

Generalized Estimating Equations

Test name: Generalized Estimating Equations

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Generalized Estimating Equations\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in GEE model: 645

Clusters used: 27

Cluster source: school, sex, address, famsize, Pstatus

Continuous outcome: G3

Binary outcome: gee_pass = 1 if G3 >= 10

Numeric predictors: G1, G2, studytime, failures, absences, age

Categorical predictors: school, sex, address

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

Main family/link: Gaussian with identity link.

Working correlations fitted: independence and exchangeable.

Exchangeable working-correlation parameter: -0.0051

Inference: GEE robust sandwich standard errors.

Decision guide: $p < 0.05$ and a robust confidence interval not crossing zero indicates stronger evidence.

Reporting guide: state outcome, cluster subject, working correlation, family/link, coefficients, robust SEs, CIs and p-values.

Charts generated: 8

GEE model fit summary

model	family	working_cor	n	clusters	min_cluster	max_cluster	mean_cluster	scale	working_cor	pseudo_r_sq	rmse	mae
Gaussian GE	Gaussian	Independent	645.0000	27.0000	2.0000	128.0000	23.8889	1.5708		0.8521	1.2436	0.7821
Gaussian GE	Gaussian	Exchangeab	645.0000	27.0000	2.0000	128.0000	23.8889	1.5710	-0.0051	0.8521	1.2436	0.7817

