| | | |
|---|---|---|
| **Institution:** London Metropolitan University | | |
| **Unit of Assessment:** 10 Mathematical Sciences | | |
| **Title of case study:** The contribution of the Generalized Additive Models for Location, Scale and Shape (GAMLSS) to the environment and financial sector. | | |
| **Period when the underpinning research was undertaken: 2000-2020** | | |
| **Details of staff conducting the underpinning research from the submitting unit:** | | |
| **Name(s):** | **Role(s) (e.g. job title):** | **Period(s) employed by submitting HEI:** |
| Mikis D. Stasinopoulos | Principal Research Fellow<br>Reader<br>Professor in Statistics<br>Associate Lecturer<br>Professor of Mathematics | 01.09.1989 - 31.08.2007<br>01.09.2007 - 31.07.2011<br>01.08.2011 - 30.09.2016<br>08.02.2017 - 19.01-2020<br>20.01.2020 - present |
| Robert A. Rigby | Researcher<br>Associate Lecturer<br>Professor of Mathematics | 01.09.1979 - 30.09.2016<br>09.02.2017 - 19.01.2020<br>20.01.2020 - present |
| **Period when the claimed impact occurred:** 2014-2020 | | |
| **Is this case study continued from a case study submitted in 2014?** NO | | |

## 1. Summary of the impact (indicative maximum 100 words)

This impact is based upon the creation of a new statistical methodology that deepens and textures our ability to harvest information from a regression analysis. The GAMLSS methodology allows the modelling of a response variable using any theoretical distribution including highly skewed or/and kurtotic distributions and the modelling of all the parameters of the distribution.

The main beneficiaries are practitioners dealing with highly skewed and kurtotic data. Such data sets are very common in environmental and financial communities. This case study demonstrates that GAMLSS models were used in several important instances leading to better understanding and changes in policies.

## 2. Underpinning research (indicative maximum 500 words)

The generalized additive models for location, scale, and shape (GAMLSS) were introduced by Stasinopoulos and Rigby in 2005, in a Royal Statistical Society read paper (with discussion), [R1].

Prior to the development of GAMLSS methodology, response variables in a regression framework were often modelled based on the assumption of the exponential family of distributions. For continuous response variables, typically, this would be the normal, the gamma or the inverse Gaussian distribution. For discrete data this would be the Poisson or the binomial distribution. However, many response variables are either heavy tailed or/and highly skewed. For example, high frequency data in econometric or finance, river flows, web traffic data and text classification data. In GAMLSS, the distribution of the response variable can be continuous, discrete or mixed. A mixed distribution is a continuous distribution with discrete parts such as, for example, a zero

inflated gamma distribution. GAMLSS also allows all the parameters of the distribution to be modelled as linear or non-linear smooth functions of the exploratory variables. Those parameters are often location, scale and shape parameters. Since all parameters are modelled, the distribution of the response variable can vary in shape according to explanatory variables. This results in more realistic models in which both rare events as well as the centre of the distribution can be modelled, accurately. This leads to better understanding of how such incidences occurred and and therefore can be used to inform more effective policies which help to mitigate risks.

GAMLSS solves problems inherited from the stricter assumptions of the popular GLM and the GAM regression models: i) heterogeneity in variance, ii) positive or negative skewness, iii) `platy' or `lepto' kurtosis, iv) over or under dispersion, v) excess or shortage of zeros and vi) variance-mean relationships. Case (iii) above is closely related to the theory of "heavy tails", common in financial and environmental data where `rare' events do occur more often. Cases (iv) to (v) are common to discrete response variables. The monograph [R4] is a testimony of the commitment of Stasinopoulos and Rigby in developing appropriate theoretical distributions to solve real problems. The monograph contains details of more than 100 distributions. It also shows how extra distributions can be generated and applied to real data examples.

London Metropolitan University is the main hub for the current and future development of GAMLSS. All co-authors of the GAMLSS books are visiting researchers and contributed to the work of the Statistics Operation Research and Mathematics centre of London Met. Dr Voudouris, is chief data officer for Argus media, Dr De Bastiani is a lecturer at the Federal University of Pernambuco, Brazil, and Professor Heller, is working at the Clinical Trials Centre of University of Sydney.

