# PMDS504L: Regression Analysis and Predictive Models

## LS Estimation on Restriction On Parameters

Dr. Jisha Francis

Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology
Vellore Campus, Vellore - 632 014
India

# Introduction

- One of the basic objectives in statistical modeling is to find good estimators of parameters.
- In multiple linear regression models $Y = X\beta + \epsilon$, the Ordinary Least Squares Estimator (OLSE)

$$\hat{\beta} = (X'X)^{-1}X'Y$$

is the Best Linear Unbiased Estimator (BLUE) of $\beta$.

# Improving OLSE

- Several approaches have been explored to improve the OLSE.
- One approach involves utilizing extraneous or prior information.
- Prior information may be available from:
    1. Theoretical considerations.
    2. Past experimental experience.
    3. Empirical investigations.
    4. Extraneous sources.

# Types of Constraints in Regression

Prior information can be expressed in the form of:

1. Exact linear restrictions.
2. Stochastic linear restrictions.
3. Inequality restrictions.

# Introduction to Exact Linear Restrictions

- Prior information about regression coefficients may be available from extraneous sources.
- This information can be expressed in the form of exact linear restrictions as:

$$\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\beta}$$

- Where:
  - $\boldsymbol{r}$ is a $q \times 1$ vector.
  - $\boldsymbol{R}$ is a $q \times k$ matrix with $\text{rank}(R) = q$ and $q < k$.
  - The elements in $\boldsymbol{r}$ and $\boldsymbol{R}$ are known.

# Examples of Exact Linear Restrictions

**Example 1: Two restrictions with** $k = 6$

$$\beta_2 = \beta_4$$

$$\beta_3 + 2\beta_4 + \beta_5 = 1$$

$$\boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 \end{bmatrix}$$

# Examples of Exact Linear Restrictions (cont.)

**Example 2: Restriction with $k = 3$**

$$\beta_2 = 3$$

$$\boldsymbol{r} = \begin{bmatrix} 3 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

# Incorporating Prior Information

- The Ordinary Least Squares Estimator (OLSE) is given by:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- OLSE does not use prior information and does not satisfy the restrictions $\boldsymbol{r} \neq \boldsymbol{R}\boldsymbol{b}$.
- Objective: Combine sample information and prior information to derive an improved estimator of $\beta$.

# Restricted Least Squares Estimation

**Restricted Least Squares Estimation** enables the simultaneous use of sample and prior information. The objective is to choose $\boldsymbol{\beta}$ such that the error sum of squares is minimized subject to linear restrictions:

$$\boldsymbol{R}\beta = \boldsymbol{r}$$

This problem can be solved using the Lagrangian multiplier technique.

# Lagrangian Function

The Lagrangian function is defined as:

$$S(\beta, \boldsymbol{\lambda}) = (\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta) - 2\lambda'(\boldsymbol{R}\beta - \boldsymbol{r})$$

where $\boldsymbol{\lambda}$ is a $k \times 1$ vector of Lagrange multipliers.

# Differentiation Results

Using the results that if $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors and $\boldsymbol{A}$ is a suitable matrix, then:

$$\frac{\partial(\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a})}{\partial \boldsymbol{a}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{a},$$
$$\frac{\partial(\boldsymbol{a}'\boldsymbol{b})}{\partial \boldsymbol{a}} = \boldsymbol{b}$$

we have:

$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = 2X'X\beta - 2X'y - 2R'\lambda = 0 \quad (*)$$
$$\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \boldsymbol{R}\beta - \boldsymbol{r} = 0$$

# Restricted Regression Estimator

Substituting $\boldsymbol{\lambda}$ back into equation (*), we get:

$$\beta = \boldsymbol{b} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}](\boldsymbol{R}\boldsymbol{b} - \boldsymbol{r})$$

This is termed the **restricted regression estimator** of $\beta$.

# Introduction to Stochastic linear restrictions

- In practical regression models, the auxiliary or prior information may involve randomness.
- This randomness leads to the formulation of **stochastic linear restrictions**.
- The standard form of such restrictions is given by:

$$r = R\beta + V$$

where:

- $r$ is a $q \times 1$ vector (known),
- $R$ is a $q \times k$ matrix (known),
- $V$ is a $q \times 1$ vector of random errors.

# Assumptions

- The vector $V$ captures the randomness in the prior information.
- The following assumptions are made:

$$E(V) = 0$$
$$E(VV') = \psi$$
$$E(V\epsilon') = 0$$

- Here:
  - $\psi$ is a known $q \times q$ positive definite matrix.
  - $\epsilon$ is the disturbance term in the multiple regression model $y = X\beta + \epsilon$.

# Important Observation

- The expectation of the prior restriction becomes:

$$E(r) = R\beta$$

- This indicates that the stochastic restriction generalizes the exact restriction $r = R\beta$ by allowing some random variation.
- This approach is particularly useful when prior information is imprecise.

## Conclusion on Stochastic linear restrictions

- Stochastic linear restrictions provide a more flexible framework for incorporating prior information.
- They are particularly relevant when the exact information assumption does not hold.

# Introduction to Pure and Mixed Regression Estimation

- Consider the multiple regression model:

$$y = X\beta + \epsilon$$

- $n$ observations and $k$ explanatory variables $X_1, X_2, \ldots, X_k$.
- The ordinary least squares (OLS) estimator of $\beta$ is:

$$\hat{\beta} = (X'X)^{-1}X'y$$

- This is termed the **pure estimator**.

