Regressions

Noé Debrois

June 16, 2024

Multivariate Regressions

- **Generalizing the Simple Model**: Extend the simple one-dimensional regression model to accommodate multiple independent variables. For example, the number of cars sold may depend on factors like car prices, public transport prices, petrol prices, and public concern about global warming.
- **Matrix Notation**: Represent the multiple regression model using vectorial notation: $y = X\beta + u$, where y is the dependent variable, X is the matrix of independent variables, β is the vector of coefficients, and u is the error term.

Multiple Regression and the Constant Term

- **Inclusion of the Constant Term**: The regression equation includes a constant term represented as a column of ones in the X matrix. The general form of the equation is $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_{r+1} x_{r+1,t} + u_t$.

Ordinary Least Squares (OLS) Estimator

- **Parameter Estimation**: The OLS method is used to estimate the parameters β by minimizing the sum of squared residuals. The OLS estimator is BLUE (Best Linear Unbiased Estimator) if certain assumptions hold.
- **Assumptions of the CLRM (Classical Linear Regression Model)**:
  1. The error term has zero mean $E(u_t) = 0$.
  2. Homoscedasticity $Var(u_t) = \sigma^2$.
  3. No autocorrelation $Cov(u_i, u_j) = 0$.
  4. The independent variables are non-stochastic.
  5. The error terms are normally distributed.

Goodness of Fit: R^2 and Adjusted R^2

- R^2: Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. R^2 values range from 0 to 1.

- **Adjusted R^2**: Adjusts R^2 for the number of predictors in the model, providing a more accurate measure when comparing models with different numbers of independent variables.

Problems with R^2 as a Goodness of Fit Measure

- **Limitations**:
  1. R^2 changes if the dependent variable is reparameterized.
  2. R^2 never decreases with the addition of more regressors.
  3. High R^2 values are common in time series regressions, which can be misleading.

The F-Test for Overall Significance

- **Testing Hypotheses**: The F-test compares the fit of a restricted model (with fewer predictors) to an unrestricted model. If the F-statistic is significantly large, we reject the null hypothesis that the restricted model is sufficient.

Violation of CLRM Assumptions

- **Detection and Consequences**: Violations of assumptions can lead to biased or inefficient estimators. Various tests (like the F-test and Chi-square test) and methods (like transforming variables or adding more data) are used to detect and correct these issues.

Multicollinearity

- **Solutions**: To address multicollinearity, one can drop collinear variables, transform them into ratios, or collect more data. Multicollinearity does not affect the goodness of fit but makes the estimation of individual coefficients unreliable.

Functional Form Misspecification

- **Ramsey's RESET Test**: This test checks if the functional form of the regression model is correct by adding higher-order terms of the fitted values. If the test indicates misspecification, transforming the data (e.g., using logarithms) might help.

Parameter Stability

- **Testing Stability**: Use parameter stability tests, like the Chow test, to check if regression coefficients are stable over time. This involves splitting the data into sub-periods and comparing the models' residual sum of squares (RSS).