Original Research Paper

Univariate Generalized Additive Models for Simulated Stationary and Non-Stationary Generalized Pareto Distribution

[1]**Mostafa Behzadi,** [1,2]**Mohd Bakri Adam and** [1]**Anwar Fitrianto**

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia*
[2]*Institute for Mathematical Research, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia*

Abstract: Generalized additive models as a predictor in regression approaches, are made up over cubic spline basis and penalized regression splines. Despite of linear predictor in GLM, generalized additive models use a sum of smooth functions of covariates as a predictor. The data which are used in this study have generalized Pareto distribution and have been simulated by inversion method. The data are generated in two types, the stationary case and the non-stationary case. The method of root mean square of errors as a method of measurement is used for comparison between power of predictions which are based on penalized regression splines as a method in univariate generalized additive models and linear regression based on maximum likelihood estimation. The finding of this research illustrates that the amount of accuracy of estimation of parameter of location in UGAM approach as an alternative promising of modelling through each specialized GPD's models, has less RMSE in compare with MLE.

Keywords: Generalized Pareto Distribution, Univariate Generalized Additive Model, Smooth Function, Penalized Regression Spline, Cubic Spline Basis, Simulated Data, Maximum Likelihood Estimation, Root Mean Square of Errors

Introduction

During early of 1990, generalized Additive models, GAMs, by attempts of Hastie and Tibshirani have been developed (Hastie and Tibshirani, 1990). The construction of GAM, is relied on back-fitting with linear smoothers. Although this model seems to be flexible, but, in practice, some difficulties like how to inference and how to select the model be revealed. To get the solution of these difficulties, the sensitive mathematical work of Wahba and her colleagues (Wahba, 1990) on generalized spline smoothing, GSS, provided a consistent and precise framework to inference. Smoothing splines, are basis for model selection (Gu and Wahba, 1991a) with GAMs (Gu and Wahba, 1991b). Computationally, the selection of these models are very expensive. Applying penalized regression splines, are a "middle approach" (Wahba and Wendelberger, 1980) to build GAMs (Eilers and Marx, 1996; Wood, 2006). Recall that GAMs is based on

smooth functions and has ability to predict. Therefore, for better prediction, a basis and "wiggliness" penalty for every smooth function in the model is selected. Univariate generalized additive models, provide a flexible and smooth approach to distinguish and identify the non-linearity's covariate effects in a several types of modelling situations and circumstances (Wood, 2006).

The simulated data are generated based on inversion method, in this study. It has been used from R software to program the approaches, problems and solutions. Generalized pareto distribution is belong to a family of continuous probability distributions and often is used to model the tails of another distributions. At first, this paper will pay attention to the root of origination of univariate generalized additive models and then will go to concentrate on parameter of location. The goal is to display the ability of amount of accuracy of estimation of parameter of location. To get the goal, the method of penalized regression spline in univariate generalized

additive models in stationary and non-stationary GPD data are compared in contrast of another approach, MLE.

Section 2, presents the basic materials in details of structure of GAM. Section 3, deals with different types of simulated GPD data and their inversions. The maximum likelihood estimation of each type of data are calculated and provided in section 3, as well. Section 4, display the implementation of GAM and goes to the details of this method like cubic spline and penalized regression spline. Section 5, is consist of results of this research. The results are included some figures and related tables. The conclusion is placed in section 6.

Basic Materials

GAM has resemblance to a GLM in which a linear combination of scatter-plot smoothers is replaced instead of linear combination of explanatory variables. The general structure of a GAM can be illustrated as bellow:

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + ... \qquad (1)$$

where, X_i^*, is a row of the model matrix for any strictly parametric model components, θ, is the corresponding parameter vector and f_j, are smooth functions of the covariates, x_k (Wood, 2006). In addition:

$$\mu_i = E(Y_i)$$

where, Y_i, is a dependent variable which is distributed as some of the exponential family distribution.

The indication of smooth function is best introduced with respect to a model which has one smooth function of one covariate. This model is:

$$y_i = f(x_i) + \varepsilon_i \qquad (2)$$

where, y_i, is an outcome or response variable and x_i is covariate, f is a smooth function and $\varepsilon_i \sim N(0, \sigma^2)$ random variables, for $i = 1, 2, ..., n$ (Wood, 2006).

To show f, as a simple predictor, one can choose a known basis function.

