Equity Fund Profitability and Sustainability Modelling using Multiple Response Regression

Raul Unnithan, Durham University

February 19, 2025

Table of Contents

Table of Contents

Introduction

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

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

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

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.

Note

$\mathbf{B} = (\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{Y})$ as from single-response regression...
Shared error correlation structure $\Rightarrow$ Response intercorrelation considered!

Table of Contents

Example of Multiple Response Linear Regression

Example

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

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

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

Table of Contents

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:*

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

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

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

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* $\bar{\mathbf{Y}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{Y}_i$.

These matrices correspond to SSE, SSR and SST, respectively!

Table of Contents

Willks' Lambda Test

Definition

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

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

where $|.|$ 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.

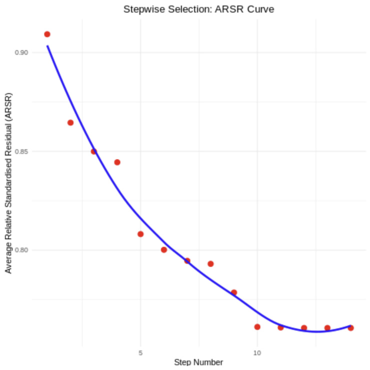
Hypothesis

H_0: The predictor does not significantly explain response mean variation.
H_1: The predictor does significantly explain response mean variation.

Table of Contents

MANOVA Stepwise Selection

- Sequential MANOVA: at each step, compute Wilks' Lambda (Λ_j) for each model, specifically: $\Lambda_j - \Lambda_{j+1}$.
- Functions for Implementing this:
  - `calculate_arsr`
  - `add_predictors`
  - `remove_predictors`
  - `stepwise_multivariate`
- Evaluate results at the end using the average relative standardised residual.



Forward Stepwise Selection Plot

Table of Contents

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?