Weighted Least Squares Regression

Test name: Weighted Least Squares Regression

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

Rows in original dataset: 649

Columns in original dataset: 33

Rows used in WLS model: 649

Outcome variable: G3

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

Categorical predictors: school, sex, address

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

OLS baseline model fitted first.

Weights created as inverse estimated residual variance from OLS residuals.

Mean scaled WLS weight: 1.0000

Minimum scaled WLS weight: 0.5477

Maximum scaled WLS weight: 1.5136

Breusch-Pagan p < 0.05 suggests heteroscedasticity, so WLS is useful to consider.

OLS adjusted R-squared: 0.8501

WLS adjusted R-squared: 0.8620

Interpretation: WLS coefficients estimate the expected outcome after giving more influence to observations with lower estimated error variance.

Decision guide: compare OLS and WLS coefficients, residual plots, and heteroscedasticity diagnostics before reporting the weighted model.

Reporting guide: state outcome, predictors, weight-construction method, coefficient estimates, 95% CIs, p-values, fit statistics, and diagnostic conclusion.

Charts generated: 7

Weighted Least Squares coefficients

model	term	coefficient	standard_error	t_value	p_value	ci95_lower	ci95_upper	decision_alpha_0_0
Weighted Least S	Intercept	-0.3660	0.7178	-0.5100	0.6103	-1.7755	1.0434	Not statistically sign
Weighted Least S	C(school)[T.MS]	-0.1518	0.1099	-1.3813	0.1677	-0.3677	0.0640	Not statistically sign
Weighted Least S	C(sex)[T.M]	-0.1680	0.0962	-1.7454	0.0814	-0.3569	0.0210	Not statistically sign
Weighted Least S	C(address)[T.U]	0.0759	0.1075	0.7060	0.4805	-0.1353	0.2871	Not statistically sign
Weighted Least S	G1	0.1419	0.0356	3.9844	0.0001	0.0720	0.2118	Statistically signific
Weighted Least S	G2	0.8561	0.0340	25.1439	0.0000	0.7892	0.9229	Statistically signific
Weighted Least S	studytime	0.0506	0.0581	0.8702	0.3845	-0.0636	0.1648	Not statistically sign
Weighted Least S	failures	-0.2766	0.0968	-2.8575	0.0044	-0.4667	-0.0865	Statistically signific
Weighted Least S	absences	0.0151	0.0104	1.4407	0.1501	-0.0055	0.0356	Not statistically sign
Weighted Least S	age	0.0436	0.0408	1.0684	0.2857	-0.0365	0.1238	Not statistically sign