## 3. References to the research (indicative maximum of six references)


R1. Rigby, R. A. and Stasinopoulos, D. M. (2005) Generalized additive models for location, scale and shape. (with discussion). *Applied Statistics*, 54, 507-554. DOI: https://doi.org/10.1111/j.1467-9876.2005.00510.x

R2. Stasinopoulos, D. M. and Rigby, R. A. (2007) Generalized additive models for location, scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23, 1-46. DOI:https://doi.org/10.18637/jss.v023.i07

R3. Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V. and De Bastiani, F., (2017) *Flexible Regression and Smoothing: Using GAMLSS in R*, Chapman & Hall/CRC, Boca Raton. DOI https://doi.org/10.1201/b21973-4

R4. Stasinopoulos, D. M., Rigby, R. A. and De Bastiani, F., (2018), GAMLSS: a distributional regression approach, *Statistical Modelling*, 18, 248-273.https://doi.org/10.1177/1471082x18759144

R5. Rigby R. A., Stasinopoulos D. M., Heller G., and De Bastiani. F. (2019). *Distributions for modelling location scale and shape: Using GAMLSS in R.* Chapman & Hall/CRC, Boca Raton. https://doi.org/10.1201/9780429298547-3


## 4. Details of the impact (indicative maximum 750 words)

The benefits of a more realistic modelling of phenomena, by GAMLSS, leads to a better understanding of how those incidences occurred and therefore supports more effective policies

which help to mitigate risk. The beneficiaries have been environmental, economic or regulatory agents, wishing to understand and prevent potentially damaging events.

**Financial Sector risk modelling**

In 2015, the International Monetary Fund (IMF) used GAMLSS for a stress testing exercise of the US economy [S1]. A key mission of the IMF is to periodically conduct assessments under the Financial Services Action Program (FSAP). "*These assessments involve a considerable amount of data analysis and quantitative modelling. It is in this context that the GAMLSS work has been extremely relevant to us.*" GAMLSS was found to be highly useful in the U.S. FSAP context, because it is "*an extremely flexible model class that allowed* [them] *to: (i) utilize a wide variety of distributions to characterize the response variable and (ii) explicitly model the first four moments of these distributions as functions of exogenous conditions. As a result,* [they found] *the framework well suited to address the presence of tail-risks, nonlinearities, and deviations from the normality assumption. The default probability data used in the stress testing exercise exhibited all of these characteristics, thereby motivating the choice of the GAMLSS model class for stress testing purposes."* [S1 p86]

The GAMLSS model for the U.S. financial system suggested "*that macroeconomic, sector credit risk, and interconnectivity influence credit risk in the United States in significant—and often nonlinear-ways.*" [S1 p88] The GAMLSS results underscored *"the importance of spillovers across sectors*" and across borders. [S1 p91] The approach usefully complemented other risk analysis used in the 2015 U.S. FSAP. The analysis yielded specific findings, suggestions and **recommendations for improvement** including conducting "*regular, comprehensive stress tests for all major financial sub-sectors that capture the impact of macro-financial factors, spillover and feedback effects between institutions and interactions between solvency and liquidity*" and "*more intensive monitoring of systemic financial sector risks, including the use of market-based solvency and shortfall measures."* [S1 p10].

Speaking of his "*deep appreciation for the world class research by Professors Dimitrios Stasinopoulos and Robert Rigby, in developing*" GAMLSS, the Division Chief at the IMF observed that the "*work has yielded an important positive contribution to our activities at the International Monetary Fund*," noting that having worked at the IMF since 2000 and "*with many models and analytical tools, but the GAMLSS truly stands thanks to its combination of innovativeness, analytical rigor, and applicability."*[S2] In the context of the stress testing exercise of the US economy, he reports that the *"use of GAMLSS helped to change our understanding of the vulnerability"* of the US financial sector.

Summarising the contribution of GAMLSS to the IMF, the Division Chief states "*the innovative research by Prof. Stasinopoulos and Prof. Rigby has been useful for the IMF's work on assessing financial sector risk in the U.S. The analysis yielded specific findings and suggestions that were discussed in detail with our technical level counterparts during the FSAP."* These informed their "*higher level policy recommendations* [which were] *discussed with top policy makers at the U.S. Treasury, Federal Reserve, other key agencies as well as with other stakeholders."[*S2].

