Economics Bulletin

Volume 43, Issue 2

Non-linearities in the R&D-firm growth relationship: Evidence from a semiparametric location-scale regression approach

Drini Morina
Gottfried Wilhelm Leibniz University Hannover

Henning Lucas
Gottfried Wilhelm Leibniz University Hannover

Stefanie Heiden
Gottfried Wilhelm Leibniz University Hannover

Abstract

Based on a generalized additive model for location and scale, this paper shows that there is no positive association between R&D and average firm growth rates in the short run, unless R&D efforts are extremely high. The growth rate variance, however, increases with smaller values of R&D intensity and the magnitude is more extreme.

Citation: Drini Morina and Henning Lucas and Stefanie Heiden, (2023) "Non-linearities in the R&D-firm growth relationship: Evidence from a semiparametric location-scale regression approach", *Economics Bulletin*, Volume 43, Issue 2, pages 1155-1161
Contact: Drini Morina - drini.morina@ite.uni-hannover.de, Henning Lucas - henning.lucas@ite.uni-hannover.de, Stefanie Heiden - stefanie.heiden@ite.uni-hannover.de.
Submitted: October 13, 2022. **Published:** June 30, 2023.

1. Introduction

Most standard economic models consider research and development (R&D) as driver of growth (e.g. Romer, 1990) and firms conduct risky R&D in hope of superior performance. Especially the short-term relation between R&D and firm growth achieved large attention in the literature. Previous research (e.g. Coad & Rao, 2010) show empirical evidence of a positive, linear relationship between R&D and the *variance* of growth rates (which is a common measure of R&D risk). However, until now such a positive relationship concerning the *mean* sales growth rate was not found; instead a positive, linear link could only be revealed for very few successful innovators.[1] Hence, it remains questionable why short-term oriented firms would aim to be innovative after all? This question seems of substantial interest for shareholders, managers, employees and economists.

Considering the high number of observations available in the conventionally used data sources (such as in the Standard & Poor's Compustat Database employed in this work), it is certainly possible to deduce a more complex effect structure, moving away from linear empirical models.[2] In the R&D-firm growth literature, first attempts in this direction were made by the introduction of quadratic or cubic terms in simple linear regression models (and corresponding tests for significance; see e.g. Tsai, 2005 or Bianchini et al. 2018). But it remained doubtful whether the relationship to sales growth is equal throughout the whole range of R&D activity.

The present work is – to our knowledge – the first attempt to model this association based on a generalized additive model (GAM; Hastie and Tibshirani, 1990) using spline-based smooth functions to estimate the functional form of the relationship (Wood, 2003). The variance of sales growth is modeled as second dependent variable next to the mean, allowing for non-linearity of any form and aiming to disentangle the relations.

We find that R&D is *on average* indeed beneficial *after one year* only if R&D efforts are extremely high. The link is increasingly positive for values roughly above R&D-sales-ratios of 100%. This might arise from the necessity to build knowledge stocks over time, which can be accelerated by heavy R&D investments. Our main finding, however, is that high R&D intensities (above 13% in terms of R&D-sales-ratio) are mostly associated with larger growth rate *variance*, i.e. with more potential to grow or decline. The association we find is in summary: Low R&D efforts are on average not risky but not beneficial after one year, whereas high R&D efforts are beneficial and very risky. These findings are consistent across all sub-industries in the manufacturing sector, but turning points may vary.

2. Data

The widely-used Standard & Poor's Compustat Industrial Database is the basis of the following GAM analysis. The focus lies on US listed companies operating in manufacturing industries (Standard Industrial Classification, SIC, classes 2000–3999) in the period between 2000 and 2020. The reason for this focus is to avoid sector aggregation issues (Botazzi & Secchi, 2003),

[1] Quantile models reveal a clearly positive relationship only for the fastest-growing firms (e.g. Coad & Rao, 2008; Hölzl, 2009; Segerra & Teruel, 2014; Bianchini et al., 2018; Guarascio & Tamagni, 2019 among many others). Innovation efforts of firms at lower quantiles are, by contrast, often found to be negatively related to sales growth at common significance levels (e.g. Coad & Rao, 2008; Guarascio & Tamagni, 2019).