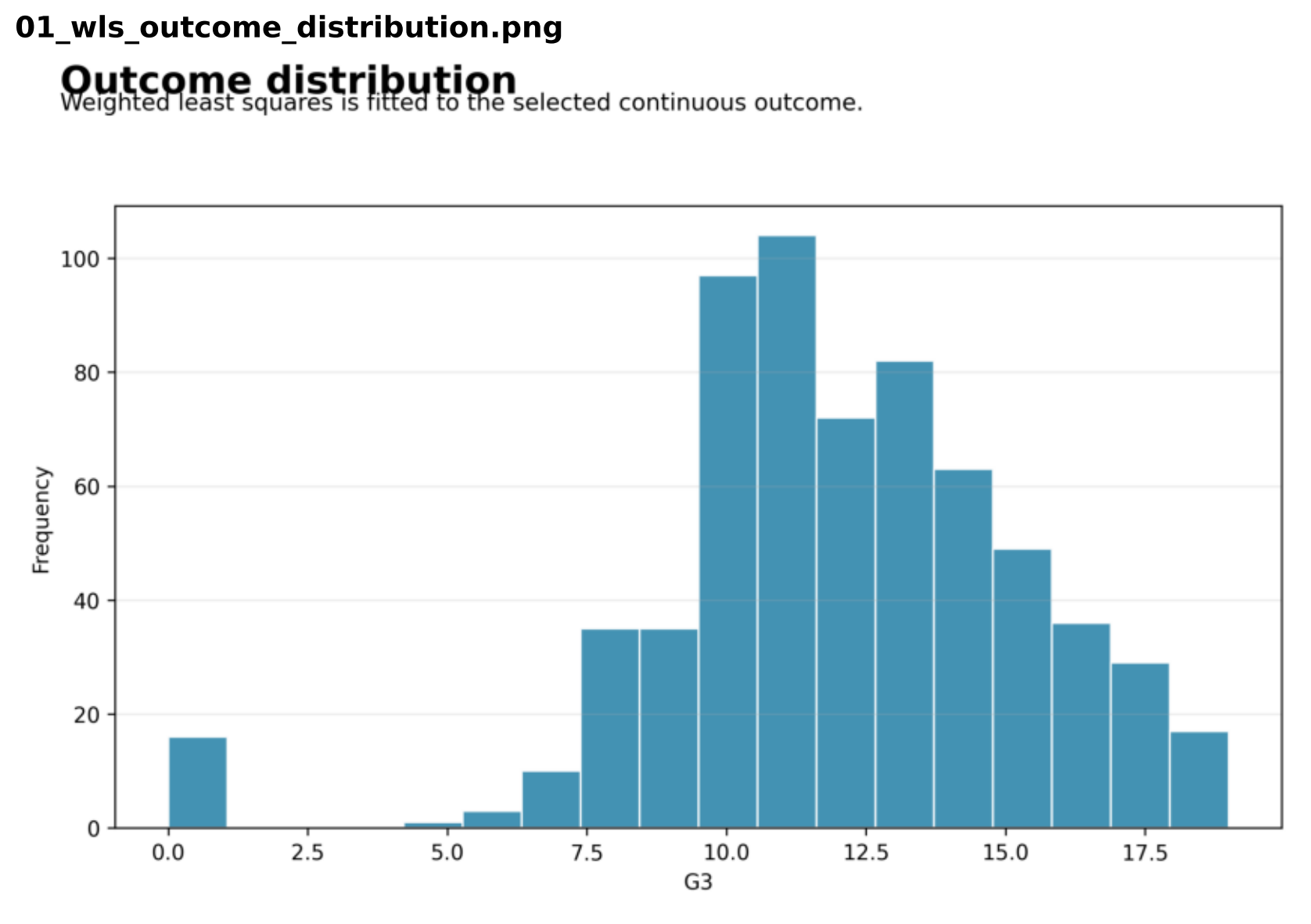
OLS vs WLS fit statistics

model	n	df_model	df_resid	r_squared	adj_r_square	aic	bic	rmse	weighted_rm	f_statistic	f_p_value
Ordinary Lea	649.0000	9.0000	639.0000	0.8522	0.8501	2142.0090	2186.7633	1.2410		409.4514	0.0000
Weighted Lea	649.0000	9.0000	639.0000	0.8639	0.8620	2052.9054	2097.6597	1.2424	1.1482	450.8106	0.0000