GEE coefficients

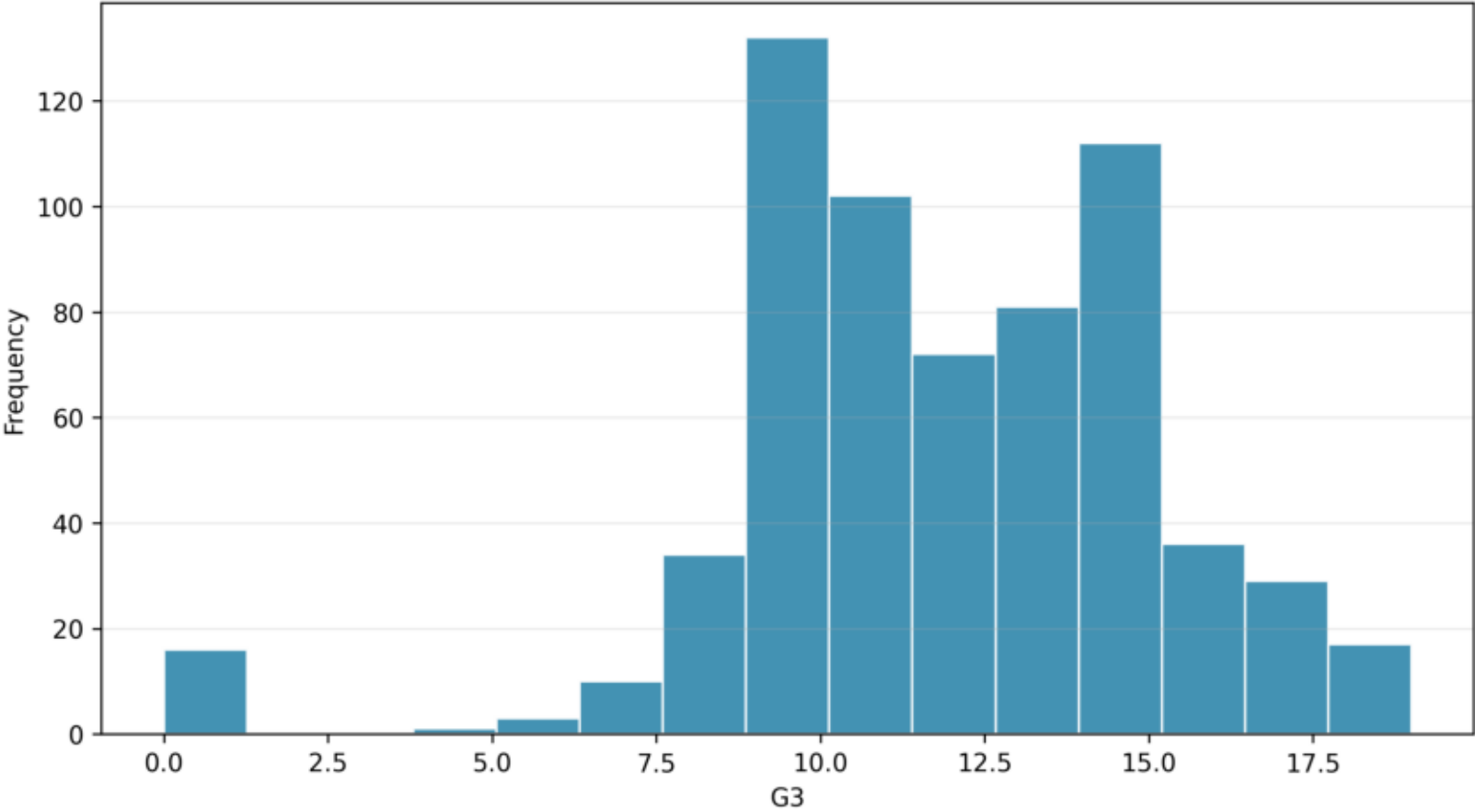
model	term	coefficient	robust_standard_	z_value	p_value	ci95_lower	ci95_upper	decision_alpha_0.05
Gaussian GEE - Ir	Intercept	-0.2078	0.6980	-0.2977	0.7659	-1.5759	1.1603	Not statistically significant
Gaussian GEE - Ir	C(school)[T.MS]	-0.1742	0.0961	-1.8135	0.0698	-0.3625	0.0141	Not statistically significant
Gaussian GEE - Ir	C(sex)[T.M]	-0.2075	0.0960	-2.1618	0.0306	-0.3956	-0.0194	Statistically significant
Gaussian GEE - Ir	C(address)[T.U]	0.0873	0.1212	0.7203	0.4714	-0.1503	0.3249	Not statistically significant
Gaussian GEE - Ir	G1	0.1328	0.0494	2.6878	0.0072	0.0360	0.2296	Statistically significant
Gaussian GEE - Ir	G2	0.8793	0.0303	28.9931	0.0000	0.8199	0.9388	Statistically significant
Gaussian GEE - Ir	studytime	0.0604	0.0511	1.1818	0.2373	-0.0398	0.1606	Not statistically significant
Gaussian GEE - Ir	failures	-0.2227	0.1256	-1.7728	0.0763	-0.4689	0.0235	Not statistically significant
Gaussian GEE - Ir	absences	0.0186	0.0139	1.3408	0.1800	-0.0086	0.0458	Not statistically significant
Gaussian GEE - Ir	age	0.0226	0.0550	0.4113	0.6808	-0.0851	0.1303	Not statistically significant
Gaussian GEE - E	Intercept	-0.1963	0.6927	-0.2834	0.7769	-1.5539	1.1614	Not statistically significant
Gaussian GEE - E	C(school)[T.MS]	-0.1727	0.0938	-1.8405	0.0657	-0.3566	0.0112	Not statistically significant
Gaussian GEE - E	C(sex)[T.M]	-0.2026	0.0735	-2.7556	0.0059	-0.3468	-0.0585	Statistically significant
Gaussian GEE - E	C(address)[T.U]	0.0748	0.1153	0.6492	0.5162	-0.1511	0.3008	Not statistically significant

Cluster summary

gee_cluster	cluster_size	mean_outcome	sd_outcome	pass_rate
1.0000	17.0000	12.6471	2.2063	0.9412
2.0000	128.0000	13.3047	2.3090	0.9766
3.0000	36.0000	13.0278	2.0631	1.0000
4.0000	40.0000	12.0000	2.9872	0.9500
5.0000	13.0000	13.5385	2.9045	0.9231
6.0000	92.0000	11.8587	2.9148	0.8261
7.0000	4.0000	14.5000	1.7321	1.0000
8.0000	31.0000	12.2903	2.2685	0.9032
9.0000	26.0000	11.8077	2.5301	0.9231
11.0000	15.0000	12.0667	2.1865	0.9333
12.0000	7.0000	14.1429	3.5790	1.0000
13.0000	8.0000	10.6250	2.1339	0.7500
15.0000	2.0000	14.0000	2.8284	1.0000
16.0000	2.0000	13.0000	4.2426	1.0000

