Rigby & Stasinopoulos (2005)
Generalized Additive Models for Location, Scale and Shape

Simon Hirsch

University of Duisburg-Essen, House of Energy Markets & Finance
Statkraft Trading GmbH

RuhrMetrics Reading Group, January 2025

Main Idea

GLM is a powerful model in linear regression space. However, variance, skewness and kurtosis of the distribution are only modeled implicitly through their dependence on μ, e.g.

- Exponential Distribution: mean $1/\lambda$, variance $1/\lambda^2$
- Bernoulli Distribution: mean p, variance $p(1-p)$
- Gamma, ...

Main Idea

GLM is a powerful model in linear regression space. However, variance, skewness and kurtosis of the distribution are only modeled implicitly through their dependence on μ, e.g.

- Exponential Distribution: mean $1/\lambda$, variance $1/\lambda^2$
- Bernoulli Distribution: mean p, variance $p(1-p)$
- Gamma, ...

Alleviate this restriction:

> *In this paper we develop a general class of univariate regression models which we call the generalized additive model for location, scale and shape (GAMLSS), where the exponential family assumption is relaxed and replaced by a very general distribution family. [RS05]*

Main Idea

- We have some independent (not i.i.d. (!)) data

$$Y = (y_1, y_2, ..., y_i, ..., y_N)$$

  with some probability density function

$$f(y_i \mid \boldsymbol{\theta}_i),$$

  where $\boldsymbol{\theta}_i = (\theta_{1i}, ..., \theta_{1K})$ is a vector for up to K distribution parameters
- Model each distribution parameter as regression equation conditional on explanatory variables and additive effects
- Note that $\boldsymbol{\theta}_i$ has the index i, i.e. the shape of $\boldsymbol{\theta}$ is $N \times K$, where K is the number of distribution parameters

This gives a very rich and flexible class of models

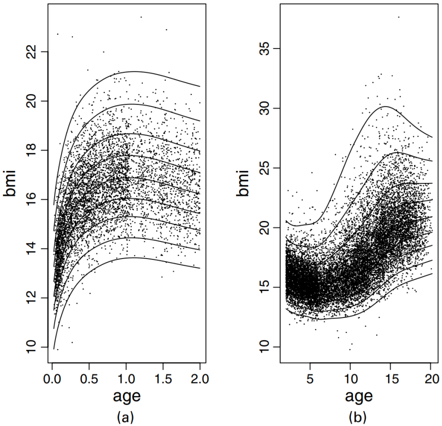

Fig. 1. Body mass index data: BMI against age with fitted centile curves

Figure: BMI vs Age with Fitted Centile Curves

Model specification

To specify a model, we need a few things:

1. A distribution with K parameters
2. K link functions $g_k(\cdot)$ for the distribution parameters

Model specification

To specify a model, we need a few things:

1. A distribution with K parameters
2. K link functions $g_k(\cdot)$ for the distribution parameters

The linear parametric GAMLSS is defined as

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k \boldsymbol{\beta}_k$$

where $\boldsymbol{\eta}_k$ are the linked predictors.

Model specification

To specify a model, we need a few things:

1. A distribution with K parameters
2. K link functions $g_k(\cdot)$ for the distribution parameters

The linear parametric GAMLSS is defined as

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k \boldsymbol{\beta}_k$$

where $\boldsymbol{\eta}_k$ are the linked predictors.
For the normal distribution $\mathcal{N}(\mu, \sigma) = \mathcal{N}(\theta_1, \theta_2)$ this looks as follows:

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \boldsymbol{X}_k \boldsymbol{\beta}_k, \\ g_1(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \boldsymbol{X}_k \boldsymbol{\beta}_k, \end{aligned}$$

usually with $g_1(\boldsymbol{\mu})$ as identity and $g_1(\boldsymbol{\sigma}) = \log(\sigma)$.

Smooth and additive effects

You can add smooth and additive effects

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J} h_{jk}(\boldsymbol{x}_{jk}) \tag{1}$$

where $h_{jk}(\boldsymbol{x}_{jk})$ is a smooth and potentially non-linear effect for the explanatory variable j for distribution parameter k.

> *The GAMLSS model (1) is more general than the GLM, GAM, GLMM or GAMM in that the distribution of the dependent variable is not limited to the exponential family and all parameters (not just the mean) are modelled in terms of both fixed and random effects. [RS05]*

Estimation

Iterative fitting approach by minimizing the penalized log-likelihood using Fisher's scoring or a Newton-Raphson scoring.

1. Score vector $\boldsymbol{u}_k = \dfrac{\partial l}{\partial \boldsymbol{\eta}_k}$
2. Working vector $\boldsymbol{z}_k = \boldsymbol{\eta}_k + \boldsymbol{W}_k^{-1} \boldsymbol{u}_k$
3. Weights $\boldsymbol{W}_k = \dfrac{\partial^2 l}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_s^T}$ resp. $\boldsymbol{W}_k = \mathrm{E}\left[\dfrac{\partial^2 l}{\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_s^T}\right]$

Repeated, weighted fit of $\boldsymbol{z}_k$ on $\boldsymbol{X}_k$ using $\boldsymbol{W}_k$.

