

Generalized Additive Model

Test name: Generalized Additive Model

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Generalized Additive Model\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in GAM model: 649

Outcome variable: G3

Smooth predictors: G1, G2, absences

Linear predictors: studytime, failures, traveltime

Categorical predictors: school, sex, address

GAM formula: $G3 \sim \text{bs}(G1, \text{df}=5, \text{degree}=3, \text{include_intercept}=\text{False}) + \text{bs}(G2, \text{df}=5, \text{degree}=3, \text{include_intercept}=\text{False}) + \text{bs}(\text{absences}, \text{df}=5, \text{degree}=3, \text{include_intercept}=\text{False}) + \text{studytime} + \text{failures} + \text{traveltime} + \text{C}(\text{school}) + \text{C}(\text{sex}) + \text{C}(\text{address})$

Linear comparison formula: $G3 \sim G1 + G2 + \text{absences} + \text{studytime} + \text{failures} + \text{traveltime} + \text{C}(\text{school}) + \text{C}(\text{sex}) + \text{C}(\text{address})$

GAM adjusted R-squared: 0.8544

Linear adjusted R-squared: 0.8511

GAM RMSE: 1.2126

Linear RMSE: 1.2368

The GAM spline additive model has lower fitted RMSE than the standard linear model.

Interpretation: spline terms allow nonlinear additive relationships between predictors and the outcome.

Decision guide: compare adjusted R-squared, RMSE, AIC and residual charts before reporting whether the GAM improved fit.

Reporting guide: state outcome, smooth terms, categorical controls, model fit, residual diagnostics and key smooth-effect patterns.

Charts generated: 8

GAM smooth terms

smooth_predictor	spline_df	basis	interpretation
G1	5.0000	cubic B-spline	nonlinear additive smooth term
G2	5.0000	cubic B-spline	nonlinear additive smooth term
absences	5.0000	cubic B-spline	nonlinear additive smooth term

GAM and linear fit statistics

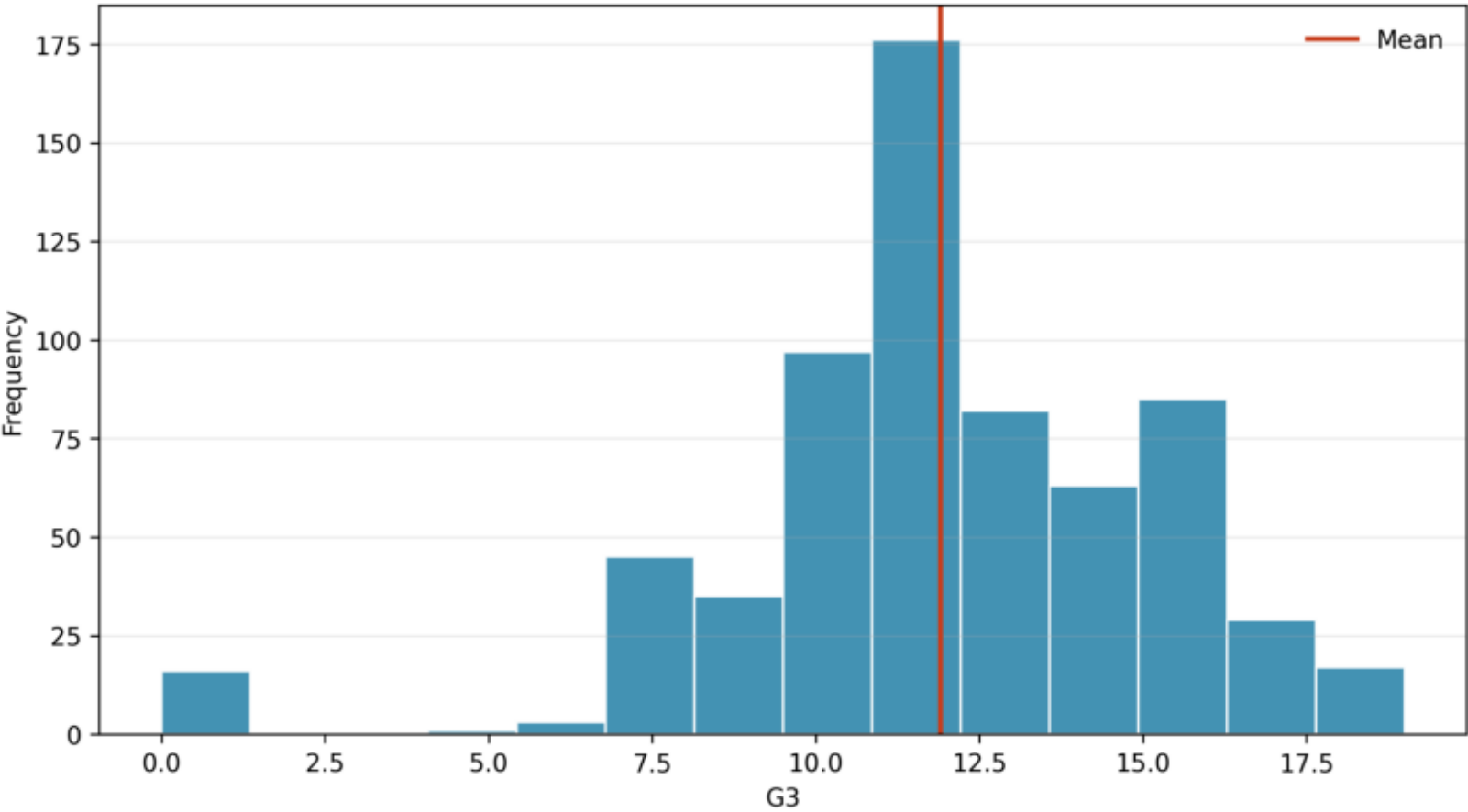
model	n	df_model	df_resid	r_squared	adjusted_r_squared	aic	bic	rmse	mae
GAM spline add	649.0000	20.0000	628.0000	0.8589	0.8544	2133.9981	2227.9822	1.2126	0.7639
Standard linear	649.0000	9.0000	639.0000	0.8532	0.8511	2137.6467	2182.4011	1.2368	0.7792

