

Regression Results Validation Checklist

Collin Zoeller

Model Specification Checks

- Correct Model Type:** Ensure the chosen model (OLS, fixed effects, Poisson, logit, etc.) is appropriate for the data.
- Dependent Variable Appropriateness:** Confirm that the dependent variable aligns with the research question and is measured correctly.
- Independent Variable Selection:** Verify that key control variables are included and justified based on theoretical or empirical reasoning.
- Functional Form:** Check whether the functional form (e.g., linear, log-linear, quadratic) appropriately captures relationships in the data.
- Interaction Terms:** Ensure that any interaction terms are meaningful and properly interpreted.
- Fixed Effects vs. Random Effects:** If using panel data, confirm whether fixed effects or random effects are correctly chosen based on the Hausman test or theoretical justification.
- Assumption Checks:** Beyond testing for multicollinearity and heteroskedasticity, ensure that other regression assumptions, such as linearity and independence of errors, are met.

Estimation Checks

- Convergence of Estimation:** Check whether the estimation algorithm successfully converged.
- Collinearity:** Examine variance inflation factors (VIF) to identify multicollinearity issues.
- Sample Size Adequacy:** Ensure that the number of observations is sufficient to justify the inclusion of independent variables.
- Correct Standard Errors:** Use robust standard errors if heteroskedasticity is present.
- Clustered Standard Errors:** If applicable, confirm that standard errors are clustered at the appropriate level.
- Endogeneity Issues:** Check for potential omitted variable bias, simultaneity, or measurement error.

Coefficient and Significance Checks

- Significance Levels:** Ensure coefficients are significant at conventional levels (1%, 5%, 10%), if applicable.
- Sign of Coefficients:** Compare coefficient signs with theoretical expectations.
- Magnitude and Interpretation:** Assess whether coefficients have economically meaningful interpretations.
- Effect Sizes:** Compare effect sizes across models to verify robustness.
- Confidence Intervals:** Examine confidence intervals to assess precision.

Model Fit and Diagnostic Checks

- R^2 , Adjusted R^2 , Pseudo- R^2 :** Check model explanatory power for OLS regressions, including Pseudo- R^2 for logit/probit models.

- Log-Likelihood and AIC/BIC:** Compare log-likelihood values and AIC/BIC for model selection.
- Heteroskedasticity:** Test for heteroskedasticity using the Breusch-Pagan or White test.
- Serial Correlation:** For time series and panel data, check for autocorrelation (e.g., Durbin-Watson, Wooldridge test).
- Functional Form Misspecification:** Conduct the Ramsey RESET test to detect omitted nonlinearities.

Residual Diagnostics

- Normality of Residuals:** Check whether residuals are approximately normally distributed.
- Residual Plots:** Inspect residual-vs-predicted plots to detect patterns.
- Leverage and Influence:** Identify high-leverage points using Cook's Distance or DFBETA.
- Homoskedasticity:** Verify that residual variance remains stable across fitted values.
- Outliers and Influential Observations:** Detect and assess the impact of outliers.

Robustness and Sensitivity Analysis

- Alternative Model Specifications:** Run models with different variable selections and compare results.
- Subsample Analysis:** Test for stability of results across different subgroups.
- Placebo Tests:** If applicable, conduct placebo tests to rule out spurious relationships.
- Instrumental Variables Validity:** If IVs are used, check for weak instruments using the first-stage F -statistic.
- Stability Over Time:** For panel data, check whether coefficients remain stable over time.

Presentation and Interpretation Checks

- Clear Tables and Figures:** Ensure results are well-formatted and clearly labeled.
- Correct Units and Transformations:** Verify that variable transformations (e.g., logs, percent changes) are consistently reported.
- Comparability Across Models:** Ensure coefficient estimates are comparable across model variations.
- Transparency in Reporting:** Clearly state estimation methods, standard errors, and key assumptions.
- Policy and Practical Implications:** If applicable, discuss the broader relevance of results.

Final Review

- Sanity Checks:** Ensure results align with economic theory and prior empirical findings.
- Replication and Reproducibility:** Verify that results can be replicated using the same dataset and code.
- Code and Documentation:** Ensure analysis scripts are well-documented and reproducible.
- Peer Review:** Have another researcher review key findings for additional validation.