[2] To employ semiparametric regression, the statistics literature suggests varying minimum sample sizes (e.g. Kain et al. 2015 or Karatekin et al. 2019), which are all exceeded by the sample analyzed in the present work.

but also for the sake of comparability with previous studies.[3] The resulting unbalanced panel contains 25,191 firm-year observations of 2,314 firms with a minimum of 5 consecutive years of data on the variables of interest. Outliers are not removed since they are substantive to the observed dispersion. Summary statistics of the sample are reported in table 1.

Growth rates, $G_{i,t}$, are calculated by differences of natural logarithms of annually total sales between two years t and $t-1$, i.e. $G_{i,t} = \log(sales_{i,t}) - \log(sales_{i,t-1})$. R&D expenditure is measured by annual costs that are related to new product development (or new service development), which is corrected by bought-in in-process R&D. R&D intensity is current R&D expenses divided by total sales. The corresponding Compustat variables are *SALE* and *XRD + RDIP*. Figure 1 shows histograms of these variables, where 3,097 (12.3%) of the firm-year observations reveal R&D intensities above 100%.

SIC Code	Num. obs.		Mean	SD	0.25th Q	Median	0.75th Q
20	607	Sales	9,583.50	16,594.46	305.10	3,736.80	8,889.10
		RDI	0.05	0.24	0.00	0.01	0.01
21	61	Sales	20,768.90	18,230.92	8,304.00	17,663.00	28,694.00
		RDI	0.01	0.00	0.01	0.01	0.01
22	84	Sales	740.50	336.21	622.93	797.27	988.11
		RDI	0.02	0.01	0.01	0.01	0.02
24	81	Sales	3,062.32	5,655.92	116.86	440.80	2,176.68
		RDI	0.03	0.06	0.00	0.01	0.01
25	261	Sales	2,757.63	4,231.48	807.96	1,617.30	2,486.60
		RDI	0.02	0.02	0.01	0.01	0.02
26	315	Sales	6,477.23	8,636.98	770.71	2,758.33	6,808.20
		RDI	0.01	0.02	0.00	0.01	0.02
27	86	Sales	1,289.00	1,289.42	254.03	909.12	1,741.18
		RDI	0.06	0.11	0.01	0.01	0.04
28	6954	Sales	1,957.75	7,448.59	5.92	41.14	499.38
		RDI	18.06	201.87	0.04	0.33	2.40
29	149	Sales	75,579.99	112,747.01	867.52	4,040.80	142,897.00
		RDI	0.16	0.56	0.00	0.01	0.03
30	392	Sales	2,139.59	4,096.78	86.19	546.74	2,370.21
		RDI	0.18	2.62	0.01	0.01	0.03
32	218	Sales	1,936.55	2,455.64	80.34	628.36	3,511.03
		RDI	0.03	0.06	0.01	0.01	0.02
33	318	Sales	1,999.02	2,424.14	511.69	1,169.84	2,343.47
		RDI	0.16	1.10	0.00	0.01	0.01
34	658	Sales	1,908.34	2,834.10	203.81	827.10	2,145.35
		RDI	0.02	0.07	0.01	0.01	0.02
35	3140	Sales	3,355.91	10,737.60	81.20	457.74	1,900.57
		RDI	0.15	1.86	0.02	0.05	0.14
36	5148	Sales	1,973.64	12,061.03	35.75	192.63	741.82
		RDI	0.59	8.63	0.04	0.11	0.21
37	1407	Sales	10,367.96	28,565.91	388.54	1,436.77	5,983.34
		RDI	0.53	11.10	0.01	0.03	0.04
38	4848	Sales	1,175.02	3,853.49	20.41	82.46	509.39
		RDI	0.97	29.61	0.05	0.09	0.18
39	386	Sales	1,289.29	2,192.57	95.88	418.95	1,089.12
		RDI	0.06	0.12	0.02	0.04	0.06

Table 1: Summary statistics for different SIC-coded industries.

[3] Especially to be comparable with the study of Coad & Rao (2010), since it is the most explicit, empirical study claiming a strictly positive, linear relationship between growth rate variance and R&D focusing on the SIC classes 2000–3999.

2. Model and Methods

Figure 1 shows log-transformed R&D intensity against sales growth rates. The relationship seems non-linear, the variance appears to increase from low intensities onwards, but the mean appears to be only greater for very high R&D intensities.