Suppose that $b_j(x)$ is the jth basis function, then f, is modelled to have a formula (Wood, 2006) like:

$$f(x) = \sum_{j=1}^{q} b_j(x) \beta_j \qquad (3)$$

for a number of unknown parameters which should be estimated. By replacement of Equation 3 into $y_i = f(x_i) + \varepsilon_i$, a linear model will be constructed.

A cubic spline can be applied as a representative for univariate functions. A cubic spline is made of combining of cubic polynomials which are connected to

each other to build a special curve. This new curve is continuous in value, as well as, in first and second derivatives (Wood, 2000). The knots of the splines are illustrated by the points where the sections have connection (Hastie and Tibshirani, 1987). What approach is applied, let the location of knots be marked by $\{x_i^* : i = 1, ..., q - 2\}$, where, $q \in N$ (Wood, 2006). For a formal spline, wherever there is some datum, the knots occur. In general, the knots could either be spread among the range of observed x, or put it at quantiles of the distribution of unique x values (Hastie and Tibshirani, 1987). With regard to the location of knots, many alternatives can occur (Gu, 2002). Let consider basis $b_1(x) = 1$, $b_2(x) = x$ and $b_i + 2 = R(x, x^*)$ for $i = 1, ..., q-2$, where:

$$R(x,z) = \frac{\left[\left(z - \frac{1}{2}\right)^2 - \frac{1}{12}\right]\left[\left(x - \frac{1}{2}\right)^2 - \frac{1}{12}\right]}{4}$$
$$- \frac{\left[\left(|x - z| - \frac{1}{2}\right)^4 - \frac{1}{2}\left(|x - z| - \frac{1}{2}\right)^2 + \frac{7}{240}\right]}{24}$$

In order to making function of "f", by applying the cubic spline basis Equation 2, will be a linear model: $Y = X\beta + \varepsilon$, where the ith row of the model matrix is:

$$X_i = \left[1, x_i, R(x_i, x_1^*), R(x_i, x_2^*), ..., R(x_i, x_{q-2}^*)\right]$$

therefore, the model could be approximated by least squares (Wood, 2002; Gu and Kim, 2002).

Instead of fitting the model by minimizing:

$$\|y - X\beta\|^2$$

where, Y, is a dependent variable and X, is model matrix. We are able to minimize:

$$\|y - X\beta\|^2 + \lambda \int_0^1 [f''(x)]^2 d(x)$$

in which λ, displays the smooth parameter in where the integrated square of second derivative penalizes models that are too wiggly. Always, the penalty can be written as a quadratic form in β, i.e.:

$$\int_0^1 [f''(x)]^2 d(x) = \beta^T S \beta$$

where, S, is a matrix of known coefficients. Here, the form of spline basis is a little complicated and this complication proves its worth. Notice that: $S_{i+2, j+2} = R(x_i^*, x_j^*)$ for $\{i, j = 1, ..., q - 2\}$, while, the first

two rows and columns of *S* are 0. Accordingly, indeed, the penalized regression spline fitting problem is to minimize:

$$\|y - X\beta\|^2 + \lambda\beta^T S\beta$$

with regard to *β*. In practice, the explanations above, is not useful for computation. The method of orthogonal matrix offers greater numerical stability. In practical solution, the bellow formula can be applied:

$$\left\| \begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} X \\ \sqrt{\lambda}B \end{bmatrix} \beta \right\|^2 = \|y - X\beta\|^2 + \lambda\beta^T S\beta$$

where $\vec{B}$, is any square root of the matrix *S*, note that $B^T B = S$ (Wood, 2006).

This research is used from generalized cross validation, GCV, as an evaluator method to trade off the validity and accuracy of modelling. GCV has computational benefits, in addition, has benefits in terms of invariance (Wahba, 1990; Gu, 1992).

Simulated Data

This section illustrates the special GPD distribution function briefly. The GPD distribution function is belong to exponential distributions functions family and has three parameters, location: $\mu \in R$, scale: $\sigma > 0$ and shape $\xi \in R$ (Arnold, 2011). If the parameters of shape and location be equal to zero, then generalized Pareto distribution is equal to exponential distribution. On the other hand, the GPD is equal to Pareto distribution where shape is considered to be $\xi > 0$ and location is regarded to be $\mu = \frac{\sigma}{\xi}$ (Coles *et al.*, 2001).

