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.

Cook, Li and Chiaromonte (2010) introduced response envelopes, and employed a subspace to envelop the material information and eliminate the immaterial variation. This method results in significant efficiency gains when the immaterial variation is larger than the material variation. Based on the work of Cook, Li and Chiaromonte (2010), several papers have extended the idea of enveloping to more general settings, and have proposed new models to accomplish greater efficiency gains (Su and Cook (2011, 2012, 2013); Cook, Helland and Su (2013); Cook and Zhang (2015, 2016, 2018); Khare, Pal and Su (2017); Li and Zhang (2017); Zhang and Li (2017); Ding and Cook (2018); Pan, Mai and Zhang (2019); Zhu and Su (2020)).

The partial envelope model introduced by Su and Cook (2011) results in a parsimonious method for multivariate linear regression when some of the predictors are of special interest. In contrast to the standard model, it has the potential for extensive efficiency gains in the estimation of the coefficients for the chosen predictors. The partial envelope model is a variation of the envelope model proposed by Cook, Li and Chiaromonte (2010), but it pays close attention to part of the predictors. It has looser restrictions and can further enhance efficiency. The envelope estimator decreases to the standard estimator when $r \leq p$ and β has rank r. There is no possibility of an efficiency gain in this case. Nevertheless, the partial envelope removes this restriction, offering gains even when $r \leq p$.

The scaled envelope model proposed by Cook and Su (2013) is scale-unchanging, and achieves efficiency gains except in the case of the initial envelope model. This is achieved by incorporating a scaling matrix into the model, so scale transformations are taken into account during estimation. It maintains the advantages of the original envelope methods in terms of efficiency and is invariant to scale variations. Cook and Su (2016) employed the relationship between partial least squares and envelopes to exploit new method-scaled predictor envelopes that involve predictor scaling into partial least squares-type applications. By estimating suitable scales, the scaled predictor envelope model estimators provide efficiency gains, surpassing those offered by partial least squares, and further decrease prediction errors.

The scaled envelope model pays close attention to all predictors. However, we focus on part of the predictors, looser restrictions, and further improving efficiency. Our research is motivated by the study of Cook and Su (2013). In order to further improve the efficiency of the parameter estimation and keep the scale invariant, we first apply the dimension reduction ideas of the partial envelope model to focus on predictors of special interest. Next, we combine the partial envelopes with the scaled envelopes to form the scaled partial envelope

model, the estimators of which show promising performance in our simulation studies and real-data analysis. From the simulation results, when the envelope subspace is equal to the full space, we find that the standard model and the scaled envelope model have the same standard deviations. In addition, the scaled partial envelope estimators exhibit a remarkable efficiency gain over the partial envelope estimators, and the latter have an obvious efficiency gain over the ordinary least squares estimators and the scaled envelope estimators, regardless of whether the error is normal or nonnormal. In the real-data analysis, we also see the advantages of the scaled partial envelope model. When the part of the response variable Y that is material to all predictors is no less variable than the immaterial part, but the part of Y that is material to part of the predictors is much less variable than the immaterial part, the scaled partial envelope estimators have a significant advantage over the standard model estimators, partial envelope estimators, and scaled envelope estimators.

The rest of the paper is organized as follows. Section 2 reviews the envelope model, partial envelope model, and scaled envelope model. Furthermore, the scaled partial envelope model is proposed. Section 3 demonstrates the maximum likelihood estimators and identifiability for the scaled partial envelope model parameters. Section 4 describes the theoretical properties of the scaled partial envelope estimator. In Section 5, we discuss selecting the partial envelope dimension u_1. Simulation studies are carried out to compare our proposed model with the standard model, partial envelope model, and scaled envelope model in Section 6. A real-data example is given in Section 7. Section 8 concludes the paper. The proofs of the propositions are provided in the Supplementary Material.

