Generalized Least Squares Regression

Test name: Generalized Least Squares Regression

Dataset: D:\DATA ANALYSIS\H Regression Tests and Models\Generalized Least Squares Regression\dataset.csv

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in GLS model: 649

Outcome variable: G3

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

Categorical predictors: school, sex, address

Formula: G3 ~ G1 + G2 + studytime + failures + absences + age + Medu + C(school) + C(sex) + C(address)

Main model: feasible generalized least squares using inverse estimated error variance weights.

Baseline model: ordinary least squares.

Breusch-Pagan LM p-value: 0.000157

White LM p-value: 0.000348

Durbin-Watson statistic: 1.8723

GLS correction is strongly motivated by heteroscedasticity diagnostics.

GLS fitted with diagonal covariance matrix from estimated variance.

GLSAR rho estimate: 0.0661

Interpretation: compare OLS and GLS/FGLS coefficients, standard errors, p-values, residual spread, and model fit.

Reporting guide: report outcome, predictors, OLS diagnostics, GLS weighting method, coefficient table, model comparison, and residual charts.

Charts generated: 9

GLS model comparison

model	n	df_model	df_resid	r_squared	adj_r_squared	aic	bic	rmse	mae	durbin_watso	weighting_or_si
OLS baseline	649.0000	10.0000	638.0000	0.8524	0.8501	2143.3400	2192.5698	1.2403	0.7791	1.8723	No GLS correcti
Feasible GLS	649.0000	10.0000	638.0000	0.8695	0.8674	2037.0053	2086.2351	1.2423	0.7721	1.8716	Inverse estimat
GLS diagonal	649.0000	10.0000	638.0000	0.8695	0.8674	2037.0021	2086.2318	1.2423	0.7721	1.8716	GLS fitted with
GLSAR autoc	649.0000	10.0000	637.0000	0.8500	0.8477	2137.7207	2186.9335	1.2405	0.7786	1.8668	Estimated AR(1