The benefits of GAMLSS are demonstrated by its widespread usage by statisticians in non-academic settings, particularly within the financial regulatory sector:
i) The **European parliament** report, [S3], aimed to provide a quantitative assessment of the macro-economic cost of a possible banking crisis within the Euro zone.

ii) The **Bank of England**, [S4], to estimate the UK mortgage market vulnerabilities from load-level data.
iii) The **Standard Chartered Bank** (SCB), [S5] to "*provide a macroeconomic balance sheet projections and corporate planning*". The SCB employs more than 86,000 people and has a presence in 60 countries and territories in Europe and Americas, Asia, the Middle East and Africa.
iv) The **Bank of America**, (BoA) "*to develop credit risk models*" specifically to benchmark "*loss given default (LGD) and exposure at default*" (EAD) [S6]. BoA has asset size of USD2.2 tn (Sep 2020), 2nd largest bank in the US, 8th largest in the world.

The team leader of Quantitative Modelling and Analytics at the SCB, reports that *"the GAMLSS toolbox is particularly appealing to us for its transparency, range of various smoothers, choice of distributions and ability to address the presence of tail-risks, nonlinearities, and deviations from the normality assumptions*". He notes that the GAMLSS *"models provide baseline view of the bank's key product performance under various macroeconomic scenarios. This information is used by the top management to plan business development more efficiently into the future and spot new business opportunities".* [S5]

**Environment and ecosystem**

Southern Water's 2019 Draft Water Resources Management Plan describes how they developed their artificial weather generator for simulating rainfall for water resource model using GAMLSS because "*these models allow the incorporation of the large-scale climate indices (NAO, SST) and determine their relation to single rain gauge sites or gridded rainfall data".* [S7] The resulting model "*is now fully parametric and predicts rainfall for all seasons directly. This has thus removed issues relating to lack of persistence which were associated bootstrap sampling of spring-autumnal rainfall*" in the earlier model. [S7 p27]. The new model not only "*produces spatially and temporally coherent monthly rainfall time series*," as a result of using GAMLSS, it also reproduces extreme rainfall patterns which is "*unique to this model and has not been attempted or demonstrated elsewhere*."[S7 p28]

Globally, seagrass meadows are one of the planet's significant carbon absorption assets as well as being nursery habitats for fish and invertebrates, important food source for marine life and helping to stabilise and maintain the seabed. The Great Barrier Reef Marine Park Authority's Marine Monitoring Program has used GAMLSS to build models to monitor the health and condition of seagrass in the Great Barrier Reef. The resulting report notes that the GAMLSS models, used to "*determine whether seagrass data collected*" could "*predict subsequent seagrass coverage*" concluded that while "*the high level of uncertainty associated with the reproductive effort index* [hindered] *the value of this indicator for assessing ecosystem health,*" the results supported "*the assertion that species diversity and productivity are good indicators of resilience*." [S9] The research statistician with CSIRO advised that it is "*essential that the condition of seagrass is measured, monitored and reported accurately*." Their models predicting seagrass abundance and reproductive effort [S9] are being used to "*inform and refine future seagrass data collection and reporting in the Great Barrier Reef to ensure that this essential part of the ecosystem is monitored appropriately and management actions*" undertaken.[S8]

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

S1. International Monetary Fund Country Report No. 15/173 "UNITED STATES FINANCIAL

SECTOR ASSESSMENT PROGRAM STRESS TESTING—TECHNICAL NOTE" July 2015, https://www.imf.org/external/pubs/ft/scr/2015/cr15173.pdf.

S2. Testimonial letter of Division Chief, International Monetary Fund (25 November 2020).

S3. European Parliamentary Research Service "Making the European Banking Union Macro-Economically Resilient - Cost of Non-Europe Report" June 2015, https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2015)558771

S4. Bank of England Staff Working Paper No. 703 "A tiger by the tail: estimating the UK mortgage market vulnerabilities from loan-level data" December 2017, https://www.bankofengland.co.uk/working-paper/2017/a-tiger-by-the-tail

S5. Testimonial letter Team Lead – Quantitative Modelling & Analytics, Standard Chartered Bank (3 February 2021).

S6. Testimonial letter of Dr Edward Tong, working as vice president for model risk management at Bank of America in New York, from 2015 to 2018, (14 July 2020).

S7. Southern Water "Draft Water Resources Management Plan 2019 Annex 3: Supply Forecast February 26 2018, version 2", http://www.hwa.uk.com/site/wp-content/uploads/2017/12/dWRMP19_SWS_Annex-3-Supply-forecast.pdf

S8. Testimonial letter from a Research Statistician at CSIRO Data61 (3 February 2021).

S9. Lawrence, E., Gladish, D. 2019, Assessment of reproductive effort as an indicator of seagrass health for the Marine Monitoring Program. http://elibrary.gbrmpa.gov.au/jspui/handle/11017/3539