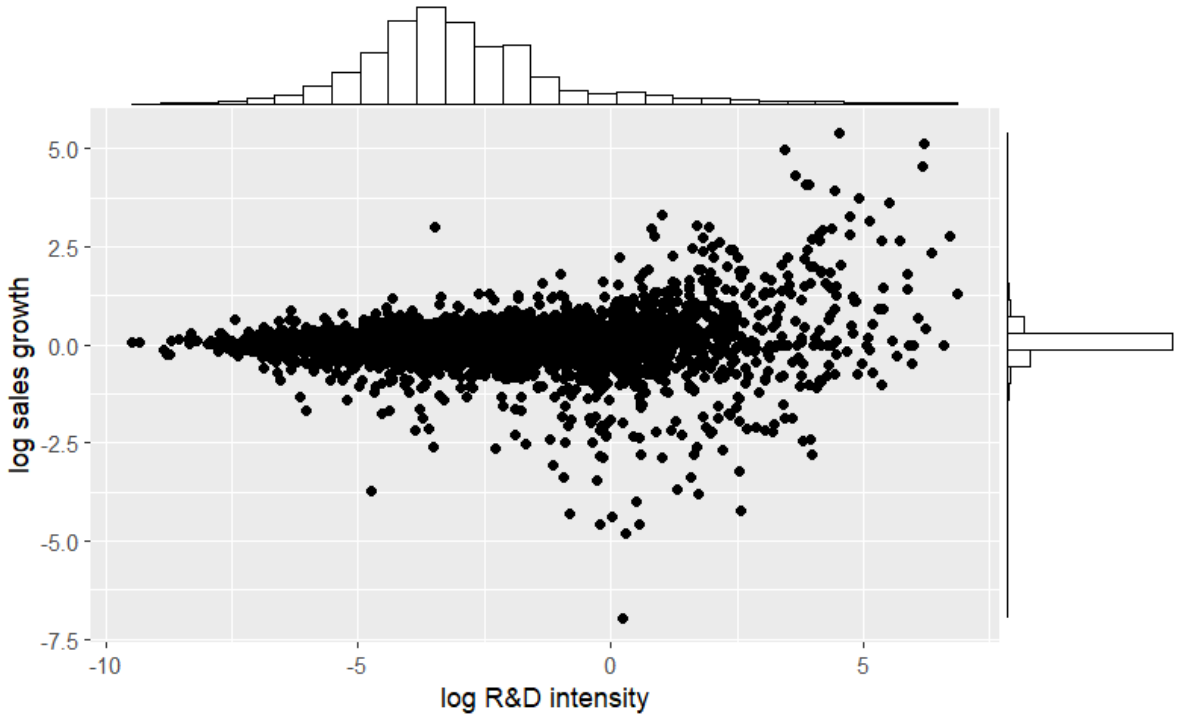

Figure 1: R&D intensities against sales growth rates, both log-transformed

The generalized additive model used to analyze the short-term relationship (without any causal claim, see far below in this section) is:

$$h(G_{i,t}) = \beta_0 + f(RDI_{i,t-1}) + \beta_1 G_{i,t-1} + \beta_2 X_{i,t-1} + \beta_3 IND_i + \beta_4 YEAR_t + u_{i,t},$$

where i and t indicate the firms and years respectively, $G_{i,t}$ and $RDI_{i,t}$ are defined as above, f is a *penalized thin plate regression spline* (Wood, 2003), used to model the functional form of the fitted predictor, h is a link function and $X_{i,t}$ denotes usual control variables for firm heterogeneity (such as age and numbers of employees, which have significantly positive and negative relations to growth; corresponding Compustat variables are *IPODATE* and *EMP*).

The model also accounts for specialities of sectors and macroeconomic shocks using the fixed effects IND_i and $YEAR_t$, respectively (which are partially significant; Compustat names are *SIC* and *FYEAR*). To additionally control for growth rate autocorrelation, the model includes the lagged growth rate as a covariate (in line with Bottazzi and Secchi 2003; the relation to the lagged growth rate is positive and significant). As discussed in the recent literature, one important question is whether to also account for firm-specific effects aiming to further exploit the panel data structure. Earlier research shows that this effect is generally negligible since growth rates are hardly predictable by any variable at all and most appropriately approximated by a random walk (see Gibrat, 1931; Geroski, 2000; and more recently e.g. Van Witteloostuijn & Kolkman, 2019). In figure 2, growth rates are plotted over time and grey lines connect the

growth rates of each single firm. Figure 2 shows no noteworthy firm-specific pattern, indicating that, in line with early studies, firm-specific effects can indeed be overlooked.