This paper deals with two types of GPD data:

Stationary Case

The meaning of stationary case of GPD distribution function is the function which has determined parameters. In other words, all parameters such as *μ*, *σ* and *ξ* are equal to a specified value, in their defined domain and are constant. In order to generate simulated stationary data, the inversion method to simulate should be applied computing the inversion function of stationary case of GPD distribution function makes an ability to produce the stationary data:

$$F(x; \frac{\sigma}{\xi}, \sigma, \xi) = 1 - \left[1 + \xi\left(\frac{x - \frac{\sigma}{\xi}}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \text{ for } \xi \neq 0 \tag{4}$$

therefore, the inverse of this function will be:

$$y = \frac{\sigma}{\xi}\left[\frac{1}{(1-x)^\xi}\right], \quad \text{for } x \geq \frac{\sigma}{\xi} \tag{5}$$

$$f(x; \frac{\sigma}{\xi}, \sigma, \xi) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x - \frac{\sigma}{\xi}}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}+1\right)} \tag{6}$$

to calculate the function of MLE, the production of p.d.f is called $L(\theta)$, where, *θ*, is representative of all parameters:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma}\left[1 + \xi\left(\frac{x_i - \frac{\sigma}{\xi}}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}+1\right)} \tag{7}$$

consequently:

$$\log(L(\theta)) = -n\log(\sigma) - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\log\left[1 + \xi\left(\frac{x_i - \frac{\sigma}{\xi}}{\sigma}\right)\right]. \tag{8}$$

Calculation of maximum likelihood of stationary case of GPD distribution function gives an opportunity to estimate of parameters and obtains base on probability distribution function of GPD:

Non-stationary case

Whereas the parameter of location in this case of GPD, μ_t, is equal to $\frac{\sigma_t}{\xi}$, then, the non-stationary case of GPD has a trend in its mean while scattering of data is increased:

$$\mu \to \mu_t : \frac{\sigma}{\xi} \to \frac{\sigma_t}{\xi} \to \frac{\kappa t}{\xi}$$

The function after replacement is:

$$F\left(x, t; \frac{\kappa}{\xi}, \kappa, \xi\right) = 1 - \left\{1 + \xi\left[\frac{\left(x - \frac{\kappa}{\xi}t\right)}{\kappa t}\right]\right\}^{-\frac{1}{\xi}} \tag{9}$$

and the inversion of this function is obtained:

$$y = \frac{\kappa t}{\xi}\left[(1-x)^{-\xi}\right] \tag{10}$$

The most important usage of maximum likelihood function is estimation of unknown parameters. Indeed, the duty of this function is to estimate the two parameters included *σ* and *ξ*. By substitution, the

parameter of location automatically is calculated, $\mu_t = \dfrac{\sigma_t}{\xi}$. To get the goal:

$$f\left(x, t; \frac{\kappa t}{\xi}, \kappa, \xi\right) = \frac{1}{\kappa t}\left\{1 + \xi\left[\frac{x - \left(\frac{\kappa}{\xi}t\right)}{\kappa t}\right]\right\}^{-\left(\frac{1}{\xi}+1\right)} \tag{11}$$

suppose that θ, is representative of all parameters, therefore, in production:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\kappa t_i}\left\{1 + \xi\left[\frac{x_i - \left(\frac{\kappa}{\xi}t_i\right)}{\kappa t_i}\right]\right\}^{-\left(\frac{1}{\xi}+1\right)} \tag{12}$$

consequently:

$$\log(L(\theta)) = -n\log(\kappa) - \sum_{i=1}^{n}\log(t_i)$$
$$-\left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\log\left[1 + \xi\left(\frac{x_i - \left(\frac{\kappa}{\xi}t_i\right)}{\kappa t_i}\right)\right] \tag{13}$$

Implementation of Univariate Generalized Additive Models for Each Set of Data

The construction of GAM is based on cubic spline basis and penalized regression spline. In fact, in this structure, polynomial bases are very useful for positions which interest concentrates on attributes of f in the neighborhood of a single determined point (Gu, 1992). However, when the query of interest relates to f on its whole domain ([0,1]), the polynomial bases have some difficulties on the border. The splines bases fulfill well in this circumstances, greatly, because spline basis can be illustrated to have good approximation theoretic properties.