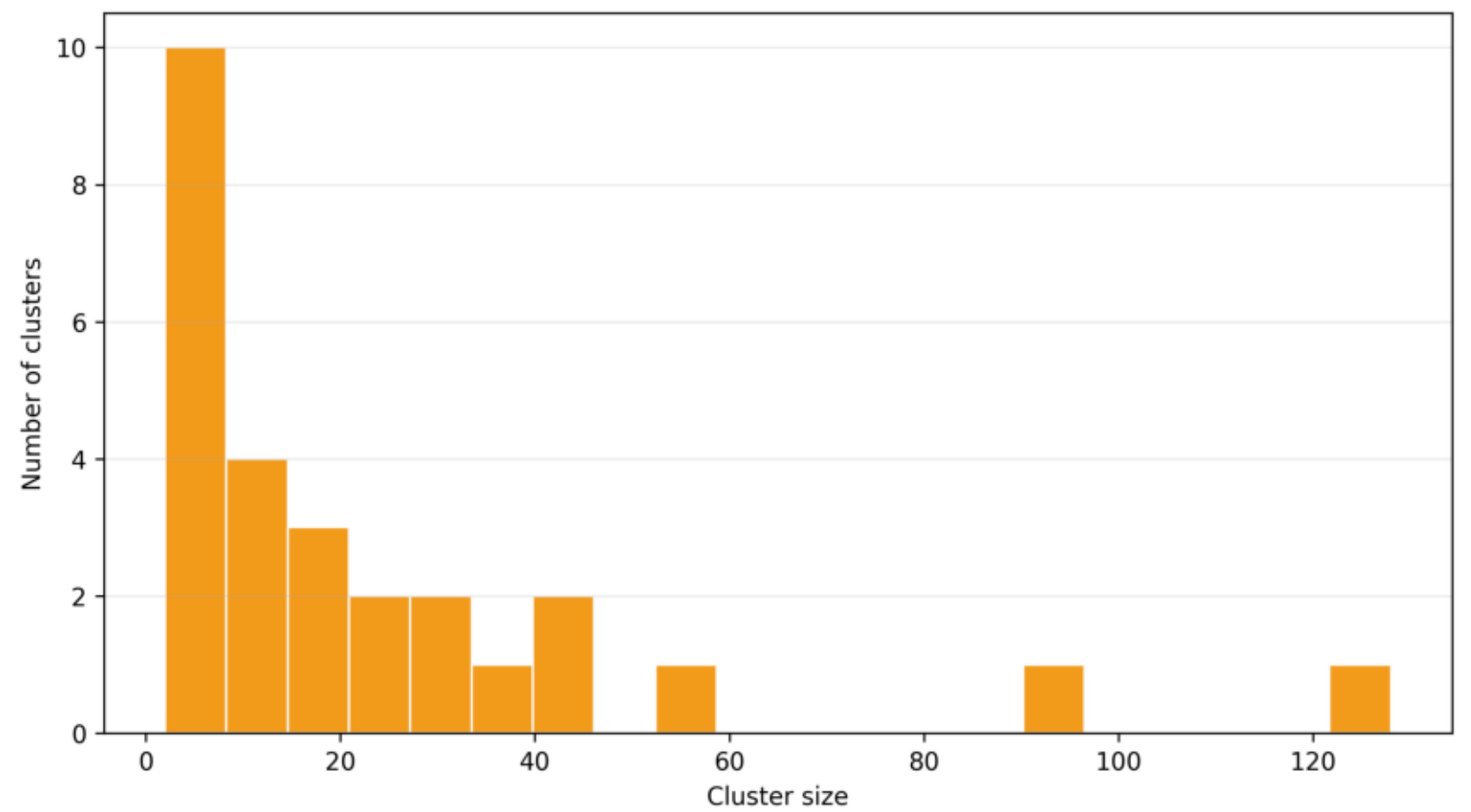
GEE continuous outcome distribution

The Gaussian GEE model uses the continuous outcome with cluster-robust inference.