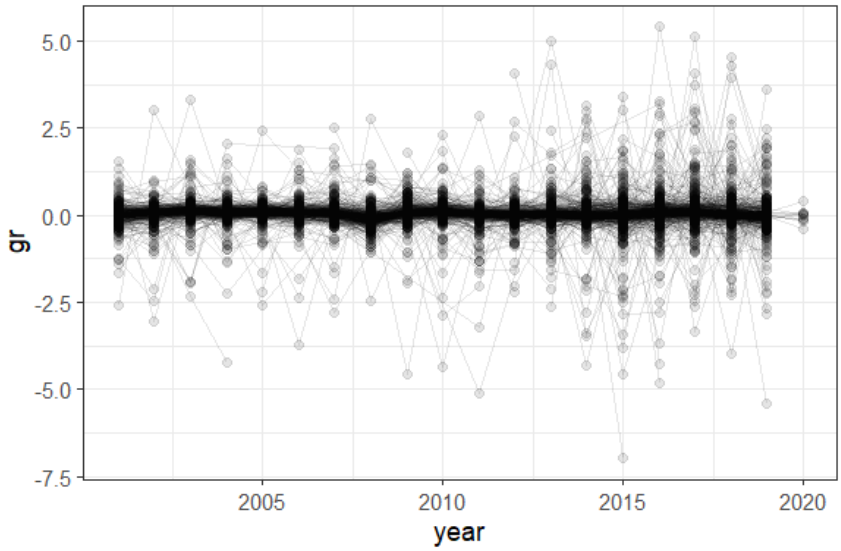

Figure 2: Growth rates (gr) over years, connected with grey lines for each firm

Another argument is that specific effects for more than 700 firms lead to heavy computational complexity, which is unfavorable in view of the complex and computationally intensive spline-based fitting procedure applied in the analysis. The main equation shown above is not estimable with (random or fixed) firm-specific effects due to the high number of firms and low number of data points in comparison. Hence in the following analysis, the empirical model does not include firm-specific effects, first and foremost with aim to ensure comparability with most of the body of literature, such as e.g. Coad & Rao (2010), Hölzl (2009), Falk (2010), Mazucatto & Demirel (2012), Segerra & Teruel (2014) Guarascio & Tamagni (2019). However, despite of these arguments a second model is fitted using a wider sample with the same design but focusing on the high-tech sector and additional firm-specific effects μ_i, especially to be in line with some very recent studies that consider those effects likewise, such as e.g. Calvino (2020), Bachianni et al. (2018) and Coad et al. (2021). The model with additional firm-specific dummies reveals very similar results to those presented in this work, additionally indicating that an omitted variable bias may not be present.

Hence, the following analysis is based on a cross-sectional GAM for the mean and variance using a Gaussian location-scale model with an identity-logb-link function and Generalized Cross Validation (GCV; Craven and Wahba, 1979; Wahba, 1990). Penalized thin plate regression splines are used to model the partial effect function of R&D intensity due to their superior performance compared to other conventional smoothers (Wood et al. 2016). The smoothing parameters which enforce sparsity or smoothness are chosen by GCV automatically, but to avoid overfitting, the dimension of the basis of smooth terms is limited to be 5.[4] The analysis is mainly focused on the "shape" of the partial effect function of R&D intensity rather than on numerical outcomes.

However, another potential endogeneity issue in this context is possible reversed causality. Even though the growth rate in t is modeled by lagged variables, such as R&D intensity in $t-1$, Bellemare et al. (2015) show that reversed causality cannot be excluded reliably in *any* case,

[4] The R software (R Development Core Team, 2020) package "mgcViz" (Wood et al. 2018) is used for fitting.

even though it is rather unlikely compared to a contemporaneous modeling approach.[5] Hence, the following section interprets all modeled relationships as *associations* rather than causal links.