First of all, it should be written a function which has ability to take a sequence of an array of x values to make a model matrix for the spline. Secondly, the only thing that remains is to select a sequence of knots, x_i^*. Briefly, it is displayed how to deal with these steps inside the programming:

- get the knots
- get the penalty matrices
- get the model matrix

Stationary Data

The feature of histogram and density function of GPD in stationary case has been displayed in Fig. 1a. As it is observed in Fig. 1b, there are six predicted colored lines. Each sub-figure represents a number of specified knots which could describe the function of GAM only for parameter of location. The fitting of the model seems enough reliable and acceptable. Adjusting and qualifying the basis dimension, q, is arbitrary. In other words, the selection of degree of model smoothness which is the number of knots +2, is basically optional (Wood, 2006; Durrleman and Simon, 1989). The arbitrary selection needs to more research and means that if a satisfactory theory for modelling with smooth function is to be extended, the number of knots must be shown. The smoothing parameter, λ, controls the specification between model fit and model smoothness (Kim and Gu, 2004). If $\lambda = 0$, means that there is no penalty and if $\lambda \to \infty$, means that something like a straight line is going to estimate f (Wood and Augustin, 2002; Eilers and Marx, 2010).

To apply the estimated function effectively, selection of q, x_j^* and λ are very significant. Hence, it must be considered to an important matter which emphasizes that the selecting of number of knots should be large enough, which basis functions could provide enough flexibility to illustrate $f(x)$. Accordingly, neither the precise of knots locations nor the exact selection of number of knots has a big influence on the model fit. As regards, the selected λ, has a great role in determining model flexibility and finally the approximated form of $\hat{f}(x)$. By determining $\lambda = 10^{-8}$ as a default to originate the program, then $\hat{\lambda}$, with respect to minimization is obtained by using this formula (Wood and Augustin, 2002):

$$\hat{\lambda} = 1.5^{\{i-1\}} * \lambda \tag{14}$$

According to Equation 14, $\hat{\lambda}$ for stationary data is obtained:

$$\hat{\lambda} = 1.5^{43-1} * 10^{-8} = 0.2479$$

Note that the program begins by a chosen value as a default for λ, empirically. In continue, according to the GCV function, the fitted model is depicted. In Fig. 1c the minimum of GCV function has occurred in point of 43 and V (i) is equal to 2.396. Indeed, this point is the global landmark of this function. According to this obtained landmark, it is possible to guess that the fitted line has a big curve along itself. The numbers of curves during the fitted line are a reaction of global and local landmarks of GCV function. Figure 1d is the reference for application of minimum GCV function.
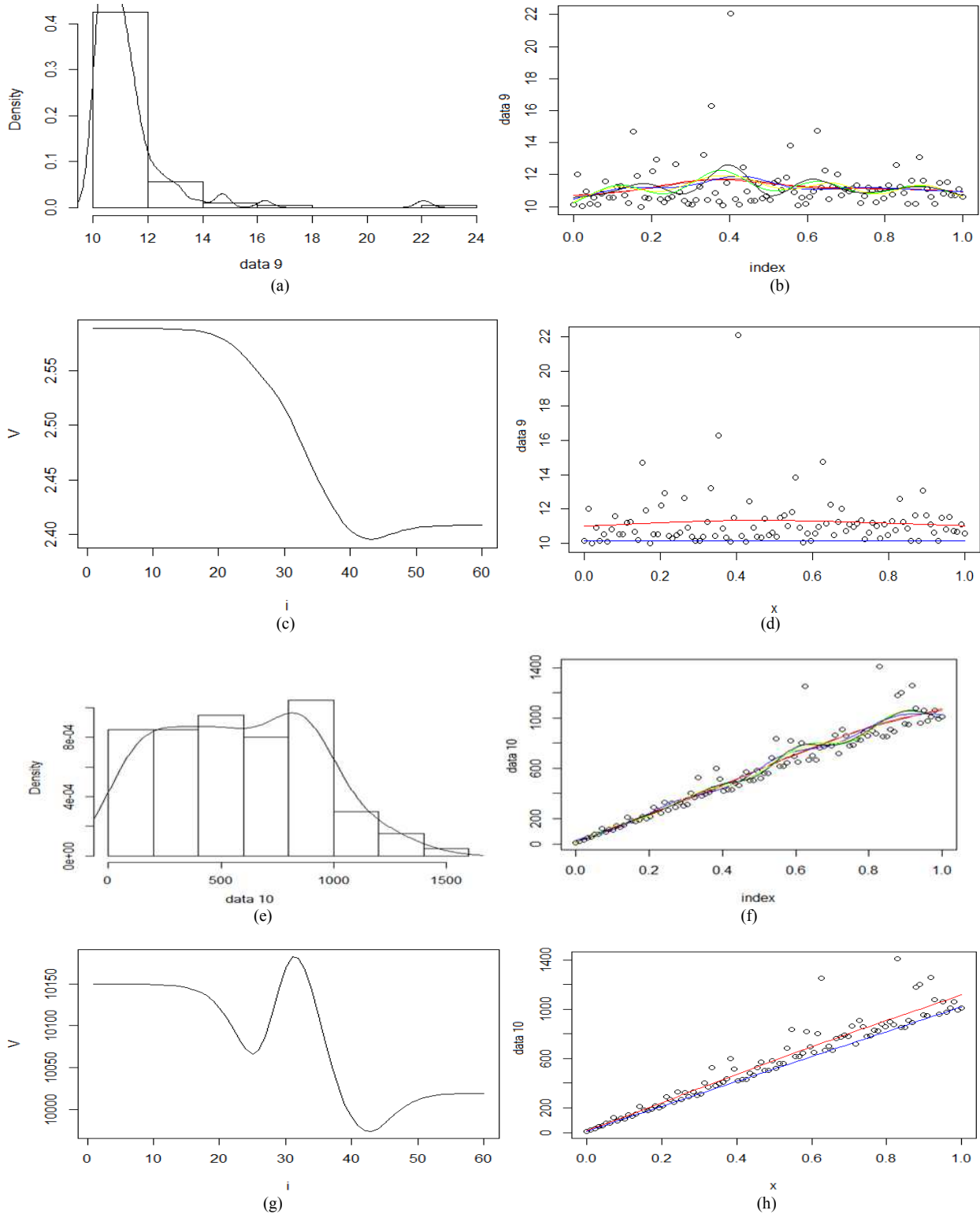
Figure 1. (a, e) density function and histogram of GPD in stationary case and non- stationary case respectively. (b, f) regression spline fit in stationary case and in non-stationary case, the brown, red, blue, green, yellow and black depicted graph has applied 4, 5, 6, 7, 8 and 9 knots respectively. (c, g) Min GCV functions, stationary: i = 43, V (i) = 2.396 and in non-stationary: i = 43, V (i) = 9974.124. (d, h) optimal fitted model for parameter of location in stationary case and in non-stationary case, blue line is depicted based on MLE's estimation and red line is based on estimation of PRS in GAM. n = 100

Non-Stationary Data

The non-stationary of GPD distribution function has a trend in its parameter of location, μ_t (Hamilton, 1994). This parameter, itself, is a new function and has two parameters such as σ and ξ. In this case σ_t is equal to κt, therefore the μ_t will be equal to $\frac{\kappa}{\xi}t$. The shape parameter can take any value in the range of real number and the scale parameter can take any non-negative real number. As the μ_t shows itself, it is a linear function without intercept. In this function the slope is the $\frac{\kappa}{\xi}t$. Figure 1h proves that there is no intercept for the predicted lines based on this case of GPD non-stationary data in the origin of the axes. With respect to the properties of μ_t, the data have two specifications simultaneously, the first specification is along the coefficient of slope and the second one is scattering generally more and more by increasing the value of "*t*" in each step (Dominici *et al.*, 2002). The histogram and density function of GPD non-stationary function is illustrated in Fig. 1e. The six PRS fitted models are located in Fig. 1f. With refer to Fig. 1f, each fitted model from these six depicted PRS, displays a determined number of knots. Based on Equation 14 the computing of λ leads to 0.2479. Figure 1g, illustrates that there is a global landmark at $i = 43$ and the value of minimum GCV function at this point is 9974.124. Hence, based on global and other local landmarks, the final fitted model is impressed proportionally to these landmarks Fig. 1h.

