



Multiple Response Regression: Equity Fund Profitability and Sustainability Modelling

Raul Unnithan

Durham University

Supervisor: Richard Crossman



Introduction and Motivation

Single-response regression models 1 response against 1 predictor. Single-response regression can be extended to multiple predictors, but **what if we need to model more than one response?**

One approach is to run several independent single-response models. However, when responses are correlated, this approach ignores their covariance, leading to suboptimal predictions. Multiple-response regression (MRR) addresses this, improving predictions.

However, **selecting the most relevant predictors** needs to be considered. This report looks at **(Sequential) MANOVA Stepwise Selection** and applies it to an **equity fund dataset** with 2 correlated responses. Its predictions are evaluated on the **ARSR**, the average normalised root mean square error divided by standard deviation.

Multiple Response Linear Regression Model

A **Multiple Response Linear Regression (MRLR) Model** is defined as:

$$\mathbf{Y}_{(n \times m)} = \mathbf{X}_{(n \times p)} \mathbf{B}_{(p \times m)} + \mathbf{E}_{(n \times m)} \quad (1)$$

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

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.[1]

The key change in MRR is that the errors are multivariate normally distributed: $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{E}})$, allowing it to consider response covariance in modelling, unlike single-response multiple linear regression.

Multiple Response Linear Regression Example

Here is a simple example to further this understanding. Suppose $n = 4$, $p = 3$, and $m = 2$, Equation 1 becomes:

$$\mathbf{Y}_{(4 \times 2)} = \mathbf{X}_{(4 \times 3)} \mathbf{B}_{(3 \times 2)} + \mathbf{E}_{(4 \times 2)},$$

which can be written in matrix form as:

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \\ y_{41} & y_{42} \end{bmatrix}_{(4 \times 2)} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix}_{(4 \times 3)} \begin{bmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}_{(3 \times 2)} + \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \\ e_{31} & e_{32} \\ e_{41} & e_{42} \end{bmatrix}_{(4 \times 2)}$$

where $E(e_i) = 0$ and $\text{Cov}(e_i, e_k) = \sigma_{ik} \mathbf{I}$ for $i, k = 1, 2$. The Covariance of the Errors, $\Sigma_{\mathbf{E}}$, is calculated here as:

$$\Sigma_{\mathbf{E}} = \frac{1}{n - p} \mathbf{E}^{\top} \mathbf{E} \Rightarrow \Sigma_{\mathbf{E}} = \mathbf{E}^{\top} \mathbf{E}.$$

Therefore, $\Sigma_{\mathbf{E}}$ is a 2 x 2 dimensional matrix with the variance and covariance of the errors for the two responses.

Multiple Analysis of Variance

Multivariate Analysis of Variance (MANOVA) extends ANOVA to analyse multiple response variables simultaneously. Unlike separate univariate ANOVAs, MANOVA considers the **response covariance matrix** through the total response variation, which is an **unscaled form** of it.

The total response variation, \mathbf{T} , is partitioned into 2 components:

1. The **explained** Sum of Squares and Cross-Products (SSCP) matrix, \mathbf{H} , which is the response variation accounted for by predictors.
2. The **unexplained** SSCP matrix, \mathbf{W} , which is the response variation not accounted for by predictors.[3]

As specified, these 2 matrices relate via: $\mathbf{T} = \mathbf{H} + \mathbf{W}$.

The **explained SSCP matrix**, \mathbf{H} can be calculated as follows:

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

where $\bar{\mathbf{Y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{Y}_i$ is the mean response vector.

The **unexplained SSCP matrix**, \mathbf{W} , in turn, is derived as follows:

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

where \mathbf{Y} and $\hat{\mathbf{Y}}$ are the observed and predicted response matrices.