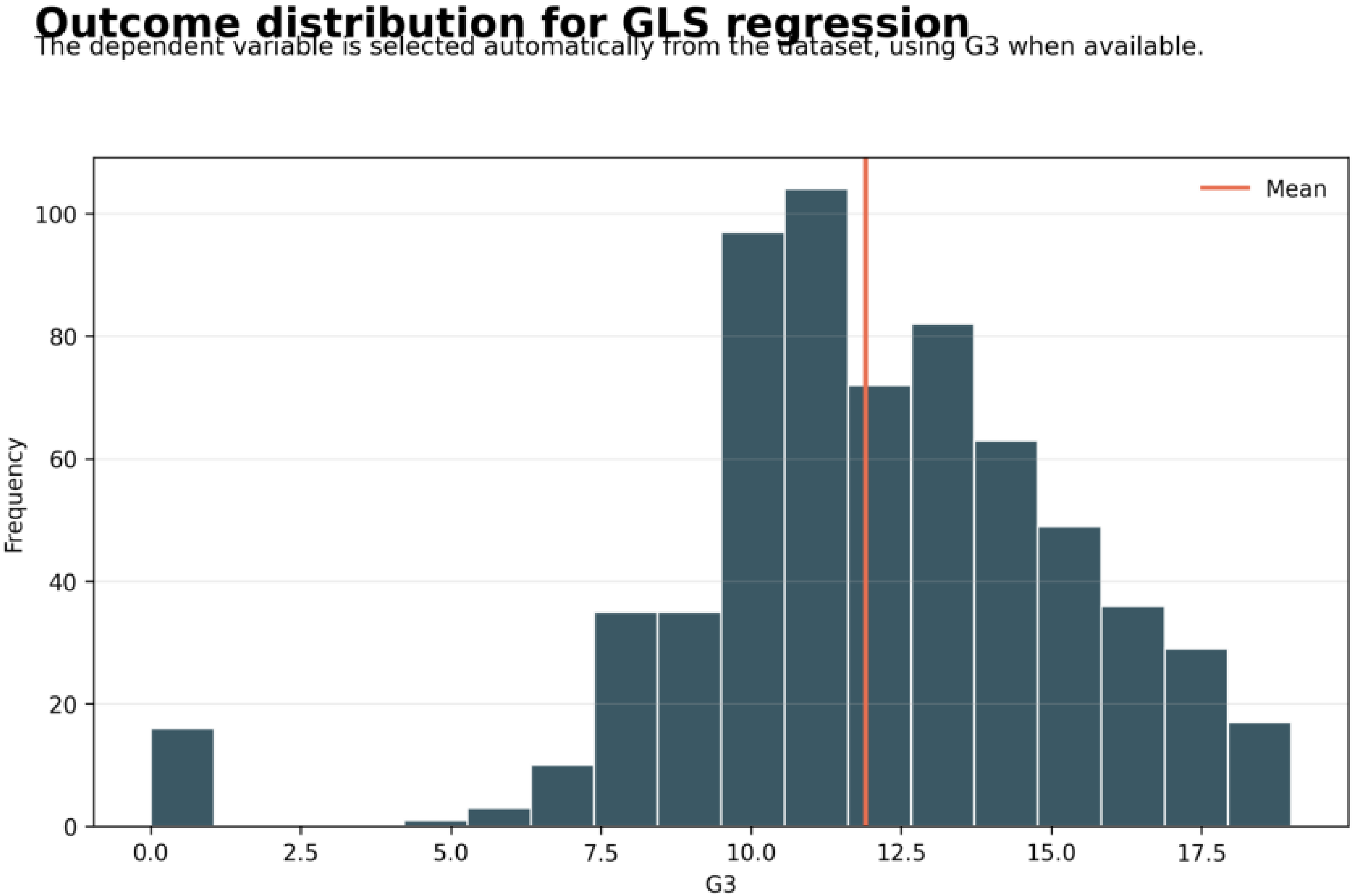
GLS coefficients

model	term	coefficient	standard_error	t_or_z_value	p_value	ci95_lower	ci95_upper	decision_alpha_0_05
OLS baseline	Intercept	-0.1646	0.7837	-0.2101	0.8337	-1.7035	1.3742	Not statistically sig
OLS baseline	C(school)[T.MS]	-0.1788	0.1191	-1.5015	0.1337	-0.4126	0.0550	Not statistically sig
OLS baseline	C(sex)[T.M]	-0.1917	0.1045	-1.8349	0.0670	-0.3969	0.0135	Not statistically sig
OLS baseline	C(address)[T.U]	0.1082	0.1151	0.9402	0.3475	-0.1178	0.3343	Not statistically sig
OLS baseline	G1	0.1337	0.0368	3.6344	0.0003	0.0615	0.2060	Statistically signific
OLS baseline	G2	0.8814	0.0342	25.7364	0.0000	0.8142	0.9487	Statistically signific
OLS baseline	studytime	0.0641	0.0633	1.0121	0.3119	-0.0602	0.1884	Not statistically sig
OLS baseline	failures	-0.2315	0.0950	-2.4377	0.0151	-0.4179	-0.0450	Statistically signific
OLS baseline	absences	0.0177	0.0112	1.5882	0.1127	-0.0042	0.0397	Not statistically sig
OLS baseline	age	0.0222	0.0436	0.5089	0.6110	-0.0634	0.1078	Not statistically sig
OLS baseline	Medu	-0.0379	0.0467	-0.8112	0.4176	-0.1297	0.0539	Not statistically sig
Feasible GLS / WL	Intercept	-0.3473	0.7111	-0.4884	0.6254	-1.7436	1.0490	Not statistically sig

OLS diagnostics

diagnostic	statistic_type	statistic	p_value	interpretation
Breusch-Pagan heteroscedastic	LM statistic	34.4086	0.0002	Evidence of non-constant varianc
Breusch-Pagan heteroscedastic	F statistic	3.5719	0.0001	Evidence of non-constant varianc
White heteroscedasticity	LM statistic	106.8797	0.0003	Evidence of non-constant varianc
White heteroscedasticity	F statistic	1.8634	0.0001	Evidence of non-constant varianc
Durbin-Watson residual depend	DW statistic	1.8723		Values near 2 suggest weak first-o
Jarque-Bera normality context	JB statistic	9864.1705	0.0000	Residuals deviate from normality

