Statistica Sinica **33** (2023), 1-21
doi:https://doi.org/10.5705/ss.202020.0352

SCALED PARTIAL ENVELOPE MODEL IN MULTIVARIATE LINEAR REGRESSION

Jing Zhang[1,2], Zhensheng Huang[1] and Lixing Zhu[3,4]

[1]*Nanjing University of Science and Technology,* [2]*Chuzhou University,*
[3]*Beijing Normal University at Zhuhai and* [4]*Hong Kong Baptist University*

Abstract: Inference based on the partial envelope model is variational or non-equivariant under rescaling of the responses, and tends to restrict its use to responses measured in identical or analogous units. The efficiency acquisitions promised by partial envelopes frequently cannot be accomplished when the responses are measured in diverse scales. Here, we extend the partial envelope model to a scaled partial envelope model that overcomes the aforementioned disadvantage and enlarges the scope of partial envelopes. The proposed model maintains the potential of the partial envelope model in terms of efficiency and is invariable to scale changes. Further, we demonstrate the maximum likelihood estimators and their properties. Lastly, simulation studies and a real-data example demonstrate the advantages of the scaled partial envelope estimators, including a comparison with the standard model estimators, partial envelope estimators, and scaled envelope estimators.

Key words and phrases: Dimension reduction, grassmannian, scaled envelope model, partial envelope model, scale invariance.

1. Introduction

The standard multivariate linear regression model with a $p \times 1$ non-stochastic predictor X and an $r \times 1$ stochastic response Y can be represented as

$$Y = \alpha + \beta X + \varepsilon, \tag{1.1}$$

where $\alpha \in \mathbb{R}^r$ is the unknown intercept, $\beta \in \mathbb{R}^{r \times p}$ is the unknown coefficient matrix, and the error vector ε has mean zero and covariance matrix $\Sigma > 0$, and is independent of X. The data involve n independent values Y_i of Y, which are observed at corresponding values X_i of $X (i = 1, \ldots, n)$. In general, we assume that the predictor is centered in the sample. The model is a cornerstone of multivariate statistics. Here, we focus on the interrelation between X and Y using the regression coefficient matrix $\beta \in \mathbb{R}^{r \times p}$. As such, our interest lies in estimating β.

Corresponding author: Zhensheng Huang, School of Science, Nanjing University of Science and Technology, Nanjing, 210094, Jiangsu, P. R. China. E-mail: stahzs@126.com.