4. Results and Discussion

The non-linearity of the association between R&D intensity and growth rates is highly significant; the fit was modeled with 3.98 estimated effective degrees of freedom (EDFs) for the mean, 3.99 estimated EDFs for the variance and p-values of smooth components (based on a Wald-type test; see Wood, 2013) resulted to be lower than 2e-16. The deviance explained by the model is 9.52%. Figure 3 illustrates the functional form of the partial effect of R&D intensity on the mean at the left-hand side and the effect on the variance at the right-hand side. The grey areas display 95% confidence intervals around the mean shape of the effect.

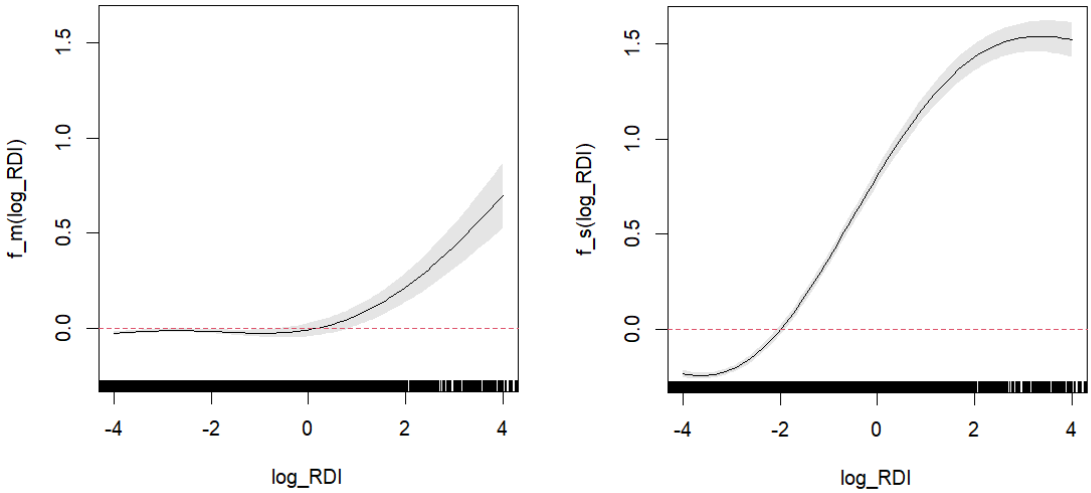

Figure 3 (in color): Partial effect plots of log-transformed R&D intensity on the mean (f_m) and variance (f_s) of growth rates with shaded confidence intervals.

The results of the GAM reveal an increasing positive association for the mean and the variance only from a certain value of R&D intensity onwards. The exact values depend on the setting and hyper parameters, but they are approximately -2 for the variance and slightly above 0 for the mean (which are about 0.13 and above 1 on the R&D-sales-ratio scale, respectively). A fitted *linear* model, however, would suggest an overall significantly positive (average) effect on both the mean and the variance of growth rates (as in many previous studies, e.g. Del Monte & Papagni (2003) concerning the mean or Coad & Rao (2010) concerning the variance). Estimation results for the linear effects of R&D intensity are displayed in table 2.

GAM without smooth terms		
Dependent variable	Mean	Variance
Coefficients of $RDI_{i-1,t}$	0.0004 ***	0.0210 ***
(Standard errors)	(0.001)	(0.002)
Deviance explained		0.707
AIC		17524.85

*Table 2: Summary of a linear GAM estimation ('***' denote significance at the 1% level)*

[5] Although the panel approach with firm-fixed effects can for the most part account for endogeneity arising from unobserved firm heterogeneity, it can yet not secure unidirectional causality.

The deviance explained by the linear GAM is 9.15% and the AIC of the non-linear model is by 10,640.6 smaller than the AIC of the linear GAM[6]; that is, the smooth GAM is preferable and indicates that OLS estimation results are dominated by the outstanding effect of highly R&D-intensive firms.

We summarize our findings as follows: On average, R&D is not generally beneficial after one year. This is consistent with parts of previous literature: "Need it be reminded, an innovation strategy is even more uncertain than playing a lottery" (Coad & Rao, 2008). On the contrary, we find that a firm might benefit on average if R&D intensities are extremely high (roughly above 100%). This might arise from the necessity to build knowledge stocks over time, which can be accelerated by heavy R&D investments in the short run. Our main finding, however, is that R&D mostly links with growth rate variance at high R&D intensities (above 13%)

5. Conclusion

It has been shown that applying a GAM for location and scale on a large manufacturing sector dataset over the period 2000–2020 would overthrow the hypothesis that R&D investments have generally positive (or negative) links with mean growth rates or growth rate variance: The relation strongly depends on the level of R&D intensity.