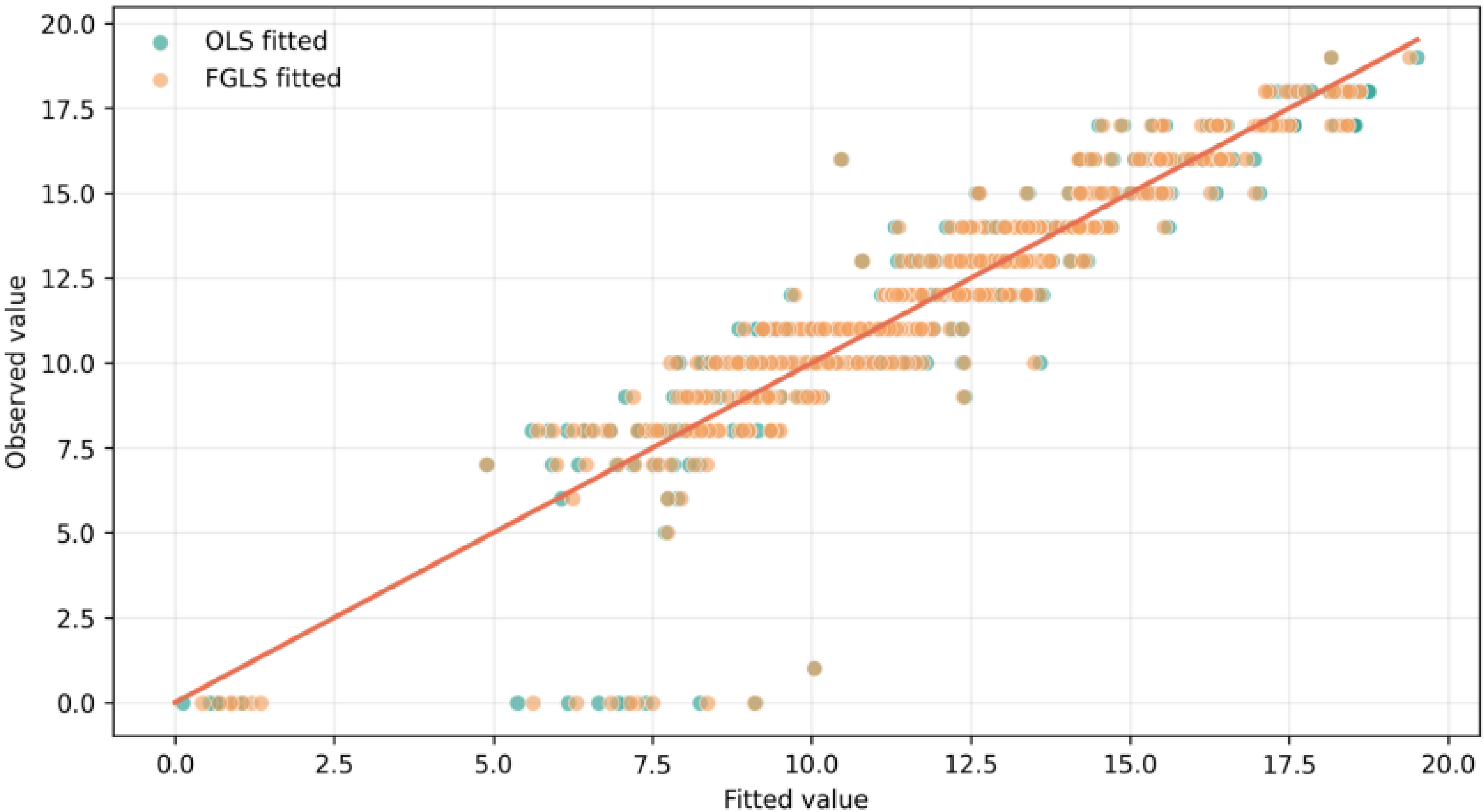
01_gls_outcome_distribution.png



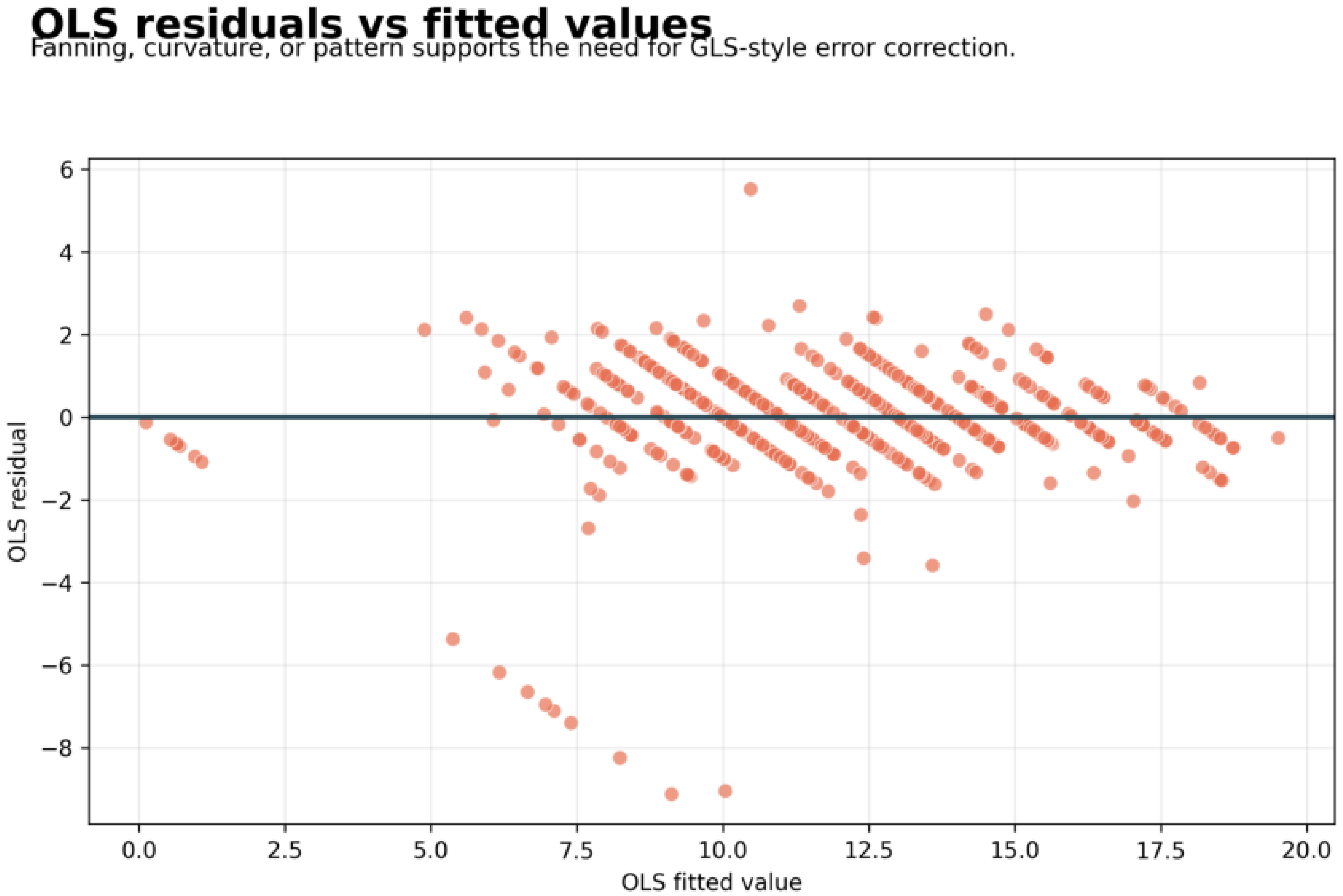
02_gls_observed_vs_fitted_ols_fgls.png

Observed vs fitted values: OLS and feasible GLS

The diagonal line represents perfect prediction; closer points indicate better fitted values.