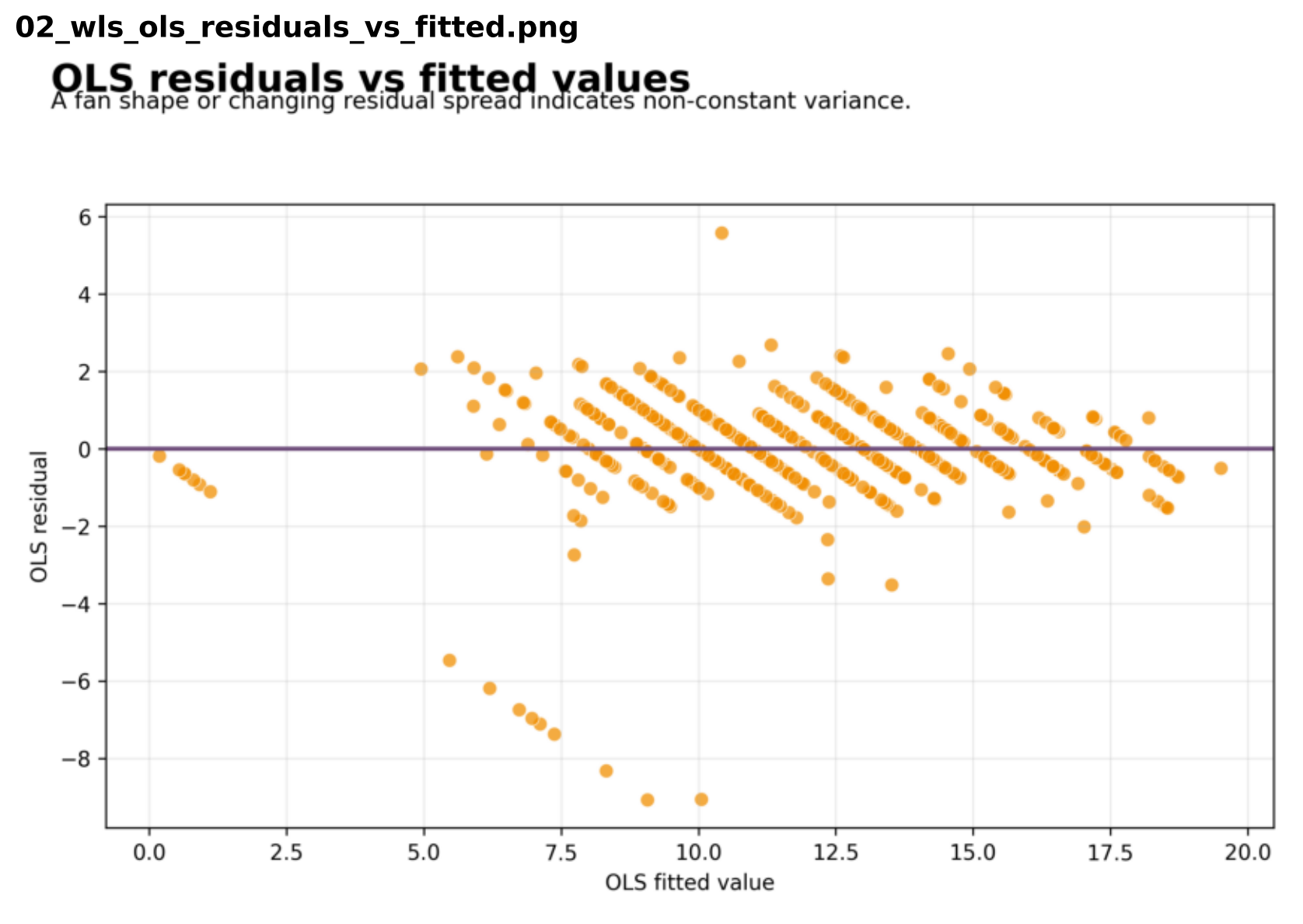
WLS diagnostics

diagnostic	statistic	p_value	interpretation
Breusch-Pagan LM test on OLS residual	34.0147	0.0001	p < 0.05 suggests heteroscedasticity and
Breusch-Pagan F test on OLS residuals	3.9270	0.0001	p < 0.05 suggests non-constant residual v
OLS residual SD	1.2419		Baseline unweighted residual spread.
WLS residual SD	1.2434		Residual spread after inverse-variance we

