Linear regression and time series econometrics

ROHU00120 “Data and text mining based on artificial intelligence research applied supporting accounting and financial decision-making (using R statistics and Python)”

János Szenderák

UNIVERSITY of DEBRECEN

Introduction

- Regression analysis is one of the central methods of econometrics.
- Provides quantitative relationships between variables.
- Foundation of hypothesis testing, prediction, and policy evaluation.

What rergession Does

- Measures relationships: e.g. income $\sim$ education, demand $\sim$ price.
- Separates systematic part (explained by regressors) from random part (error).
- Allows forecasting and policy analysis.

The Regression Model

The basic form of the linear regression model.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- y_i: dependent (explained) variable
- x_i: explanatory variable
- u_i: error term capturing unobserved factors

Role of the Error Term

The error term is the most critical part of the equation. Why do we need it?

- Represents measurement errors, omitted variables, and randomness.
- Assumptions on u_i determine whether OLS gives valid results.
- Example: income explained by education, but ability and effort are hidden in u_i.

OLS Principle

Choose $\hat{\beta}_0, \hat{\beta}_1$ to minimize squared errors:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Residuals

$$\hat{u}_i = y_i - \hat{y}_i$$

- Residuals estimate unobserved error terms.
- OLS ensures that their mean is zero.

Deriving the Estimators

Normal equations:

$$\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \sum x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Solutions:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Interpretation of Coefficients

- β_0: expected y when $x = 0$ (sometimes not meaningful).
- β_1: average change in y if x increases by 1.
- Example: each extra year of education raises expected income by β_1 units.

Sampling Variation

- Estimates $\hat{\beta}$ vary with samples.
- We describe their distribution with standard errors.
- Standard error = estimated standard deviation of $\hat{\beta}$.

UNIVERSITY of DEBRECEN

Testing Individual Parameters

$$H_0 : \beta_j = 0 \quad vs. \quad H_1 : \beta_j \neq 0$$

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

If $|t|$ large, reject H_0.

Confidence Intervals

$$\hat{\beta}_j \pm t_{\alpha/2} \cdot se(\hat{\beta}_j)$$

Interpretation: with 95% confidence, the true β_j lies in this interval.

Joint Significance: F-Test

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k-1)}$$

Tests whether a group of variables contributes jointly to the model.

Decomposition of Variance

$$SST = SSE + SSR$$

- SST: total variation
- SSE: explained by regression
- SSR: residual

R^2

$$R^2 = 1 - \frac{SSR}{SST}$$

Measures proportion of variance explained by the model.

Adjusted R^2

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

Accounts for number of regressors.

CLM Assumptions

1. Linearity in parameters
2. Random sampling
3. No perfect multicollinearity
4. Zero conditional mean: $E(u|X) = 0$
5. Homoscedasticity: $Var(u|X) = \sigma^2$
6. Normality (for small samples)

Gauss–Markov Conditions

- The first five conditions are also called the Gauss–Markov conditions.
- Condition 6) is not critical, since in the case of a sufficiently large sample it can be omitted (but in the case of a small sample it is important, which is becoming increasingly rare in the economic and financial field).

Importance of the Conditions

These conditions are extremely important for verifying two factors.

- We want the estimator function to be unbiased.
- Since numerous unbiased estimators can be constructed, we further want the variance of the estimator function to be the smallest (that is, to be the most efficient).

If conditions 1–4 are satisfied, the estimator is unbiased. If conditions 1–5 are satisfied, the OLS estimation is the Best Linear Unbiased Estimator (BLUE).

UNIVERSITY of DEBRECEN

Gauss–Markov Theorem

Gauss–Markov theorem

If CLM conditions 1–5 are satisfied, the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ give the BLUE of the population parameters $\beta_0, \beta_1, \dots, \beta_k$.

Implication of the Theorem

This theorem states that among linear and unbiased estimators, the variance of the OLS estimator is the smallest.

It makes an extremely strong statement: if these conditions are satisfied, then no estimator function will perform better than OLS.

Gauss–Markov Theorem

- OLS is BLUE: Best Linear Unbiased Estimator.
- What does this mean?

Specification Issues

Omitted Variables

- Leaving out relevant regressors causes bias.
- Especially severe if omitted variable is correlated with included ones.

Irrelevant Variables

- Adding unnecessary regressors does not bias estimates.
- But it increases their variance.

UNIVERSITY of DEBRECEN

Specification Issues

Multicollinearity

- High correlation among regressors makes estimates imprecise.
- Variance Inflation Factor (VIF) often used as a diagnostic.

Heteroskedasticity

- Non-constant error variance invalidates standard errors.
- Solution: robust (heteroskedasticity-consistent) SEs.

Level–Level Model

$$y = \beta_0 + \beta_1 x + u$$

β_1: unit change in y per unit change in x.

Log–Level Model

$$\ln y = \beta_0 + \beta_1 x + u$$

β_1: approximate percent change in y per unit change in x.

Level–Log Model

$$y = \beta_0 + \beta_1 \ln x + u$$

β_1: change in y for 1% change in x.

Log–Log Model

$$\ln y = \beta_0 + \beta_1 \ln x + u$$

β_1: elasticity of y with respect to x.

Measurement Error

If observed $x^* = x + v$, then

$$y = \beta_0 + \beta_1 x^* + (u - \beta_1 v)$$

This correlation between regressor and error leads to attenuation bias.

Firm Loan Example

- Dependent: loan amount.
- Explanatory: revenue, size, age, ownership, industry.

Conclusions

- OLS is the fundamental tool of econometrics.
- Inference relies on assumptions: unbiasedness and efficiency.
- Functional forms and time series require extensions.

Thank you!