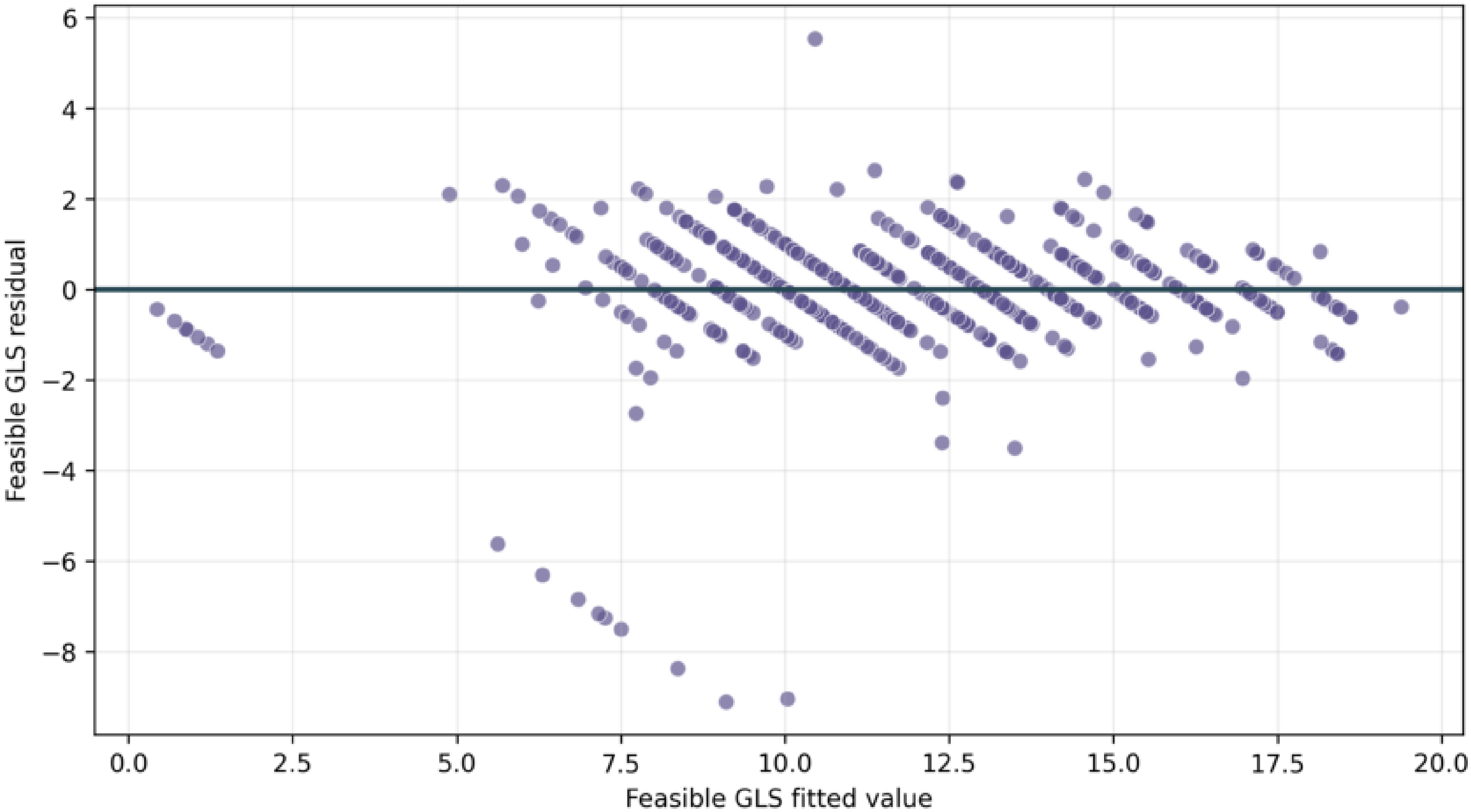
03_gls_ols_residuals_vs_fitted.png



04_gls_fgls_residuals_vs_fitted.png

Feasible GLS residuals vs fitted values