01_wls_outcome_distribution.png



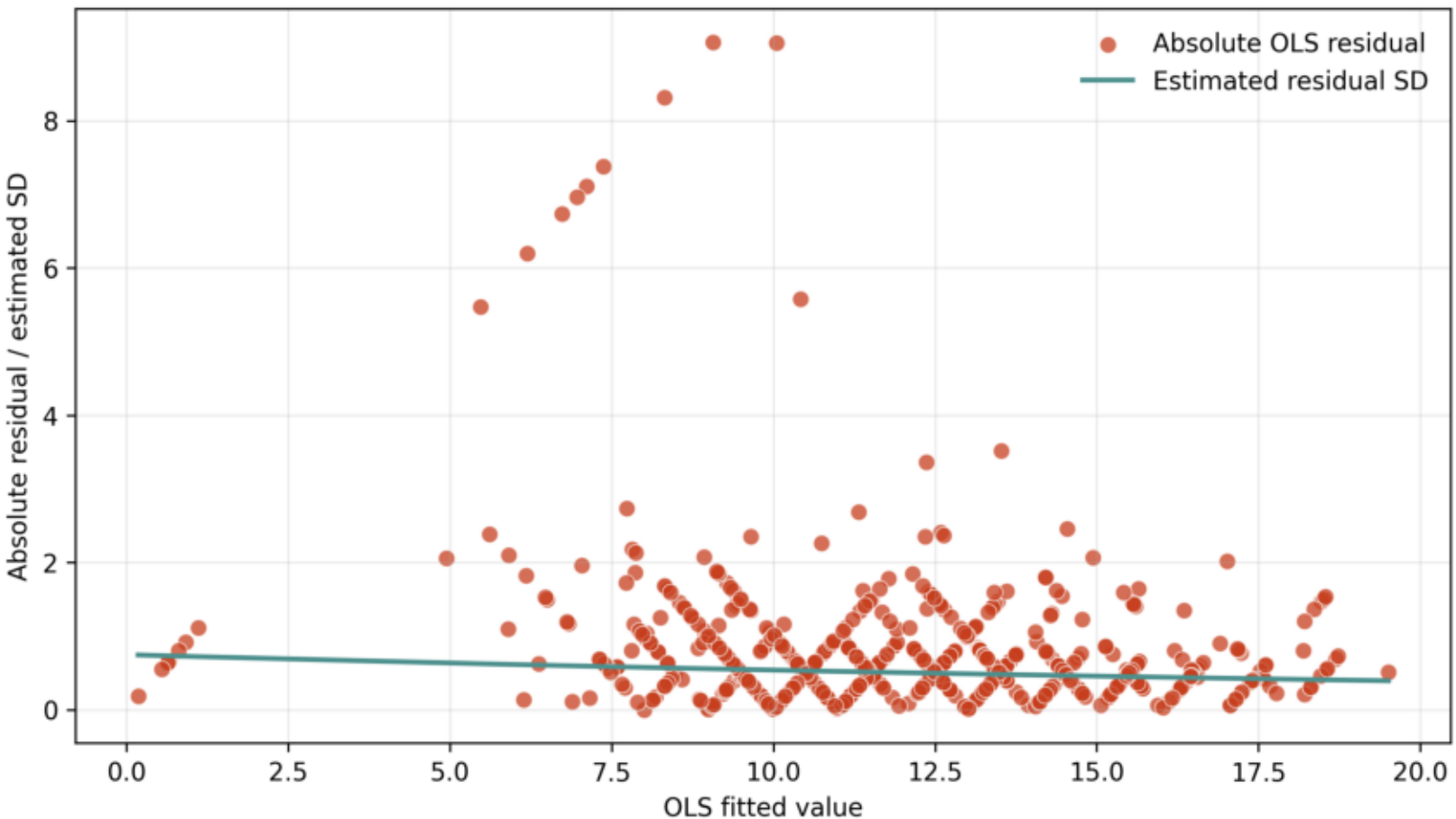
02_wls_ols_residuals_vs_fitted.png