However, the causality in this observed relationship may not surely be unidirectional. It is possible that extensive R&D efforts will pay off one year after investing but it is reversely possible that firms expecting extraordinary growth spend overproportionate efforts in R&D. Following the explanation in Coad & Rao (2010) for the variance relation, one might likewise argue that heavy investments in R&D may increase the growth rate variance, since R&D is very uncertain, but it is on the other hand possible that firms in turbulent situations may devote extensive efforts to risky R&D strategies. Future research on the causality would be welcome. Further work could also aim to throw light on (time-dependent) structural implications. More specified (and advanced) models could ensure definite numerical values for the relations shown in this paper, especially in view of practical suggestions.

References

Bianchini, S., Pellegrino, G., Tamagni, F. (2018). Innovation complementarities and firm growth. Industrial and Corporate Change 27 (4): 657-676

Calvino, F. (2020). Technological innovation and the distribution of employment growth: A firm-level analysis. Industrial and Corporate Change 28(1):177-202

Coad, A., Rao, R. (2008). Innovation and Firm Growth in High-Tech Sectors: A Quantile Regression Approach. Research Policy 37(4): 633-648

Coad, A., Rao, R. (2010). R&D and firm growth rate variance. Economics Bulletin 30: 702-708

Coad, A., Segarra-Blasco, A., Teruel, M. (2021). A bit of basic, a bit of applied? R&D strategies and firm performance. The Journal of Technology Transfer 46(6):1-26

[6] Separated OLS (or MAD) regression, as applied by e.g. Coad & Rao (2010), which are not estimated in a Gaussian location-scale framework, reveal even inferior goodness-of-fits.

Craven, P., Wahba, G. (1979). Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. Numerische Mathematik. 31.

Falk, M. (2010). Quantile estimates of the impact of R&D intensity on firm performance. Small Business Economics 35 (2): 1-19

Fasiolo, M., Goude, Y., Nedellec, R., Wood, S. (2017). Fast Calibrated Additive Quantile Regression. Journal of the American Statistical Association 116 (535)

Guarascio, D., Tamagni, F. (2019). Persistence of innovation and patterns of firm growth. Research Policy 48 (6): 1493-1512

Hagedoorn, J., Cloodt, M. (2003). Measuring Innovative Performance: Is There an Advantage in Using Multiple Indicators? Research Policy 32 (8): 1365-1379

Hastie, T.J., Tibshirani, R.J. (2017). Generalized Additive Models. 10.1201/9780203753781.

Hölzl, W. (2009). Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries. Small Business Economics 33 (1): 59-75

Kaiser, U. (2009). Patents and profit rates. Economics Letters 104 (2): 79-80

Kile, C., Phillips, M. (2009). Using Industry Classification Codes to Sample High-Technology Firms: Analysis and Recommendations. Journal of Accounting, Auditing & Finance. 24. 10.1177/0148558X0902400104

Mazzucato, M., Demirel, P. (2012). Innovation and Firm Growth: Is R&D Worth It? Industry and Innovation 19 (1)

Monte, A., Papagni, E. (2003). R&D and the growth of firms: Empirical analysis of a panel of Italian firms. Research Policy 32 (6): 1003-1014

Moschella, D., Tamagni, F., Yu, X. (2019). Persistent high-growth firms in China's manufacturing. Small Business Economics 52 (3): 573-594

Romer, P. (1990). Endogenous technological change. J. Polit. Econ. 98 (5), S71–102

Segerra, A., Teruel, M. (2014). High-Growth Firms and Innovation: an empirical analysis. Small Business Economics 43 (4)

Tsai, K. (2005). R&D productivity and firm size: A nonlinear examination. Technovation 25 (7): 795-803

Wahba, G. (1990). Spline Models for Observational Data. SIAM, Philadelphia

Wood, S. (2003). Thin Plate Regression Splines. Journal of the Royal Statistical Society Series B. 65. 95-114.

Wood, S. (2013). On p-values for smooth components of an extended generalized additive model. Biometrika. 1. 10.1093/biomet/ass048.

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

Wood S. N., Fasiolo M., Nedellec R., Goude Y. (2018). Scalable visualisation methods for modern Generalized Additive Models. arXiv:1809.10632.