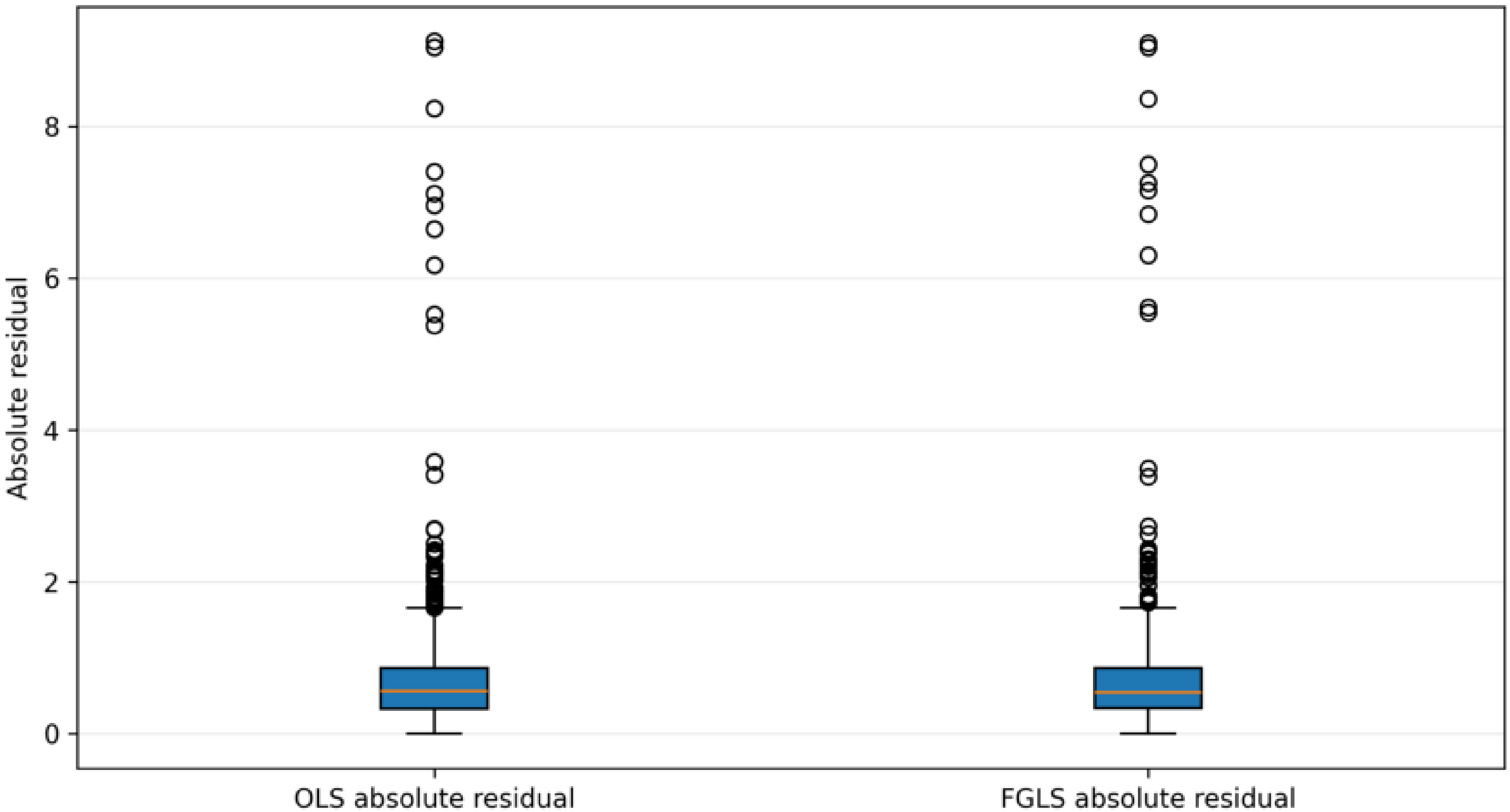
The GLS-weighted model should reduce non-constant residual spread compared with OLS.



