

Multiple Response Regression: Equity Fund Profitability and Sustainability Modelling

Raul Unnithan, Durham University

February 26, 2025

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Why is Multiple Response Regression needed?

- Single-response regression can extend to cover multiple responses.
- What happens when these responses are correlated?
- MRR accounts for these correlations, improving predictions.

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Multiple Response Linear Regression

Definition

$$\mathbf{Y}_{(n \times m)} = \mathbf{X}_{(n \times p)} \mathbf{B}_{(p \times m)} + \mathbf{E}_{(n \times m)}$$

with

$$E(e_i) = 0 \quad \text{and} \quad \text{Cov}(e_i, e_k) = \sigma_{ik} \mathbf{I} \quad i, k = 1, 2, \dots, m$$

where, \mathbf{I} is the identity matrix, n is the number of observations, m is the number of responses and p is the number of predictors.

Key Ideas:

Covariance Structure: $\text{Cov}(\mathbf{Y}) = \mathbf{B} \text{Cov}(\mathbf{X}) \mathbf{B}^T + \text{Cov}(\mathbf{E})$

Shared error covariance \Rightarrow Response covariance considered!

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example**
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Example of Multiple Response Linear Regression

Example

Take $n = 4$, $p = 3$, and $m = 2$, in matrix form we get:

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \\ y_{41} & y_{42} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} \cdot \begin{bmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \\ e_{31} & e_{32} \\ e_{41} & e_{42} \end{bmatrix},$$

where $E(e_i) = 0$ and $\text{Cov}(e_i, e_k) = \sigma_{ik} \mathbf{I}$ for $i, k = 1, 2$.

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)**
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Multiple Analysis of Variance (MANOVA)

Definition

MANOVA assesses the impact of the predictors on the response variation.

Theorem

MANOVA uses 3 matrices: represented as \mathbf{W} , \mathbf{H} and \mathbf{T} , respectively:

$$\mathbf{W} = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}),$$

$$\mathbf{H} = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}),$$

$$\mathbf{T} = \mathbf{W} + \mathbf{H} = (\mathbf{Y} - \bar{\mathbf{Y}})^\top (\mathbf{Y} - \bar{\mathbf{Y}}),$$

where \mathbf{Y} is the $n \times m$ matrix of observed response values, $\hat{\mathbf{Y}}$ is the $n \times m$ matrix of predicted response values from the model and $\bar{\mathbf{Y}}$ is the overall mean response vector, $m \times 1$, given by $\bar{\mathbf{Y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{Y}_i$.

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test**
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Wilks' Lambda Test

Definition

MANOVA has multiple tests it can use, but Wilks' Lambda was chosen:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|},$$

where $|\cdot|$ is the determinant.

Wilks' Lambda F-Statistic

Definition

Wilks' Lambda is changed to an F-statistic for hypothesis testing:

$$F = \frac{(1 - \Lambda)/m}{\Lambda/(n - m - 1)},$$

where m is the number of response variables and n is the total sample size.

Hypotheses:

H_0 : The predictors do not significantly explain response mean variation.

H_1 : The predictors do significantly explain response mean variation.

Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection**
- 7 Conclusion and Further Study

Sequential MANOVA

Definition

- Sequential MANOVA extends MANOVA to nested models.
- At each step, compute the Wilks' Lambda (Λ_j) value for each model, specifically: $\Lambda_j - \Lambda_{j+1}$.

How is Sequential MANOVA used?

- Sequential MANOVA is combined with stepwise selection
- This selects the best Multiple Response Linear Regression Model.

Results

Model	ANRMSE
Multiple Response Linear Regression using all the Predictors	0.7606
Forward Stepwise Selection - Sum of all the Predictors	0.7606
Backward Stepwise Selection - One predictor removed	0.7924
Bidirectional Stepwise Selection - Sum of the Predictors	0.7606
Bidirectional Stepwise Selection - including Interaction Terms	0.5613
Bidirectional Stepwise Selection - including Non-Linear Terms	0.6893

Table 1. Selection Methods and their ANRMSE Values

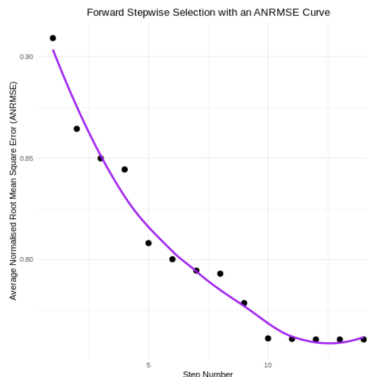


Table of Contents

- 1 Introduction
- 2 Multiple Response Linear Regression
- 3 Multiple Response Linear Regression Example
- 4 Multiple Analysis of Variance (MANOVA)
- 5 Wilks' Lambda Test
- 6 Sequential MANOVA Stepwise Selection
- 7 Conclusion and Further Study

Conclusion and Further Study

- This model is a good start, but more complex models are needed.
- The whole report is a comparison across different regression models.
- Use lots of different datasets or compare more regression models.

Any Questions?