MANOVA Forward Stepwise Selection Plot

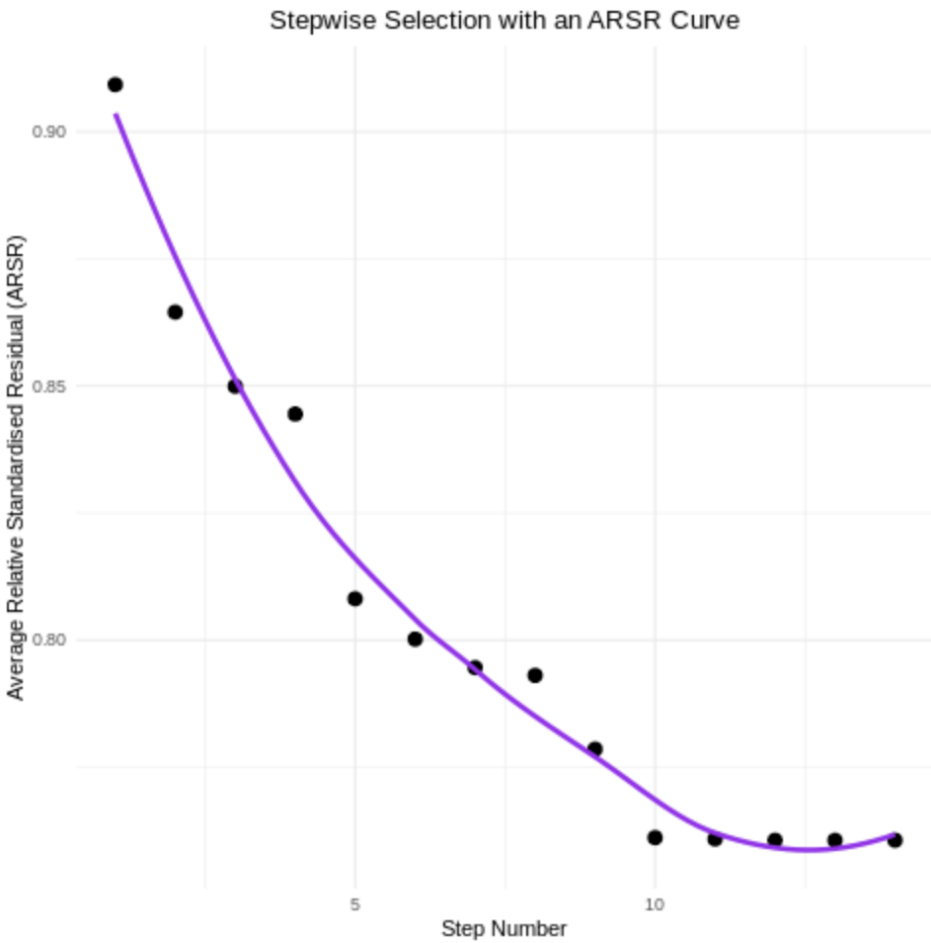


Figure 1: ARSR across Forward Stepwise Selection

This model seeks to reduce this ARSR value as much as possible. Here, ARSR constantly decreases as more predictors are added.

Wilks' Lambda Test

To test significance, MANOVA uses multiple tests. This report applies **Wilks' Lambda**, Λ , to measure how much the numerical predictors impact response variance:

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

where $|\cdot|$ denotes the determinant of each SSCP matrix. A smaller Λ (closer to 0) indicates the predictors explain more response variance, whereas a larger Λ (closer to 1) suggests the opposite.

Wilks' Lambda can then be converted into an **F-statistic** for hypothesis testing:

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

where m is the number of responses and n is the number of observations.[3] A significant F-test ($p < 0.05$) indicates that the predictors significantly affect the variation across the responses.[2]

Sequential MANOVA extends ANOVA by testing predictors in a nested sequence, computing Wilks' Lambda at each step:

$$\Lambda_1 \geq \Lambda_2 \geq \dots \geq \Lambda_m.[3]$$

A significant drop in Λ suggests the added predictor reduces unexplained variance a lot, meaning it should be a part of the model.[2] Combining this with stepwise selection forms the model.

R Implementation and Early Results

The MANOVA stepwise selection methods were coded in R. Table 1 outlines the performance of these methods on the equity fund dataset:

Model	ARSR
Full Model - MRLR 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 with Interaction Terms	0.5613
Bidirectional Stepwise Selection with Non-Linear Terms	0.6893

Table 1. Selection Methods and their ARSR Values

References

- [1] Richard Johnson and Dean Wichern. Multivariate linear regression models: Section 7.7. In *Applied Multivariate Statistical Analysis: Pearson New International Edition*, pages 360–429. Pearson Education, Limited, 6th edition, 2013. ISBN 9781292024943. URL <https://ebookcentral.proquest.com/lib/durham/detail.action?docID=5174865>.
- [2] Newsom. Multivariate analysis of variance. *Psy 522/622 Multiple Regression and Multivariate Quantitative Methods*, 2024. Winter 2024 Lecture Notes.
- [3] STAT 505 Pennsylvania State University. Lesson 8: Multivariate analysis of variance (manova), 2024. URL <https://online.stat.psu.edu/stat505/lesson/10>.