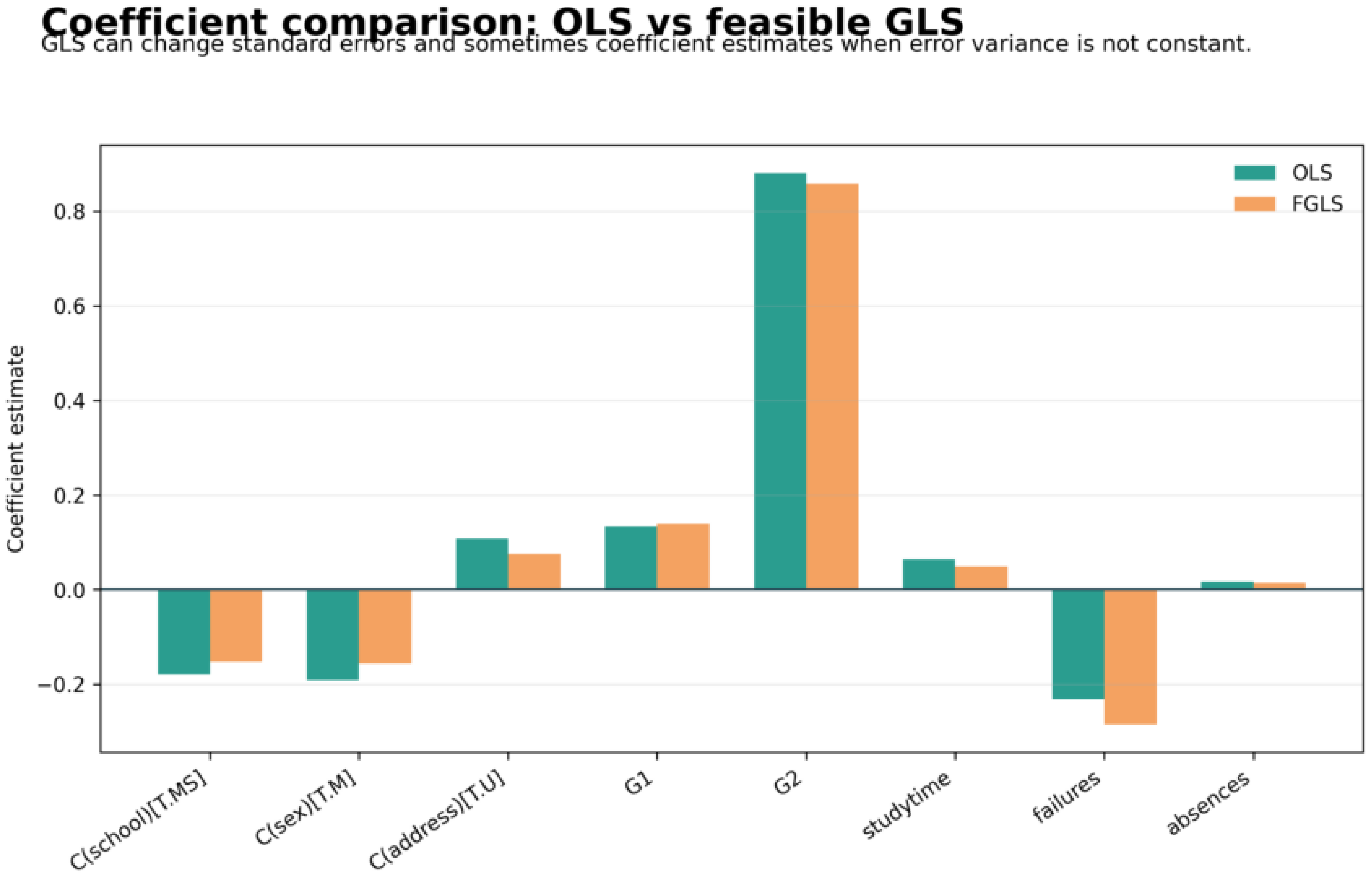
05_gls_residual_spread_comparison.png

Residual spread comparison

Lower and more stable absolute residuals support the GLS correction.