GAM coefficients

model	term	coefficient	standard_error	t_value	p_value	ci95_lower	ci95_upper	decision_alpha_0.05
GAM spline additive	Intercept	0.5163	1.0965	0.4708	0.6379	-1.6370	2.6696	Not statistically significant
GAM spline additive	C(school)[T.MS]	-0.1079	0.1221	-0.8833	0.3774	-0.3476	0.1319	Not statistically significant
GAM spline additive	C(sex)[T.M]	-0.2233	0.1027	-2.1740	0.0301	-0.4250	-0.0216	Statistically significant
GAM spline additive	C(address)[T.U]	0.1668	0.1180	1.4138	0.1579	-0.0649	0.3986	Not statistically significant
GAM spline additive	bs(G1, df=5, degrees of freedom)	-2.3537	1.7321	-1.3588	0.1747	-5.7552	1.0478	Not statistically significant
GAM spline additive	bs(G1, df=5, degrees of freedom)	0.5492	1.2126	0.4529	0.6507	-1.8319	2.9304	Not statistically significant
GAM spline additive	bs(G1, df=5, degrees of freedom)	0.8105	1.2887	0.6289	0.5296	-1.7202	3.3411	Not statistically significant
GAM spline additive	bs(G1, df=5, degrees of freedom)	1.7058	1.3441	1.2691	0.2049	-0.9337	4.3452	Not statistically significant
GAM spline additive	bs(G1, df=5, degrees of freedom)	1.9058	1.5725	1.2120	0.2260	-1.1822	4.9938	Not statistically significant
GAM spline additive	bs(G2, df=5, degrees of freedom)	2.6803	1.0890	2.4612	0.0141	0.5417	4.8189	Statistically significant
GAM spline additive	bs(G2, df=5, degrees of freedom)	7.2835	0.6757	10.7793	0.0000	5.9566	8.6104	Statistically significant
GAM spline additive	bs(G2, df=5, degrees of freedom)	12.3530	0.6557	18.8383	0.0000	11.0653	13.6407	Statistically significant
GAM spline additive	bs(G2, df=5, degrees of freedom)	15.4141	0.8004	19.2579	0.0000	13.8423	16.9859	Statistically significant
GAM spline additive	bs(G2, df=5, degrees of freedom)	15.5358	1.0342	15.0227	0.0000	13.5050	17.5667	Statistically significant