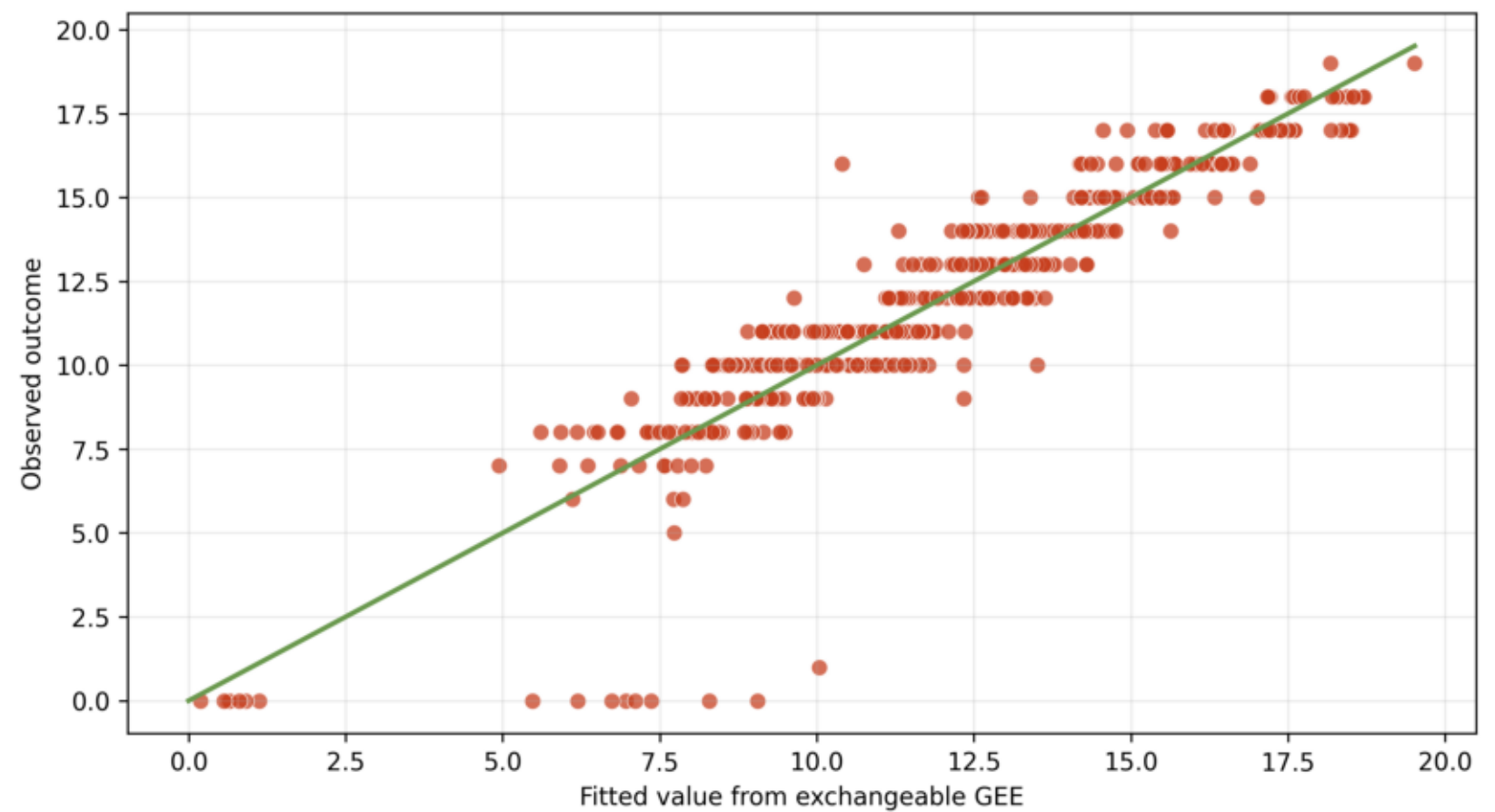
Cluster size distribution

GEE estimates account for observations grouped inside repeated clusters.