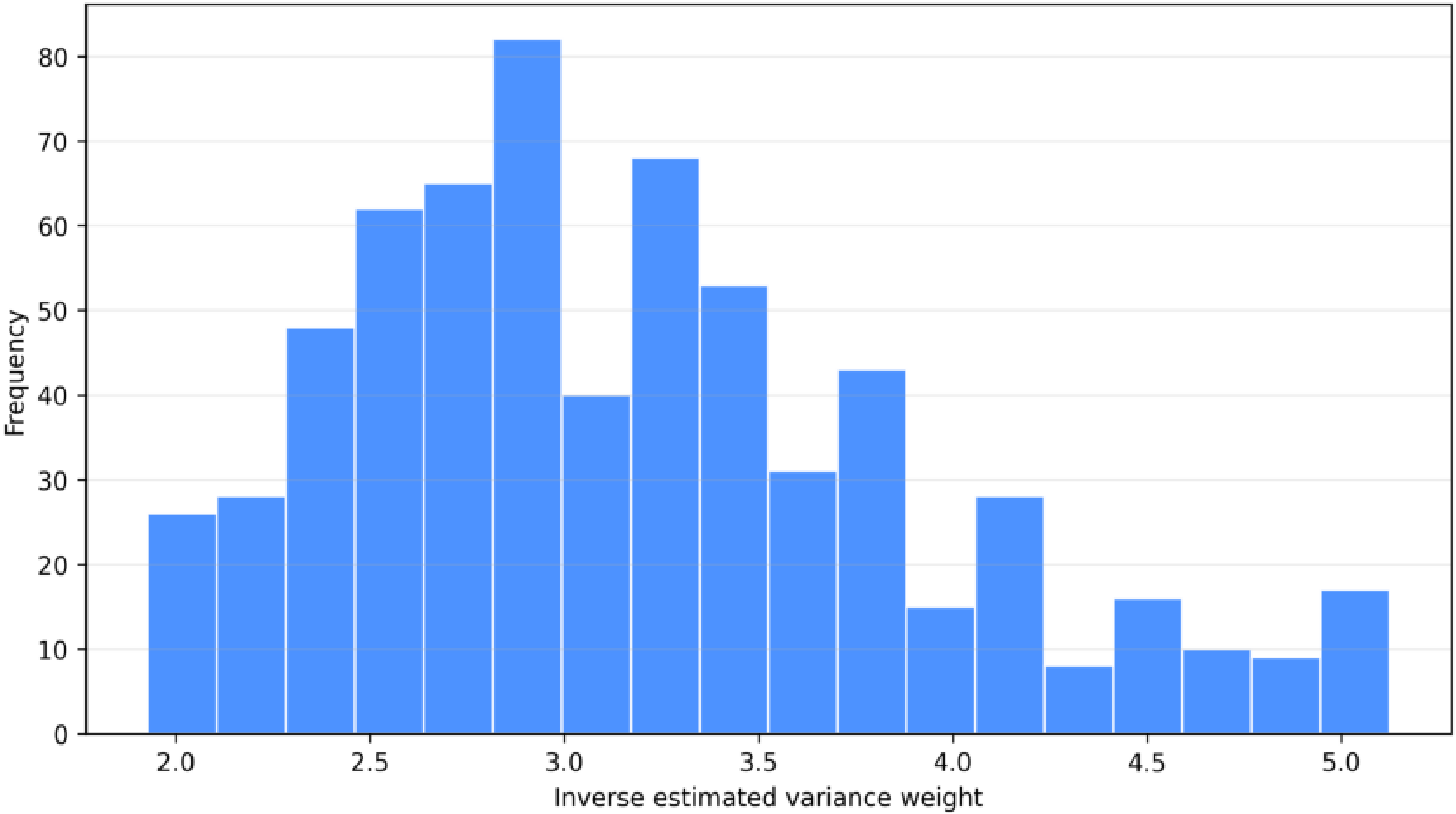
06_gls_coefficient_comparison.png



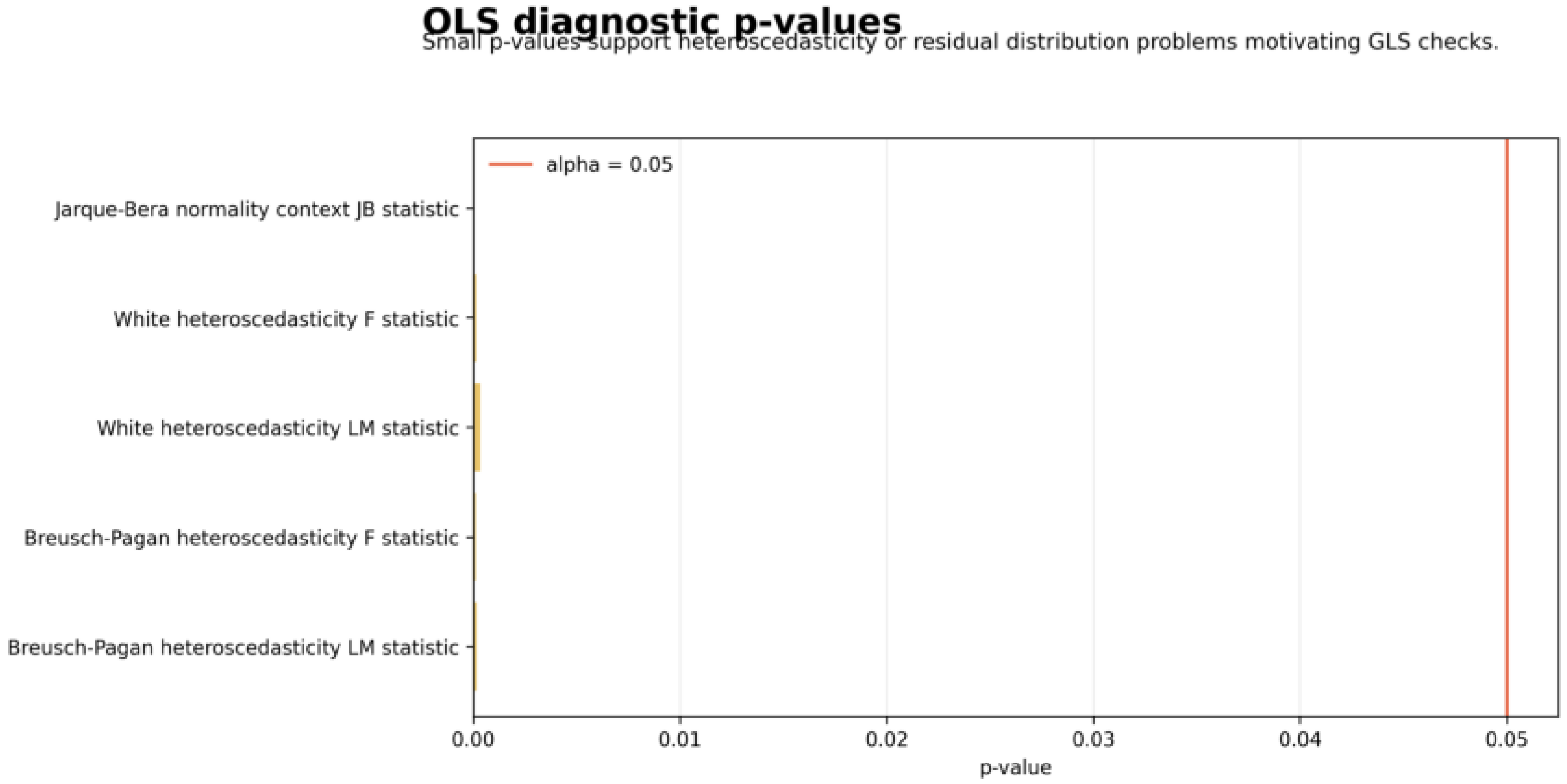
07_gls_weight_distribution.png

Feasible GLS weight distribution

Higher weights are assigned to observations with lower estimated error variance.



08_gls_diagnostic_p_values.png



09_gls_observed_fitted_by_main_predictor.png