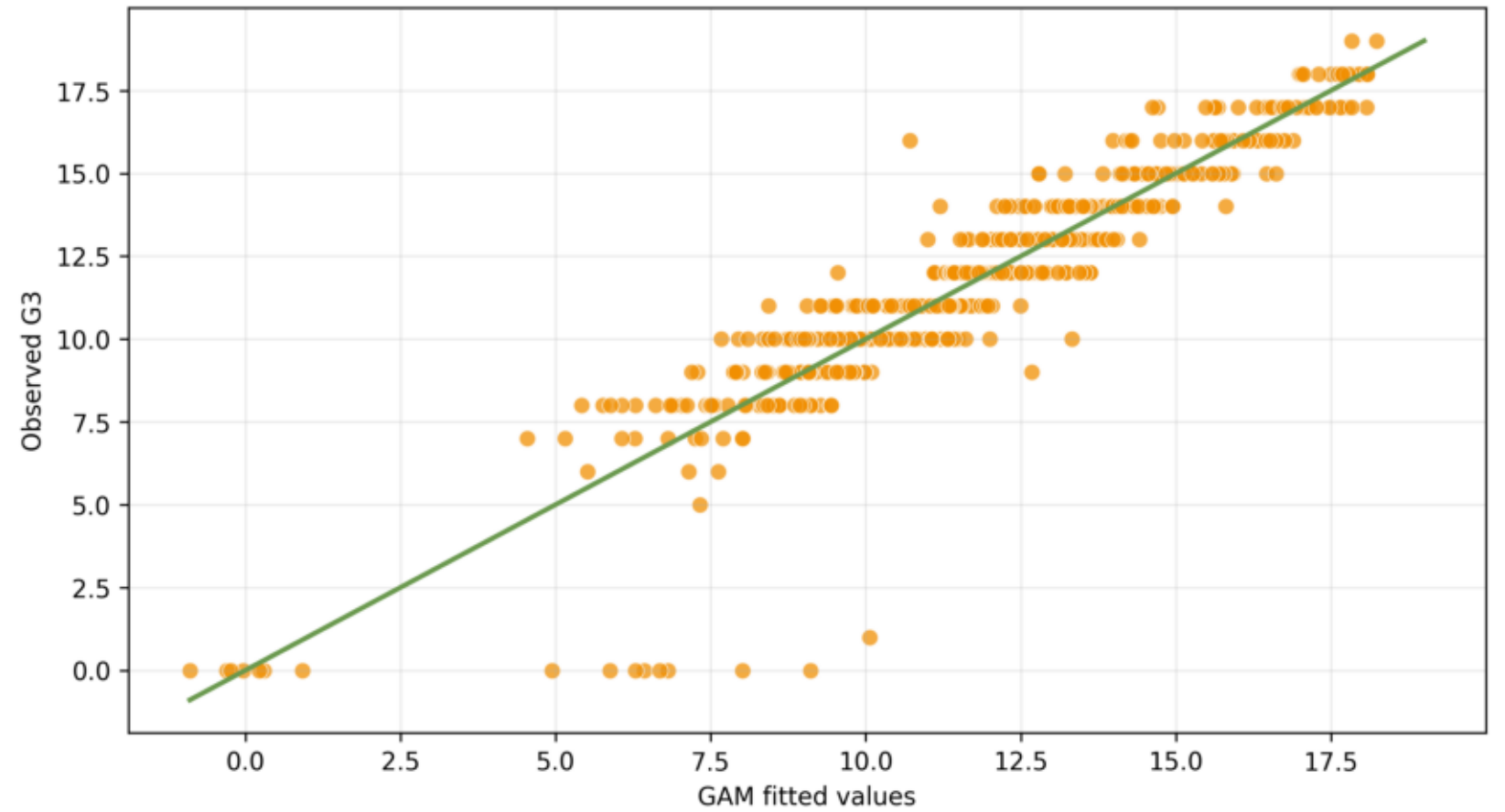
Outcome distribution for GAM

The response variable is checked before fitting the additive smooth model.



