Residual Diagnostics for Linear Regression: Shapiro–Wilk, Durbin–Watson, and Breusch–Pagan

Why check residuals?

Classical linear regression inference (t–tests, confidence intervals) assumes residuals that are (i) roughly normal, (ii) independent, and (iii) have constant variance (homoskedasticity). The three tests below provide quick, complementary checks of these assumptions.

1. Shapiro–Wilk (Normality)

What: Tests whether residuals are approximately normally distributed.

Null hypothesis: Residuals are normal.

Test statistic:

$$W \;=\; \frac{\left(\sum_{i=1}^{n} a_i \, e_{(i)}\right)^2}{\sum_{i=1}^{n} (e_i - \bar{e})^2},$$

where $e_{(i)}$ are the ordered residuals and a_i are constants derived from the expected values of normal order statistics.

Rule of thumb: $p > 0.05 \Rightarrow$ no strong evidence against normality.

If it fails: Inspect outliers, consider transformations (e.g. $\log(1+y)$), add missing terms/interactions, or use robust/bootstrapped inference.

2. Durbin–Watson (Independence)

What: Detects first–order autocorrelation in residuals (especially for time/ordered data).

Statistic:

$$DW \;=\; \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2},$$

which ranges from 0 to 4; $DW \approx 2$ indicates no autocorrelation, $DW < 2$ positive autocorrelation, $DW > 2$ negative autocorrelation.

Rule of thumb: Values in roughly $[1.5, 2.5]$ are typically acceptable for cross–sectional work.

If it fails: For time series, model the dependence (lags/AR terms), difference the series, or use HAC/ Newey–West standard errors.

3. Breusch–Pagan (Heteroskedasticity)

What: Tests whether residual variance changes with predictors (non–constant variance).

Idea: Regress squared residuals on predictors; the LM statistic is approximately χ^2_p under

$$H_0 : \; \mathrm{Var}(e \mid X) = \sigma^2 \quad \text{(constant)}.$$

Rule of thumb: $p > 0.05 \Rightarrow$ no strong evidence of heteroskedasticity.

If it fails: Transform variables (e.g. log), add missing predictors or nonlinearity (polynomial/splines), or use heteroskedasticity–robust standard errors (e.g. HC3).

Quick Reference

Check	What it tests	Desired outcome	Action if it fails
Shapiro–Wilk	Normality	$p > 0.05$	Transform, handle outliers, robust inference
Durbin–Watson	Autocorrelation	$DW \approx 2$	Add lags / HAC SEs / TS model
Breusch–Pagan	Constant variance	$p > 0.05$	Transform / add terms / robust SEs

Drop–in Python Snippet

Assumes you already computed predictions and have `y_test`, `y_pred`, `X_test`, and a list `numeric_cols`.

```
from scipy import stats
import pandas as pd
import statsmodels.api as sm
from statsmodels.stats.stattools import durbin_watson
from statsmodels.stats.diagnostic import het_breuschpagan

# residuals from your fitted model
residuals = y_test - y_pred

# 1) ShapiroWilk (normality)
sh_stat, sh_p = stats.shapiro(residuals)

# 2) DurbinWatson (autocorrelation)
dw = durbin_watson(residuals)

# 3) BreuschPagan (heteroskedasticity)
X_bp = sm.add_constant(pd.DataFrame(X_test, columns=numeric_cols))
bp_stat, bp_p, _, _ = het_breuschpagan(residuals, X_bp)

print(f"Shapiro p={sh_p:.3f} | DW={dw:.2f} | BP p={bp_p:.3f}")
```

One–Line Cheat Sheet

- Shapiro: $p > 0.05 \Rightarrow$ residuals look normal (✓).
- Durbin–Watson: $\approx 2 \Rightarrow$ no autocorrelation (✓).
- Breusch–Pagan: $p > 0.05 \Rightarrow$ constant variance (✓).