

Lecture 25: Limitations of ridge regression; summary of key formulas

Econometrics ■ Unit II: Multicollinearity and heteroskedasticity

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

Previous lecture

Lecture 24: Why ridge regression works; choice of the biasing parameter; ridge trace and HKB.

Today

- Ridge regression
 - Limitations of ridge regression
- Summary of key formulas

1. The choice of δ depends on the data. The properties derived above (bias, variance, MSE) assume that δ is a fixed constant, so they hold only approximately when δ is chosen from the data.
2. δ can be any value in $(0, \infty)$, so many values must be explored. With a computer this is not a serious problem.
3. The choice from a ridge trace is subjective and may not be unique; different people may choose different values. However, because δ is chosen where the estimates are stable, a small change in δ does not change them much. Different formulas (such as the HKB choice) may give quite different values.

4. Different coefficients may stabilise at different values of δ , and a compromise may be hard. Then *generalized ridge regression*, which uses a separate δ_j for each canonical coefficient, can be used.
5. No exact theory is available for testing hypotheses and constructing confidence intervals with ridge estimates.

- Standardised data: $X'X$ is the correlation matrix R and $X'y$ the vector of correlations with y .
- Two regressors: $\text{Var}(b_1) = \text{Var}(b_2) = \sigma^2/(1 - r^2)$,
 $\text{Cov}(b_1, b_2) = -r\sigma^2/(1 - r^2)$.
- $C_{jj} = \text{VIF}_j = 1/(1 - R_j^2)$; $\text{Var}(b_j) = \sigma^2 \text{VIF}_j$; confidence interval length grows by $\sqrt{\text{VIF}_j}$.
- $E(b'b) = \beta'\beta + \sigma^2 \sum_j 1/\lambda_j$.
- $0 \leq |R| \leq 1$; condition number $\kappa = \lambda_{\max}/\lambda_{\min}$ (100 and 1000 as guide values); condition indices $\kappa_j = \lambda_{\max}/\lambda_j$.
- Variance proportions: $\text{VIF}_i = \sum_j v_{ij}^2/\lambda_j$, $p_{ij} = (v_{ij}^2/\lambda_j)/\text{VIF}_i$.

- Principal components: $Z = XV$, $Z'Z = \Lambda$, $\hat{\beta}_{pc} = V_{k-r}\hat{\alpha}_{k-r}$.
- Ridge:
 - $\hat{\beta}_{\text{ridge}} = (X'X + \delta I)^{-1}X'y$;
 - Bias = $-\delta(X'X + \delta I)^{-1}\beta$;
 - Var = $\sigma^2(X'X + \delta I)^{-1}X'X(X'X + \delta I)^{-1}$;
 - MSE = $\sigma^2 \sum \lambda_j/(\lambda_j + \delta)^2 + \delta^2 \sum \alpha_j^2/(\lambda_j + \delta)^2$;
 - better than OLS for $0 < \delta < \sigma^2/\beta'\beta$;
 - HKB choice $\hat{\delta} = k\hat{\sigma}^2/b'b$.

Practice

- Exercise 3.8 (Exercises and Solutions booklet).
- R practical P06: Ridge regression for the Hald cement data.

Next lecture

Lecture 26: Heteroskedasticity: concept, sources and consequences for OLS.