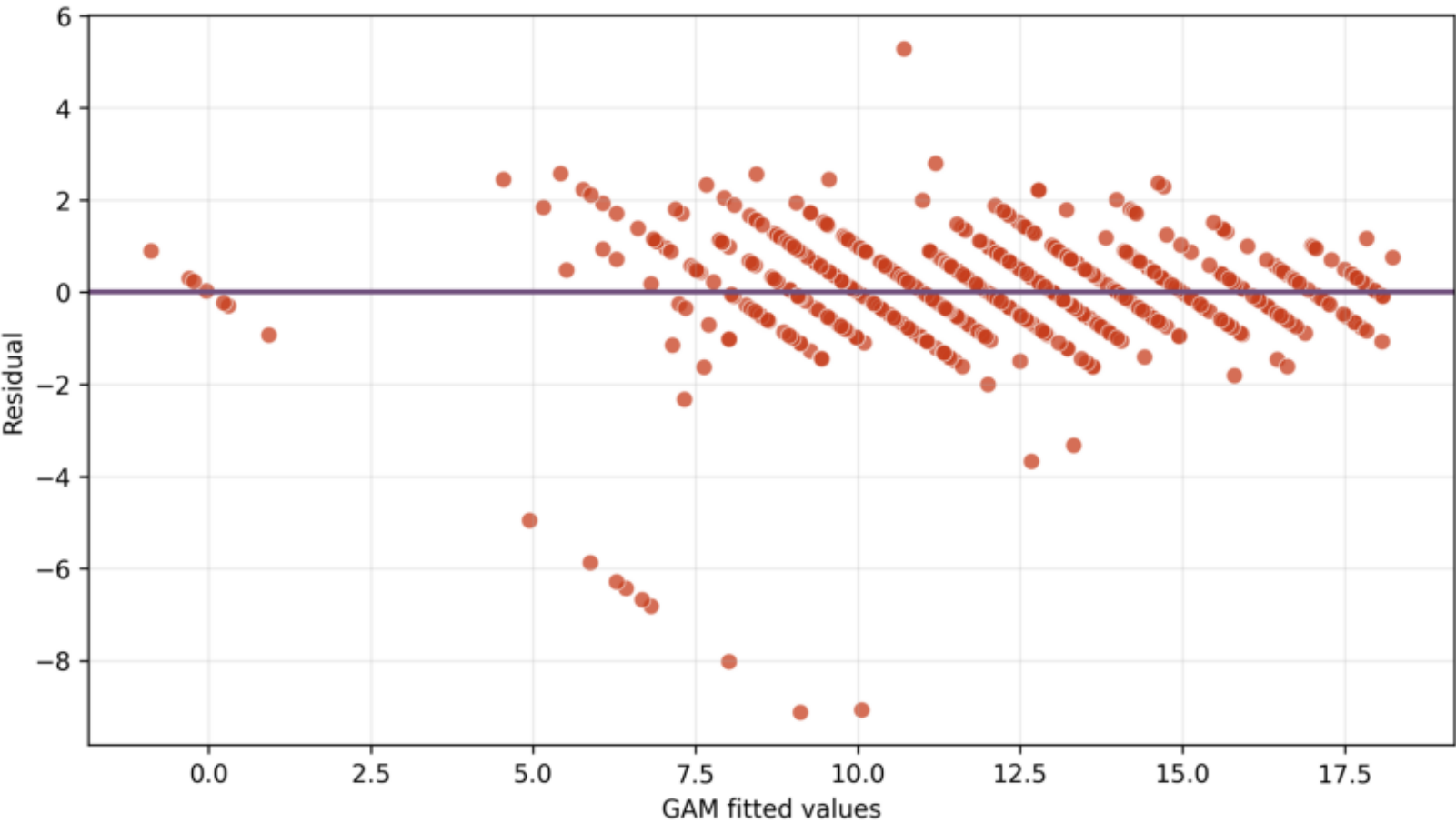
Observed vs GAM fitted values

Points close to the diagonal show stronger agreement between observed and fitted values.



