{Lap}	$[1.9398, 2.0610]$	$[0.4228, 0.5793]$

TABLE: $\mathbb{P}[N(\hat{\theta}_N - \theta_0)\hat{V}^{-1}(\hat{\theta}_N - \theta_0)^T \leq \chi^2_{0.95,2}]$

	CP
GMM_{MM}	94.8%
GMM_{Lap}	94.4%

TABLE: $A(t) = at^{\alpha}$, $b(t) = bt^{\beta}$, *Number of increments* $= 10$, $T = 10$, $N = 500$, *Number of replications* $= 500$

	$\hat{a}$	$\hat{\alpha}$	$\hat{b}$	$\hat{\beta}$
True value	1	2	1	0.5
	Mean (std)	Mean (std)	Mean (std)	Mean (std)
GMM_{MM}	1.0402 (0.0635)	1.9897 (0.0282)	1.0398 (0.0538)	0.4873 (0.0299)
GMM_{Lap}	1.0053 (0.0501)	2.0030 (0.0247)	1.0083 (0.0440)	0.5023 (0.0270)
	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$
GMM_{MM}	$[0.9138, 1.1725]$	$[1.9369, 2.0478]$	$[0.9387, 1.1457]$	$[0.4286, 0.5487]$
GMM_{Lap}	$[0.9159, 1.1081]$	$[1.9534, 2.0513]$	$[0.9301, 1.1025]$	$[0.4490, 0.5582]$

TABLE: $\mathbb{P}[N(\hat{\theta}_N - \theta_0)\hat{V}^{-1}(\hat{\theta}_N - \theta_0)^T \leq \chi^2_{0.95,4}]$

	CP (N=500)	CP (N=800)
GMM_{MM}	84.8%	87.6%
GMM_{Lap}	89.8%	93.4%

TABLE: $A(t) = at^{\alpha}$, $b(t) = b(t+c)^{\beta}$, *Number of increments* $= 10$, $T = 10$, $N = 3000$, *Number of replications* $= 500$

	$\hat{a}$	$\hat{\alpha}$	$\hat{b}$	$\hat{\beta}$	$\hat{c}$
True value	1	2	1	2	1
	Mean (std)	Mean (std)	Mean (std)	Mean (std)	Mean (std)
GMM_{MM}	1.0079 (0.0364)	1.9979 (0.0173)	1.0022 (0.0494)	1.9997 (0.0214)	1.0068 (0.0496)
GMM_{Lap}	1.0023 (0.0192)	1.9998 (0.0097)	1.0002 (0.0502)	2.0008 (0.0196)	1.0028 (0.0454)
	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$	$[Q_{.025}, Q_{.975}]$
GMM_{MM}	[0.9489, 1.0684]	[1.9693, 2.0272]	[0.9066, 1.0992]	[1.9590, 2.0397]	[0.9196, 1.0946]
GMM_{Lap}	[0.9682, 1.0439]	[1.9801, 2.0184]	[0.9022, 1.0931]	[1.9640, 2.0396]	[0.9156, 1.1064]

TABLE: $\mathbb{P}[N(\hat{\theta}_N - \theta_0)\hat{V}^{-1}(\hat{\theta}_N - \theta_0)^T \leq \chi^2_{0.95,5}]$

	CP
GMM_{MM}	87.2%
GMM_{Lap}	88%

- GMM for an EGP seems to behave well for estimating the unknown parameters. It also provides asymptotic properties.
- GMM based on the Laplace transform is more performing as shown in the empirical example.

- The next step is to compare the application of SGP and EGP to reliability.