Results and Discussions

The result of this paper is gathered concisely in Table 1 and 2. The explanation of each type of data

with consideration to related figure is categorized in below respectively:

This panel demonstrates GPD stationary data. An illustration of the fitting models to predict the parameter of location based on estimated GAM and MLE's method are located in Fig. 1d. The blue line has been depicted according to MLE's estimations and the red one has been drawn based on estimation of GAM. These data have got homoscedasticity because the data scattered with the same variance. In Table 3, has been listed some various sample sizes with n = 20, 100 and 500. The column of GPD stationary data makes the ability to test the comparison of RMSE between two estimators. In each sample size, the RMSE which is calculated by GAM's estimation is less than MLE's method. As it is observed, the structure of GAM helps to predict the parameter of location better among this type of data. By investigating the stationary column in Table 3, after increasing the sample size from 100 to 500, a decrease has happened in amount of values of GAM's estimation for obtaining RMSE.

Figure 1h displays two predictions for equation of μ_t among the GPD non-stationary data. The μ_t is equal to $\frac{\kappa}{\xi}t$. Thus, these data have got heteroscedasticity, because scattered with different variances. With respect to the Fig. 1h, the red line or GAM's prediction, is almost a straight line as the blue prediction which is depicted based on MLE. As it is realized from this panel, the prediction based on GAM has been depicted among the data, but the prediction based on MLE, has been drawn near the lower threshold of these data. Hence, by looking at this phenomenon and before searching Table 3, it is possible to guess that the estimation by GAM has less RMSE for prediction of parameter of location, rather than MLE's method.

Table 1. The estimated parameters for stationary GPD distribution based on inversion method. There are three parameters for estimation: location, scale and shape. The estimation of location is illustrated by $\hat{\mu}$, scale by $\hat{\sigma}$ and shape by $\hat{\xi}$: In this table for more accuracy, the range of estimations have been calculated up to three digits after decimal. The true values for μ and ξ are 1.2 and 0.12 respectively

Sample size	$\hat{\sigma}$	Bias	95% CI
n = 20	1.83	0.63	(1.28,2.38)
n = 100	1.37	0.17	(1.206,1.534)
n = 500	1.33	0.13	(1.261,1.399)
Sample size	$\hat{\xi}$	Bias	95% CI
n = 20	0.167	0.047	(0.117,0.217)
n = 100	0.129	0.009	(0.106,0.152)
n = 500	0.127	0.007	(0.107,0.147)
Sample size	$\hat{\mu} = \frac{\hat{\sigma}}{\xi}$	Bias	95% CI
n = 20	10.96	0.96	(9.170,12.720)
n = 100	10.65	0.65	(9.430,11.860)
n = 500	10.48	0.48	(9.350,11.610)

Table 2. The estimated parameters for non-stationary GPD distribution according to inversion method. The three parameters are: Location, scale and shape. The estimation of parameters are demonstrated for scale by $\hat{\sigma}$, shape by $\hat{\xi}$; therefore, the illustration of estimation of location will be $\mu_t = \dfrac{\hat{\sigma}_t}{\hat{\xi}}$, note that: $\hat{\sigma}_t = kt$. In this table, for more accuracy the estimations have been calculated based on three digits after decimal. True values for k and ξ are: 1.2 and 0.12 respectively

Sample size	$\hat{\kappa}$	Bias	95% CI
n = 20	1.90	0.70	(-0.746,4.546)
n = 100	1.35	0.15	(-0.260,2.980)
n = 500	1.27	0.07	(0.976,1.560)
Sample size	$\hat{\xi}$	Bias	95% CI
n = 20	0.175	0.055	(-0.073,0.433)
n = 100	0.132	0.012	(-0.026,0.239)
n = 500	0.126	0.006	(0.082,0.170)
Sample size	$\hat{\mu} = \dfrac{\hat{\kappa}}{\hat{\xi}}$	Bias	95% CI
n = 20	10.86	0.86	(9.79,11.93)
n = 100	10.26	0.26	(9.89,10.63)
n = 500	10.08	0.08	(9.95,10.21)

Table 3. A comparison by RMSE over parameter of location for GPD data in stationary and nonstationary cases between estimations based on estimated GAM and method of MLE

	RMSE Types of GPD Data			
	Stationary		Non-Stationary μ_τ	
Sample size	MLE	Est. Based on GAM	MLE	Est. Based on GAM
n = 20	4.291	1.327	88.172	26.137
n = 100	14.069	1.978	651.184	101.437
n = 500	24.194	1.637	7370.33	462.037

The column which displays the non-stationary case in Table 3, is contain the calculated RMSE for different sample sizes. According to Table 3, by increasing the number of data in each sample size, the RMSE based on MLE's method, increases much more than the RMSE which has been calculated according to GAM's estimation. In estimation based on MLE in Table 3, there is just growth in value of RMSE with raising the sample size, but in estimation according to GAM, for n = 500, there is a decrease in calculated RMSE rather than n = 100.