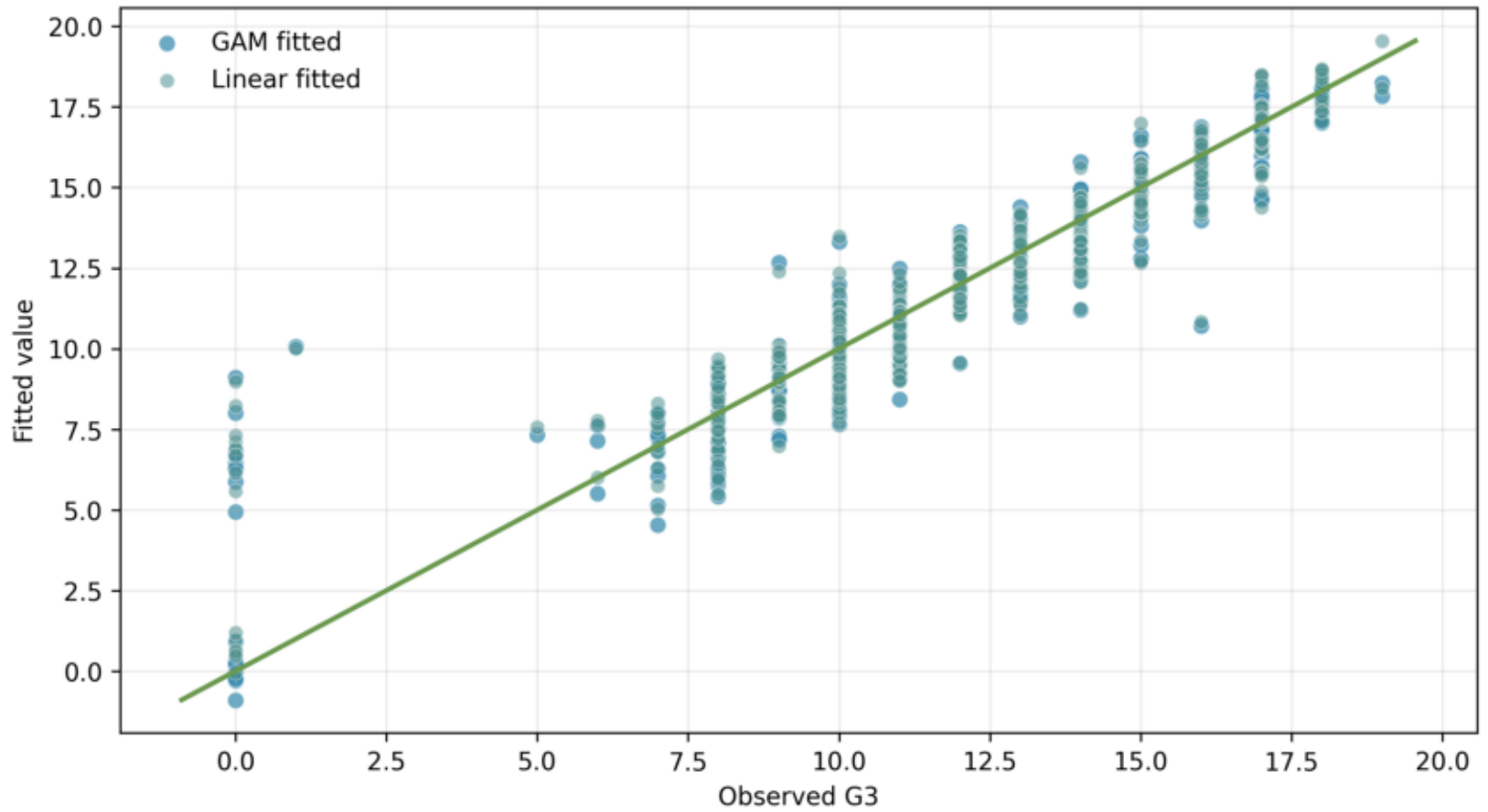
GAM residuals vs fitted values

Residual patterns help diagnose lack of fit after allowing nonlinear additive effects.