# Problem with Pure Estimator

- The pure estimator $\hat{\beta}$ does not satisfy the restriction:

$$r = R\beta + V$$

- Objective: Obtain an estimator of $\beta$ utilizing the stochastic restrictions.
- This estimator should satisfy both prior and sample information.

# Combining Prior and Sample Information

- Write the information as:

$$y = X\beta + \epsilon$$

$$r = R\beta + V$$

- Jointly expressed as:

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \epsilon \\ V \end{bmatrix}$$

- Define:

$$a = \begin{bmatrix} y \\ r \end{bmatrix}, \quad A = \begin{bmatrix} X \\ R \end{bmatrix}, \quad w = \begin{bmatrix} \epsilon \\ V \end{bmatrix}$$

# Assumptions

- $E(\epsilon) = 0, \quad E(V) = 0$
- $E(\epsilon\epsilon') = \sigma^2 I_n, \quad E(VV') = \psi$
- $E(\epsilon V') = 0$
- Jointly:

$$E(w) = 0$$

$$E(ww') = \Omega = \begin{bmatrix} \sigma^2 I_n & 0 \\ 0 & \psi \end{bmatrix}$$

- The disturbances $w$ are non-spherical (heteroskedastic).

# Generalized Least Squares Estimation

- Applying generalized least squares (GLS) to the model:

$$a = A\beta + w, \quad E(w) = 0, \quad E(ww') = \Omega$$

- The GLS estimator of $\beta$ is given by:

$$\hat{\beta} = (A'\Omega^{-1}A)^{-1}A'\Omega^{-1}A$$

- This estimator is more efficient than the OLS estimator.

# Conclusion to Pure and Mixed Regression Estimation

- Pure estimators may not satisfy stochastic restrictions.
- Mixed regression estimation incorporates both sample and prior information.
- The use of GLS yields efficient estimators when disturbances are non-spherical.

# What is Multicollinearity?

Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated with each other. This makes it difficult to estimate the relationship between the predictors and the dependent variable.

Let's explore some real-world examples where multicollinearity can arise.

# Example 1: House Price Prediction

**Predictors:** Square footage, number of rooms, and house age.

**Multicollinearity Issue:**

- Square footage and number of rooms are highly correlated since larger houses tend to have more rooms.
- Including both in the model causes unstable coefficient estimates.

# Example 2: Employee Salary Prediction

**Predictors:** Years of experience, education level, and number of certifications.

**Multicollinearity Issue:**

- Years of experience and education level are often correlated, as experienced professionals may have higher educational qualifications.
- This can lead to unreliable coefficients in salary prediction models.

# Example 3: Health Outcomes Prediction

**Predictors:** Age, blood pressure, and cholesterol levels.

**Multicollinearity Issue:**

- Blood pressure and cholesterol levels are correlated as both are risk factors for cardiovascular disease.
- Including both may make it difficult to interpret their individual effects.

# Example 4: Customer Spending Behavior

**Predictors:** Income, education level, and occupation.

**Multicollinearity Issue:**

- Income and education level are correlated since individuals with higher education tend to earn more.
- This causes redundancy in models predicting customer spending.

# Example 5: Weather Data Analysis

**Predictors:** Temperature, humidity, and dew point.

**Multicollinearity Issue:**

- Temperature and dew point are highly correlated as both measure atmospheric moisture.
- This can affect models predicting rainfall or weather patterns.

# Consequences of Multicollinearity

- **Unreliable Estimates:** The estimated coefficients become unstable and sensitive to small changes in the data.
- **Inflated Standard Errors:** This results in failed hypothesis testing, making it seem like predictors have no effect when they might.
- **Difficulty in Interpretation:** It becomes challenging to understand the effect of each predictor due to their overlap.

# Detecting Multicollinearity

Several methods to detect multicollinearity include:

- **Correlation Matrix:** Calculate correlations between all pairs of independent variables. High correlations (near 1 or -1) indicate a problem.
- **Variance Inflation Factor (VIF):** A VIF greater than 10 suggests significant multicollinearity.
- **Condition Index:** A condition index over 30 indicates potential multicollinearity.

## Dealing with Multicollinearity

Solutions to handle multicollinearity:

- **Remove Highly Correlated Predictors:** Remove one of the highly correlated variables from the model.
- **Combine Predictors:** Combine correlated variables into a single composite variable.
- **Principal Component Analysis (PCA):** Transform correlated variables into uncorrelated components.
- **Ridge or Lasso Regression:** These techniques penalize large coefficients, reducing multicollinearity.

# Example of Multicollinearity in Action

**Problem:** Predicting house prices based on square footage and number of rooms. These variables are often correlated. Including both can cause multicollinearity and lead to unstable regression coefficients.

By removing or combining predictors, the model can become more stable and interpretable.

# When Is Multicollinearity Not a Problem?

- **Prediction Focus:** If the goal is prediction, multicollinearity may not be an issue as long as the model performs well.
- **Using Regularization:** Ridge or Lasso regression can handle multicollinearity while maintaining the predictive power.

# Key Takeaways

- Multicollinearity makes it hard to separate the effects of correlated predictors.
- It can cause unreliable estimates, inflated standard errors, and interpretation problems.
- Use correlation matrices, VIF, and condition indices to detect it.
- Solutions include removing, combining predictors, PCA, or using regularization methods.

# References


- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis, Fifth Edition. Wiley.

# Thank You!

Thank you for your attention!