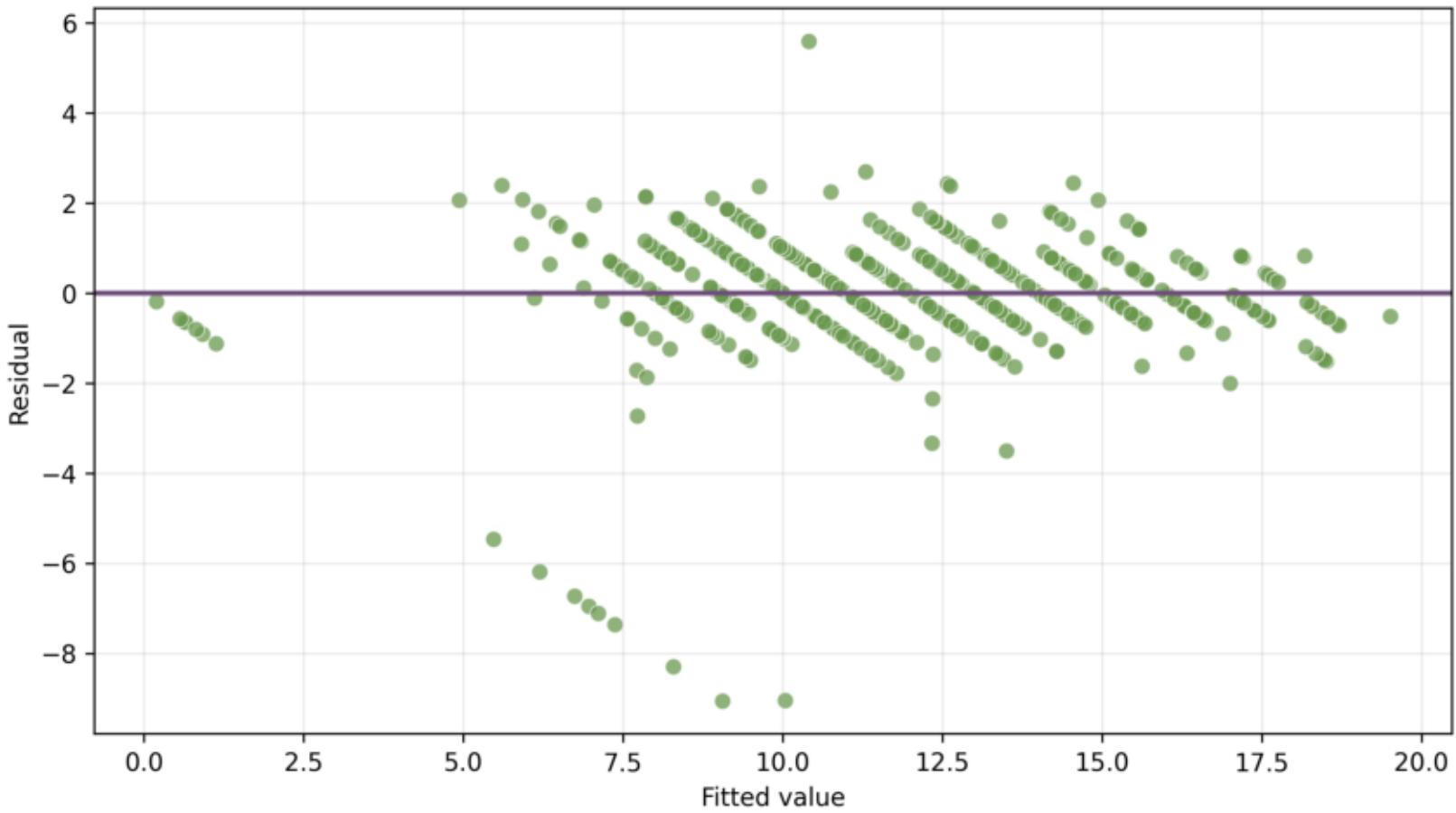
Observed vs fitted GEE values

Better fit is shown by points closer to the diagonal reference line.