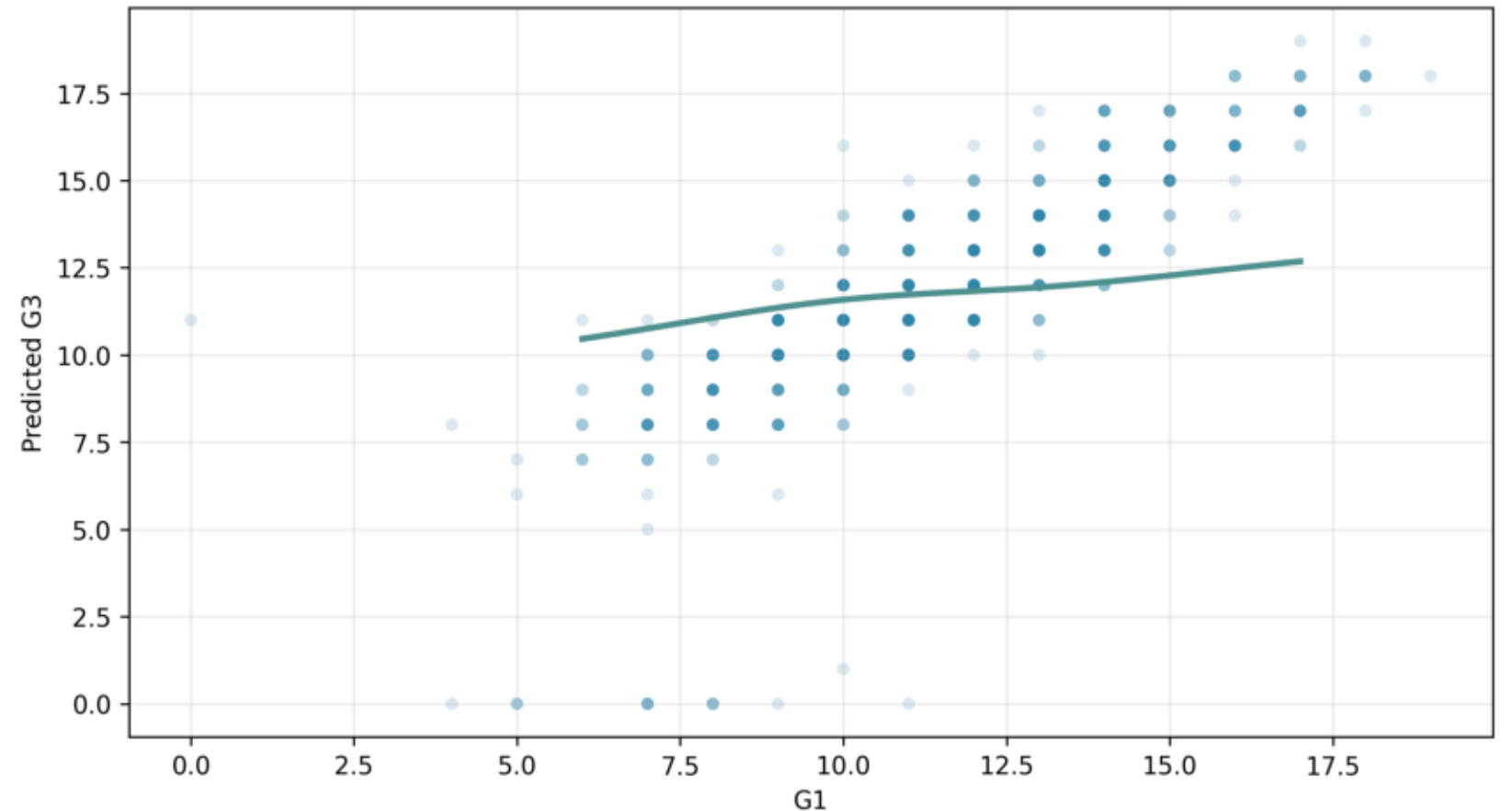
GAM versus linear model fit

This compares additive spline predictions with ordinary linear predictions.