Conclusion

The concentration and focus of studying in this study is over parameter of location. The results of this paper illustrate the extension of conception of generalized additive models in order to evaluate the accuracy of estimation of parameter of location. This development is displayed by related figures. In addition, the parameter of location is monitored by method of MLE, as well. In order to compare the amount of accuracy of estimation of parameter of location 1 is prepared. According to RMSE values, the amount of accuracy of estimation of parameter of location based on PRS in GAM is better than Method of MLE via GPD models.

In general terms, the finding illustrate that the estimation based on GAM has less root mean square of errors, hence, the estimation based on GAM is more powerful than findings based on MLE.

Acknowledgment

Funding Information

Author's Contributions

All authors equally contributed in this work.

Ethics

This paper is constructed based on our research and each step and all findings are original.

References

Arnold, B.C., 2011. Pareto and Generalized Pareto Distribution. In: Chotikapanich, D. (Ed.), Modelling Distributions and Lorenz Curves, Springer, New York.

Coles, S., J. Bawa, L. Trenner and P. Dorazio, 2001. An Introduction to Statistical Modeling of Extreme Values. 1st Edn., Springer Science and Business Media, London, ISBN-10: 1852334592, pp: 208.

Dominici, F., A. McDermott, S.L. Zeger and J.M. Samet, 2002. On the use of generalized additive models in time-series studies of air pollution and health. Am. J. Epidemiol., 156: 193-203. PMID: 12142253

Durrleman, S. and R. Simon, 1989. Flexible regression models with cubic splines. Stat. Med., 8: 551-561. DOI: 10.1002/sim.4780080504

Eilers, P.H.C. and B.D. Marx, 1996. Flexible smoothing with B-splines and penalties. Stat. Sci., 11: 89-121.

Eilers, P.H.C. and B.D. Marx, 2010. Splines, knots and penalties. Comput. Stat., 2: 637-653. DOI: 10.1002/wics.125

Gu, C. and G. Wahba, 1991a. Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. J. Scientific Stat. Comput., Volume: 12: 383-398. DOI: 10.1137/0912021

Gu, C. and G. Wahba, 1991b. Discussion: Multivariate adaptive regression splines. Annals. Stat., 19: 115-123.

Gu, C. and Y.J. Kim, 2002. Penalized likelihood regression: General formulation and efficient approximation. Can. J. Stat., 30: 619-628. DOI: 10.2307/3316100

Gu, C., 1992. Cross-validating non-Gaussian data. J. Comput. Graph. Stat., 1: 169-179. DOI: 10.2307/1390840

Gu, C., 2002. Smoothing Spline ANOVA Models. 1st Edn., Springer, New York, ISBN-10: 0387953531, pp: 289.

Hamilton, J.D., 1994. Time Series Analysis. 1st Edn., Princeton University Press, Princeton, ISBN-10: 0691042896, pp: 799.

Hastie, T. and R.M.F. Tibshirani, 1987. Generalized additive models: Some applications. J. Am. Stat. Assoc., 82: 371-386. DOI: 10.2307/2289439

Hastie, T.J. and J. Tibshirani, 1990. Generalized Additive Models. 1st Edn., CRC Press, Florida, ISBN-10: 0412343908, pp: 352.

Kim, Y.J. and C. Gu, 2004. Smoothing spline Gaussian regression: More scalable computation via efficient approximation. J. Royal Stat. Society, 66: 337-356. DOI: 10.1046/j.1369-7412.2003.05316.x

Wahba, G. and J. Wendelberger, 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. Monthly Weather Rev., 108: 1122-1122. DOI: 10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2

Wahba, G., 1990. Spline Models for Observational Data. 1st Edn., SIAM, Philadelphia, ISBN-10: 0898712440, pp: 169.

Wood, S., 2006. Generalized Additive Models: An Introduction with R. 1st Edn., CRC Press, USA., ISBN-10: 1584884746, pp: 410.

Wood, S.N. and N.H. Augustin, 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modeling. Ecol. Modell., 157: 157-177. DOI: 10.1016/S0304-3800(02)00193-X

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. Royal Stat. Society, 62: 413-428. DOI: 10.1111/1467-9868.00240