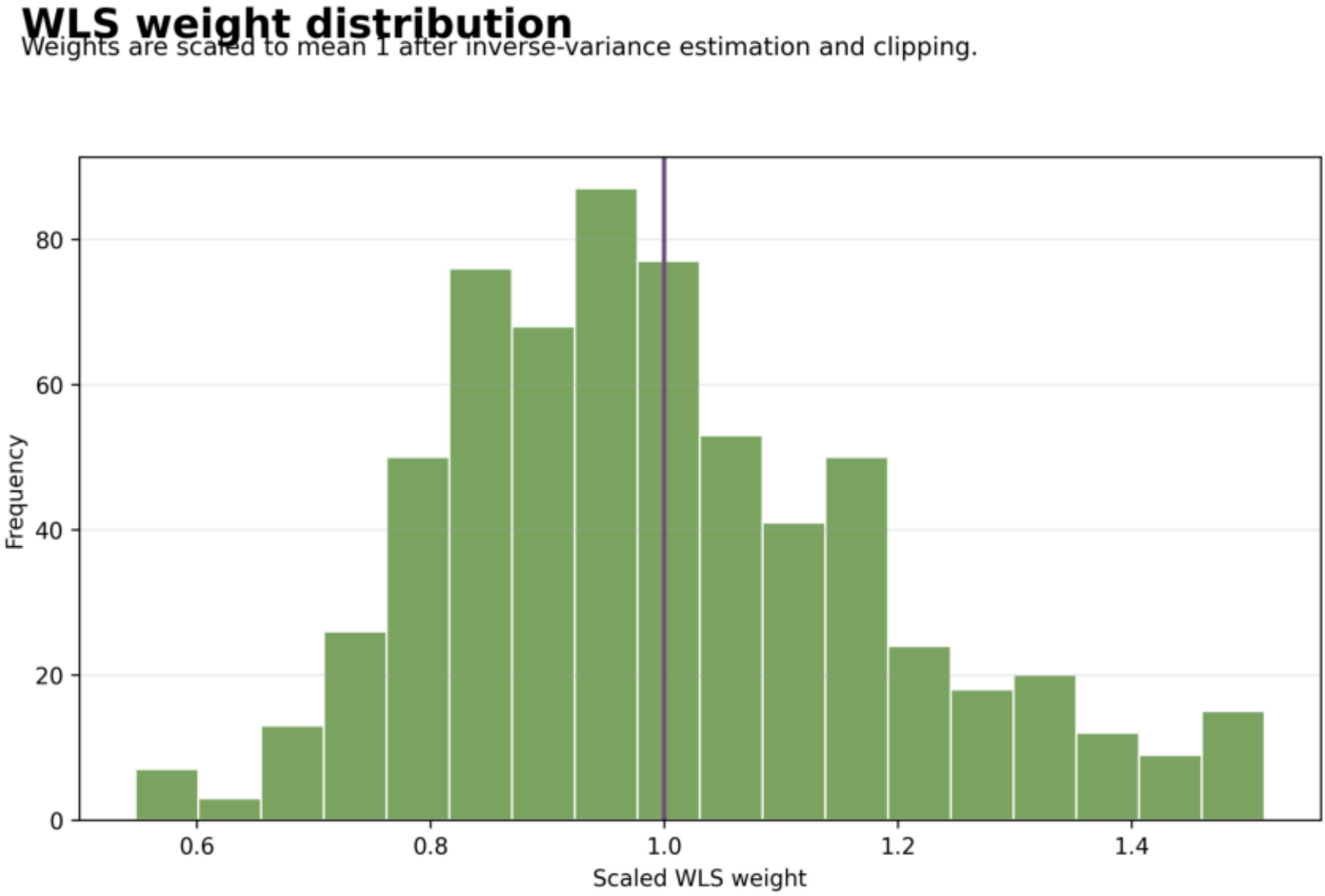
03_wls_estimated_variance_pattern.png

Estimated variance pattern