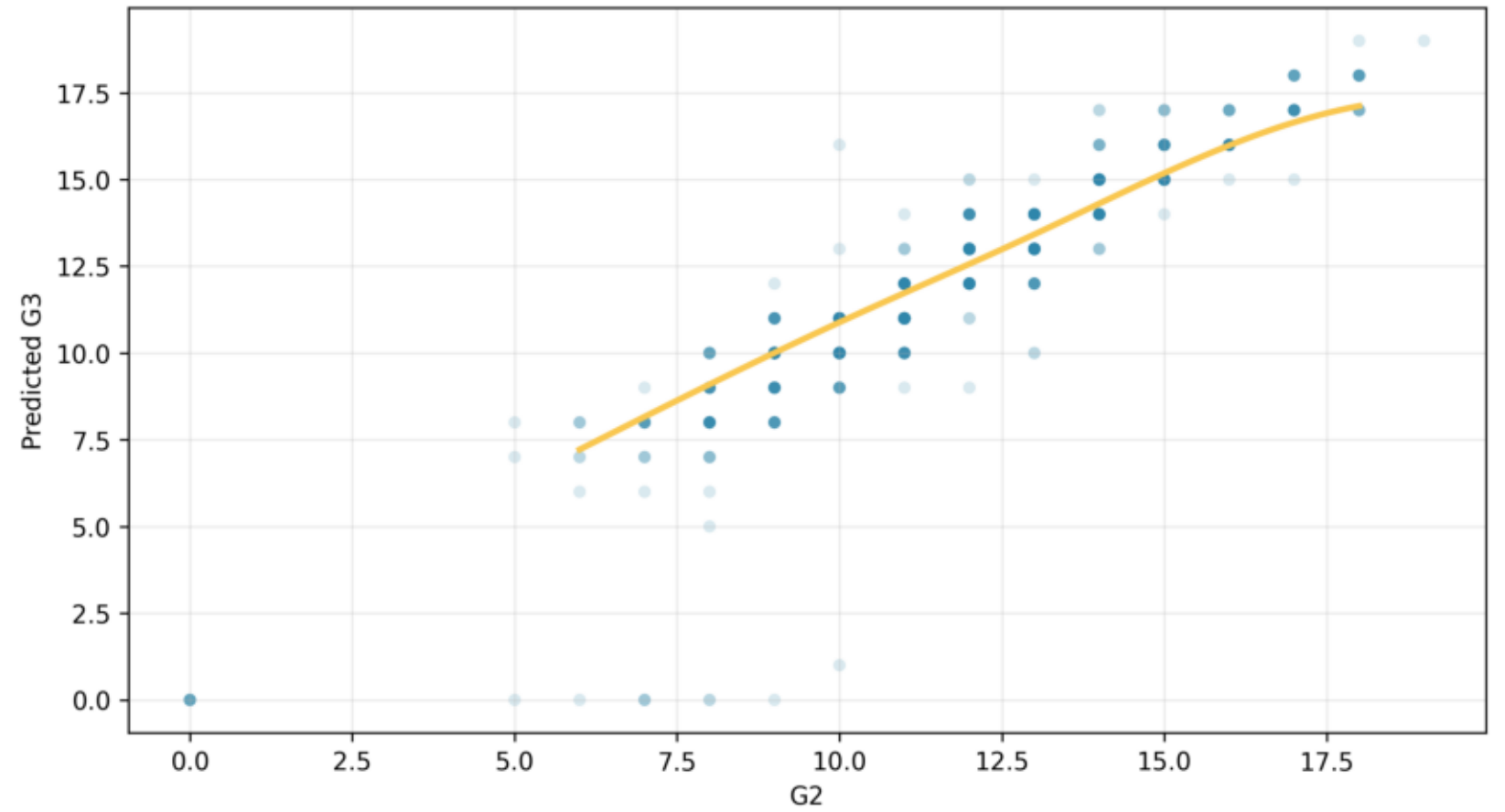
Estimated smooth effect of G1

Other predictors are held at typical values while G1 changes across its range.