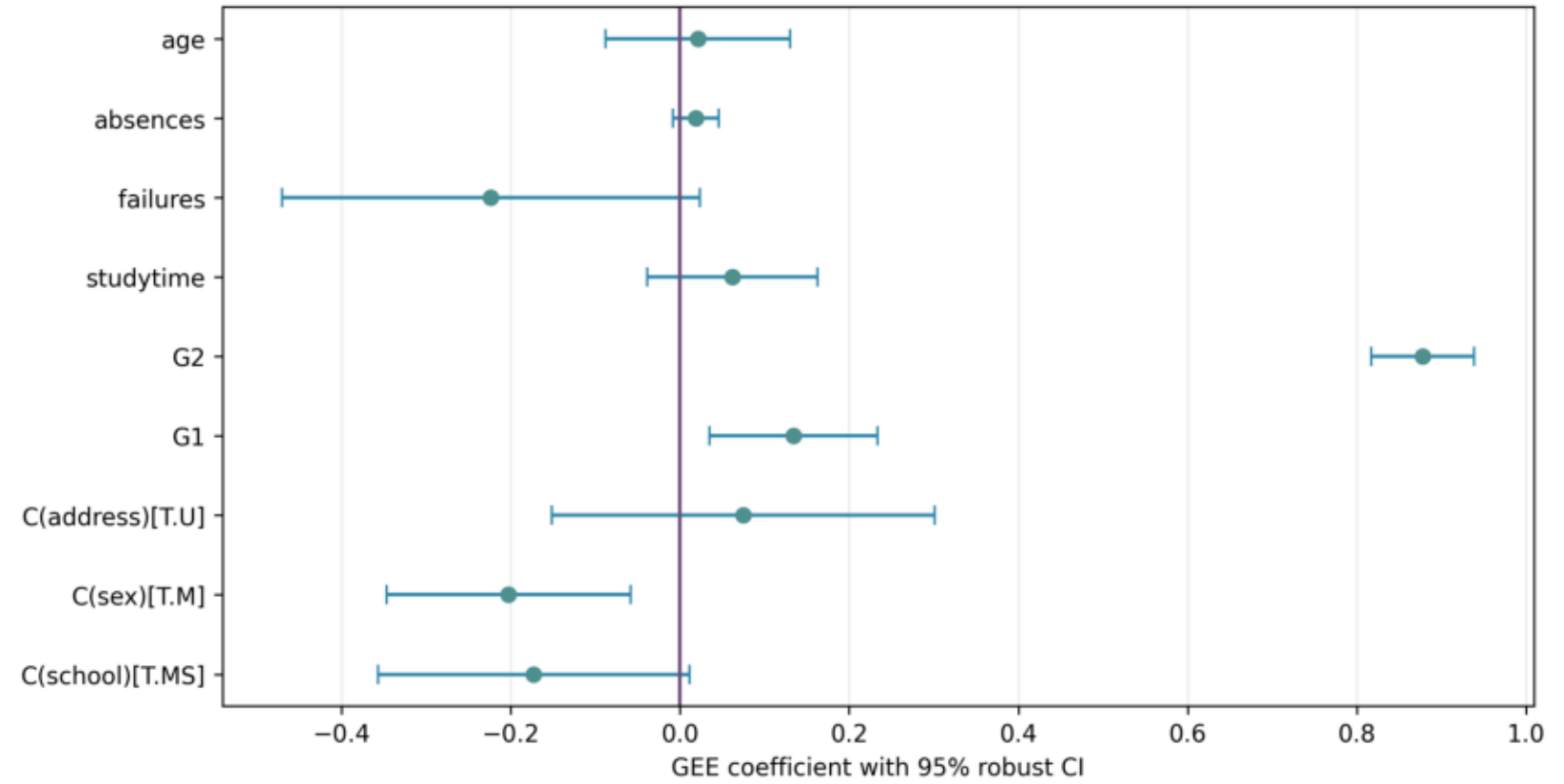
GEE residuals vs fitted values

Random scatter around zero supports a reasonable mean-model specification.