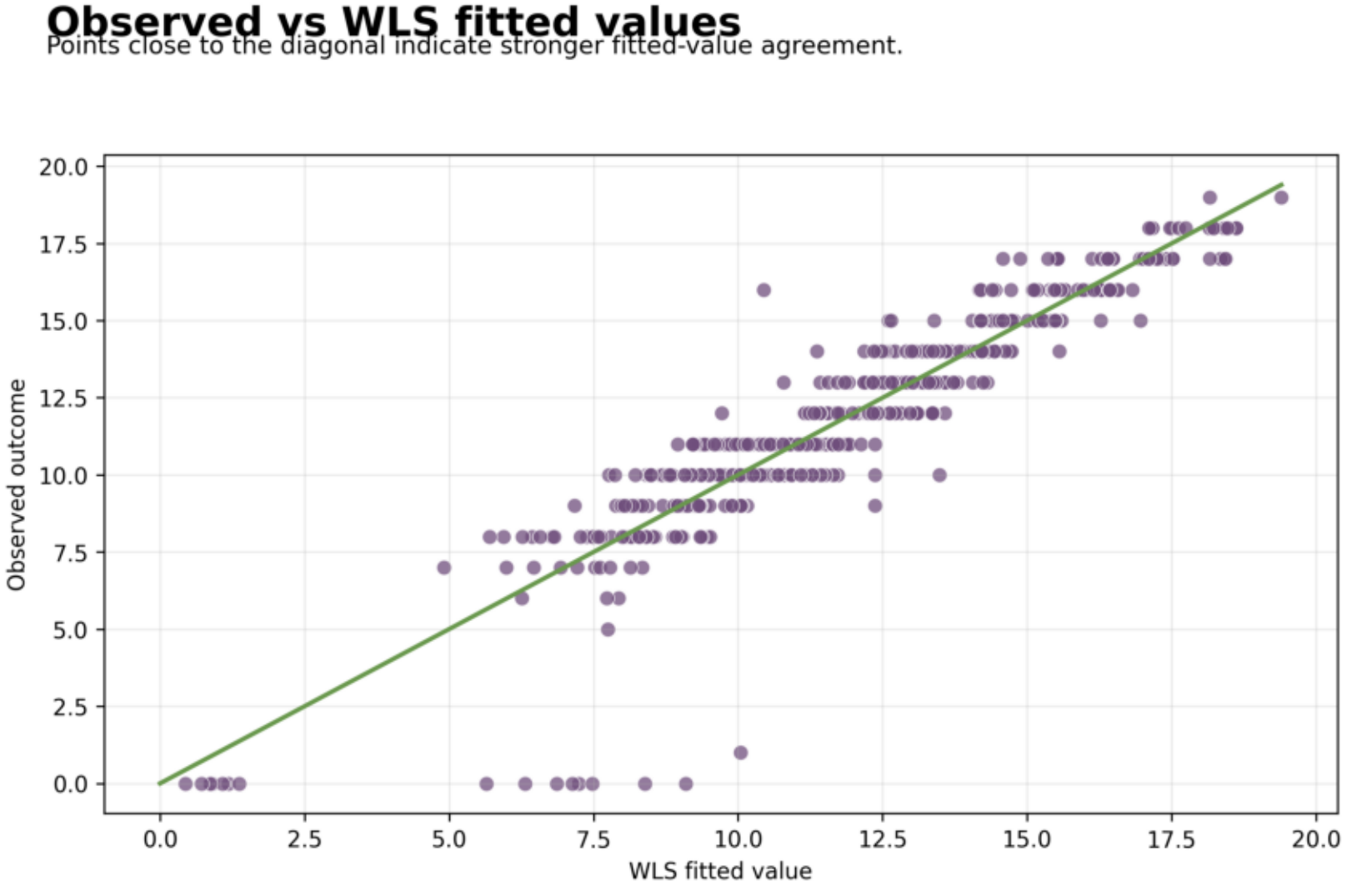
Weights are created from the estimated residual variance pattern.