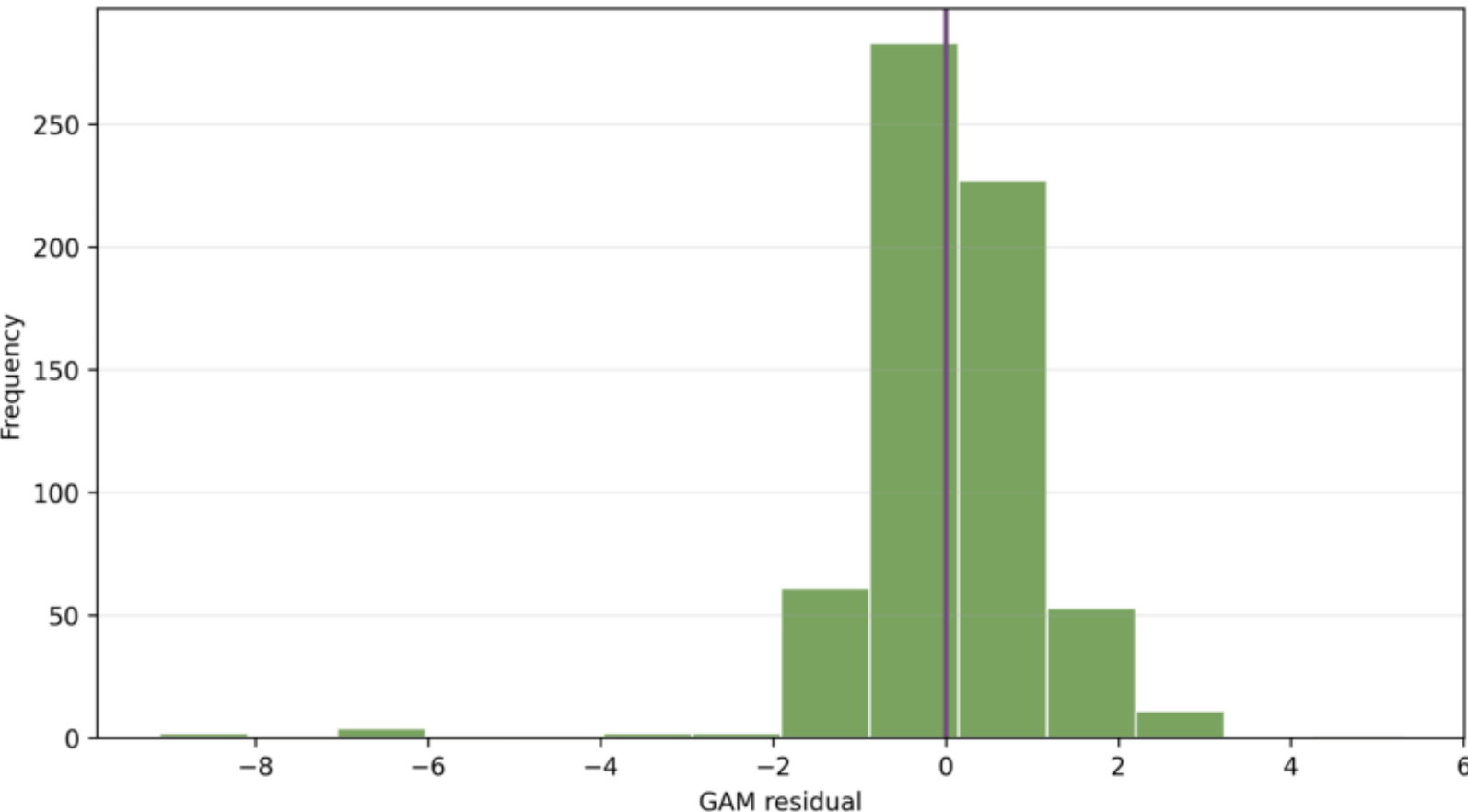
Estimated smooth effect of G2

Other predictors are held at typical values while G2 changes across its range.



GAM residual distribution

A centered residual distribution supports a more balanced fitted additive model.



Model error comparison

Lower RMSE indicates better average prediction accuracy on the fitted data.