Algorithm

Algorithm 1 RS Algorithm for Fitting GAMLSS Models

for $i = 1$ until convergence **do**

  for $k = 1$ to K **do**

    for $j = 1$ until convergence **do**

      Calculate the score vector $\boldsymbol{u}_k^{[ij]}0$

      Calculate the working vector $\boldsymbol{z}_k^{[ij]}$

      Calculate the weights $\boldsymbol{W}_k^{[ij]}$

      Do weighted fit $\boldsymbol{z}_k^{[ij]}$ on $\boldsymbol{X}_k$ to get $\beta_k^{[ij]}$

      Do back-fitting of additive and smooth effects

      Check convergence of $\boldsymbol{\theta}_k^{[ij]}$

    end for

  end for

  Check convergence of $\boldsymbol{\theta}$

end for

This has started quite something

People thought "This is cool!" and are building all kinds of extensions:

- Classic Papers [RS05; SR08]
- Regularization and LASSO [Gro+19; OB23; ZMS21]
- Bayesian [UKZ18; Uml+19]
- Boosting [HMS14; Spe+23]
- Neural Networks [Thi+24; KNS21]
- Multivariate Distributions [KK24; Mus+24; Kle+15; Str+23],
- Stochastic Gradient Descent estimation [Uml+24]
- Online Learning [HBZ24]

Have you want a review, there is the *Rage Against the Mean – A Review of Distributional Regression Approaches* [KSS23].

References I

[RS05] Robert A Rigby and D Mikis Stasinopoulos. "Generalized additive models for location, scale and shape". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 54.3 (2005), pp. 507–554.

[SR08] D Mikis Stasinopoulos and Robert A Rigby. "Generalized additive models for location scale and shape (GAMLSS) in R". In: *Journal of Statistical Software* 23 (2008), pp. 1–46.

[HMS14] Benjamin Hofner, Andreas Mayr, and Matthias Schmid. "gamboostLSS: An R package for model building and variable selection in the GAMLSS framework". In: *arXiv preprint arXiv:1407.1774* (2014).

[Kle+15] Nadja Klein et al. "Bayesian structured additive distributional regression for multivariate responses". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 64.4 (2015), pp. 569–591.

References II

[UKZ18] Nikolaus Umlauf, Nadja Klein, and Achim Zeileis. "BAMLSS: Bayesian additive models for location, scale, and shape (and beyond)". In: *Journal of Computational and Graphical Statistics* 27.3 (2018), pp. 612–627.

[Gro+19] Andreas Groll et al. "LASSO-type penalization in the framework of generalized additive models for location, scale and shape". In: *Computational Statistics & Data Analysis* 140 (2019), pp. 59–73.

[Uml+19] Nikolaus Umlauf et al. "bamlss: a Lego toolbox for flexible Bayesian regression (and beyond)". In: *arXiv preprint arXiv:1909.11784* (2019).

[KNS21] Nadja Klein, David J Nott, and Michael Stanley Smith. "Marginally calibrated deep distributional regression". In: *Journal of Computational and Graphical Statistics* 30.2 (2021), pp. 467–483.

[ZMS21] F Ziel, P Muniain, and M Stasinopoulos. "gamlss. lasso: Extra Lasso-Type Additive Terms for GAMLSS". In: *R package version* (2021), pp. 1–.

References III

[KSS23] Thomas Kneib, Alexander Silbersdorff, and Benjamin Säfken. "Rage against the mean–a review of distributional regression approaches". In: *Econometrics and Statistics* 26 (2023), pp. 99–123.

[OB23] Meadhbh O'Neill and Kevin Burke. "Variable selection using a smooth information criterion for distributional regression models". In: *Statistics and Computing* 33.3 (2023), p. 71.

[Spe+23] Jan Speller et al. "Robust gradient boosting for generalized additive models for location, scale and shape". In: *Advances in Data Analysis and Classification* (2023), pp. 1–20.

[Str+23] Annika Strömer et al. "Boosting multivariate structured additive distributional regression models". In: *Statistics in Medicine* 42.11 (2023), pp. 1779–1801.

[HBZ24] Simon Hirsch, Jonathan Berrisch, and Florian Ziel. "Online Distributional Regression". In: *arXiv preprint arXiv:2407.08750* (2024).

References IV

[KK24] Lucas Kock and Nadja Klein. "Truly multivariate structured additive distributional regression". In: *Journal of Computational and Graphical Statistics* just-accepted (2024), pp. 1–17.

[Mus+24] Thomas Muschinski et al. "Cholesky-based multivariate Gaussian regression". In: *Econometrics and Statistics* 29 (2024), pp. 261–281.

[Thi+24] Anton Frederik Thielmann et al. "Neural additive models for location scale and shape: A framework for interpretable neural regression beyond the mean". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2024, pp. 1783–1791.

[Uml+24] Nikolaus Umlauf et al. "Scalable estimation for structured additive distributional regression". In: *Journal of Computational and Graphical Statistics* (2024), pp. 1–17.