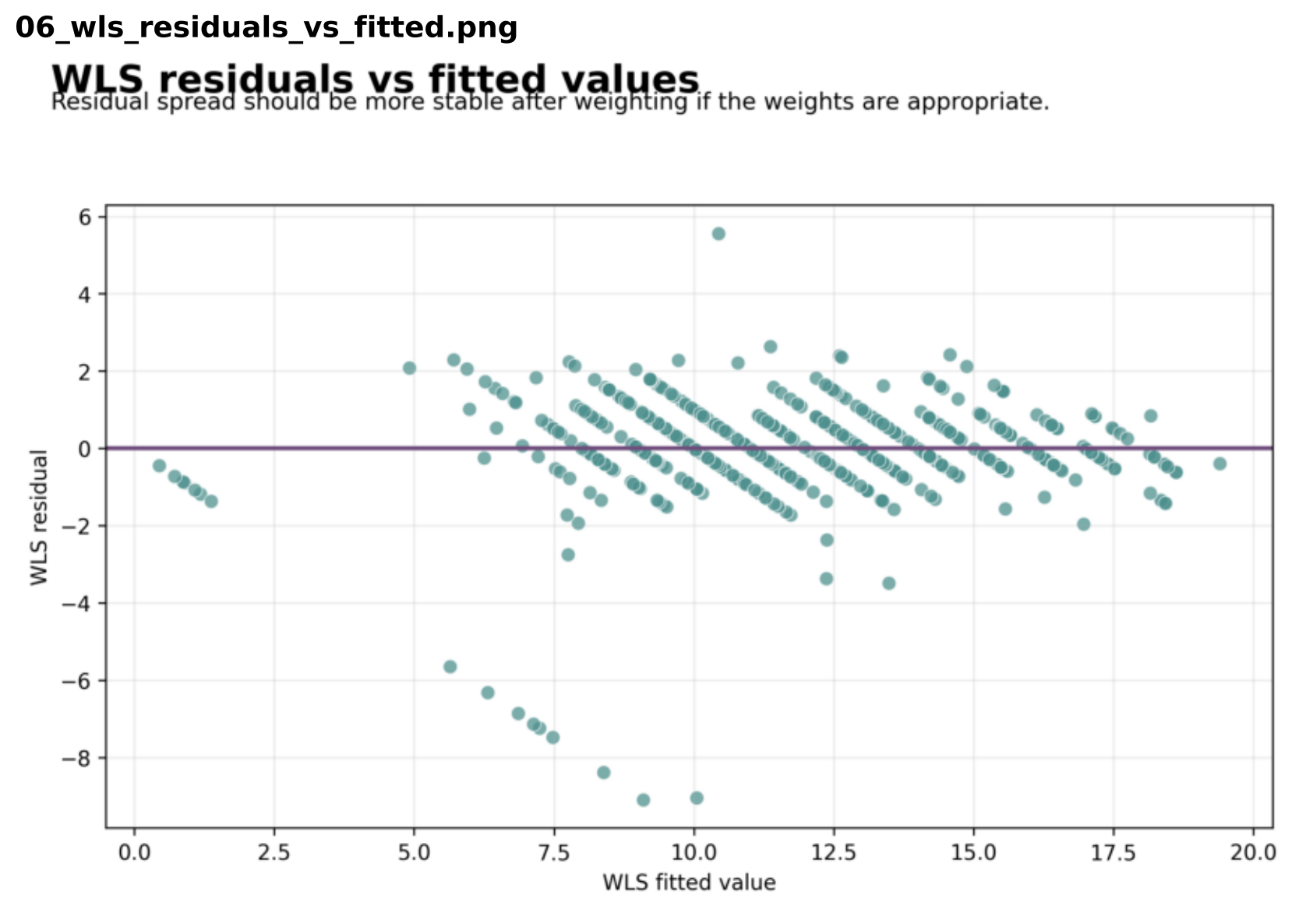
04_wls_weight_distribution.png



05_wls_observed_vs_fitted.png



06_wls_residuals_vs_fitted.png



07_wls_ols_vs_wls_coefficients.png