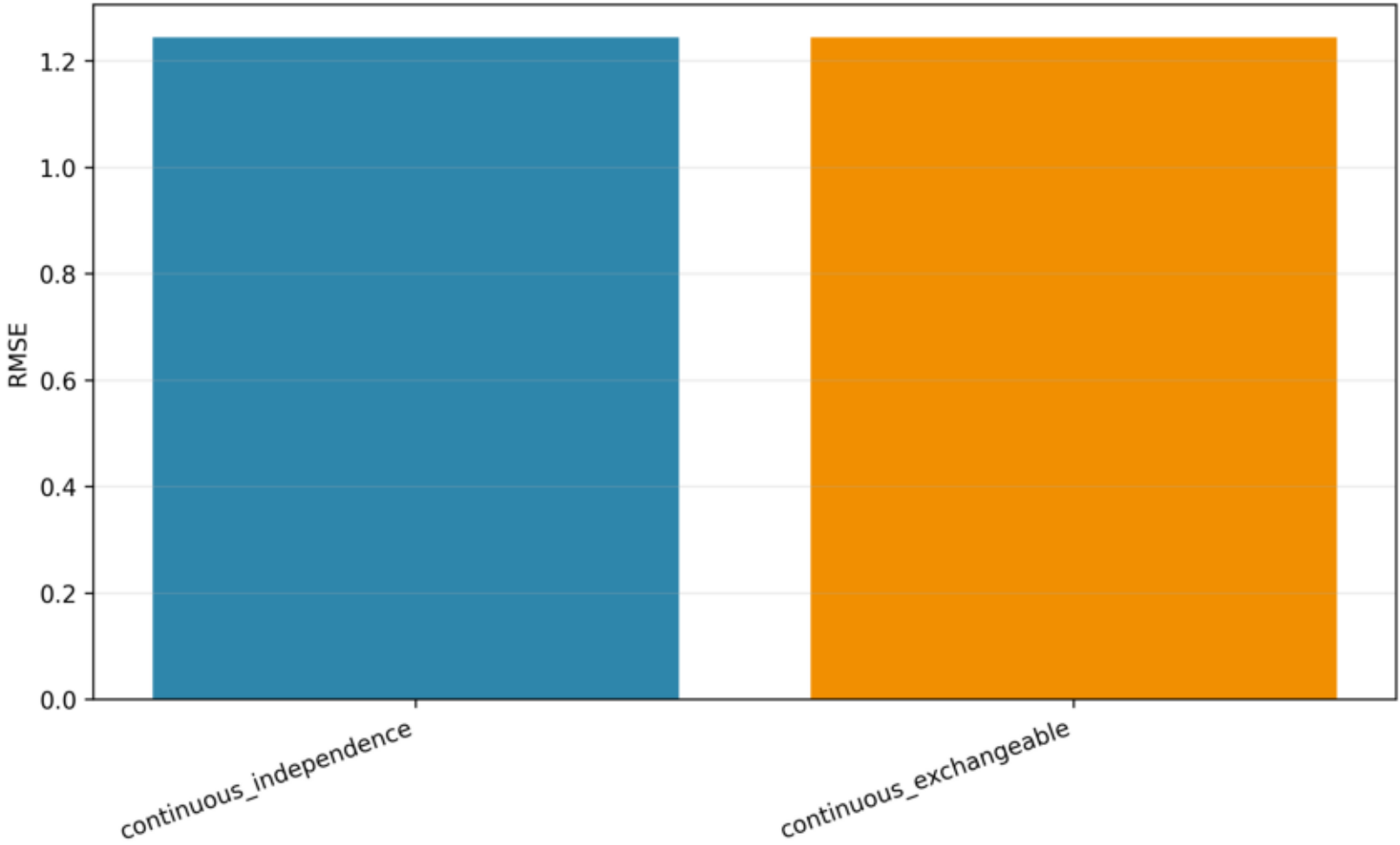
Exchangeable GEE coefficient plot

Intervals not crossing zero indicate stronger evidence for a predictor effect.



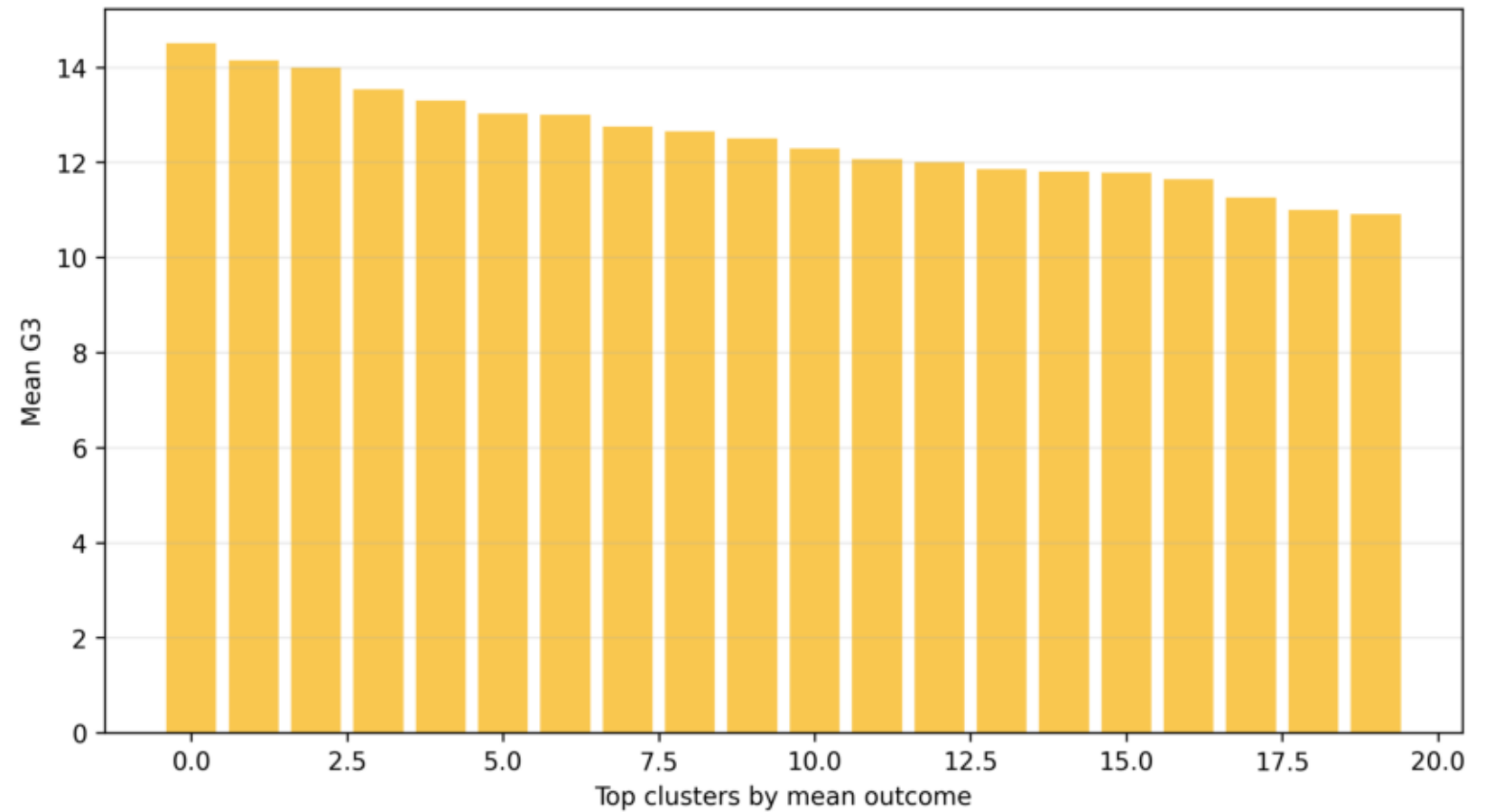
Working-correlation model comparison

Lower RMSE suggests a better mean-prediction summary for the fitted GEE model.



Cluster-level outcome means

This chart shows between-cluster variation handled by the GEE framework.



Binary GEE calibration context

The pass/fail GEE model is summarized by predicted probability groups.