2. Scaled Partial Envelope Model

2.1. Envelope model and partial envelope model

The envelope model (Cook, Li and Chiaromonte (2010)) aims to find the smallest subspace $\mathcal{E} \subseteq \mathbb{R}^r$ that satisfies the following two conditions:

$$(a)Q_{\mathcal{S}}Y|X \sim Q_{\mathcal{S}}Y, (b)Q_{\mathcal{S}}Y \amalg P_{\mathcal{S}}Y|X. \tag{2.1}$$

The sign "$\sim$" means identically distributed, and "$\amalg$" means statistically independent. The symbol $P_{(\cdot)}$ projects onto the subspace and $Q = I_r - P$. Property (a) implies that the distribution of $Q_{\mathcal{S}}Y$ does not rely on X, so $Q_{\mathcal{S}}Y$ has no information about β. Property (b) implies that $Q_{\mathcal{S}}Y$ is conditionally independent of $P_{\mathcal{S}}Y$ given X, and thus $Q_{\mathcal{S}}Y$ cannot transmit information about β via

a link with $P_{\mathcal{S}}Y$. The properties provided by (2.1) are equal to the condition $Q_{\mathcal{S}}Y|(P_{\mathcal{S}}Y, X) \sim Q_{\mathcal{S}}Y$. This expression signifies that the projection $Q_{\mathcal{S}}Y$ is not material to the estimation of β. All of the immaterial information in Y can be gained by detecting the smallest subspace $\mathcal{S}$ that meets the requirements in (2.1).

Let $\mathcal{B} = \text{span}(\beta)$. Cook, Li and Chiaromonte (2010) showed that the pair of conditions (2.1) is equal to the following two conditions:

$$(2a)\mathcal{B} \subseteq \mathcal{S}, (2b)\Sigma = P_{\mathcal{S}}\Sigma P_{\mathcal{S}} + Q_{\mathcal{S}}\Sigma Q_{\mathcal{S}}. \tag{2.2}$$

Condition (2b) holds if and only if $P_{\mathcal{S}}Y$ and $Q_{\mathcal{S}}Y$ are uncorrelated given X, and is equivalent to claiming that $\mathcal{S}$ is a reducing subspace of Σ. These conditions signify that all of the immaterial information can be obtained by choosing $\mathcal{S}$ as the intersection of all reducing subspaces of Σ that include $\mathcal{B}$. This is called the Σ-envelope of $\mathcal{B}$, and is denoted by $\mathcal{E}_\Sigma(\mathcal{B})$, or simply $\mathcal{E}$.

Let $u = \dim\{\mathcal{E}_\Sigma(\mathcal{B})\}$, and let $\Gamma \in \mathbb{R}^{r \times u}$ and $\Gamma_0 \in \mathbb{R}^{r \times (r-u)}$ denote semi-orthogonal basis matrices for $\mathcal{E}_\Sigma(\mathcal{B})$ and its orthogonal complement $\mathcal{E}_\Sigma^\perp(\mathcal{B})$, respectively. By forcing conditions (2.2) on the standard model (1.1), the coordinate form of the envelope model can be acquired as follows:

$$Y = \alpha + \Gamma\eta X + \varepsilon, \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{2.3}$$

where $\eta \in \mathbb{R}^{u \times p}$ are the coordinates of β, which is relative to Γ, and $\Omega \in \mathbb{R}^{u \times u}$ and $\Omega_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are both positive-definite matrices. In model (2.3), $\mathcal{E}_\Sigma(\mathcal{B})$ connects the mean and the covariance matrices, and it is this connection that provides the efficiency gains, particularly when the variation of the immaterial part $\Gamma_0^T Y$ is mainly larger than that of the material part $\Gamma^T Y$. The parameters in (2.3) are obtained by maximizing a normal likelihood function. Let $\widetilde{\Sigma}_Y$, $\widetilde{\beta}$, and $\widetilde{\Sigma}_{\text{res}}$ denote the sample covariance matrix of Y, the least squares estimator of β, and the sample covariance matrix of the residuals from the least squares regression of Y on X, respectively. The estimator of the envelope subspace is the span of $\arg\min\{\log|\Gamma^T\widetilde{\Sigma}_{\text{res}}\Gamma| + \log|\Gamma^T\widetilde{\Sigma}_Y^{-1}\Gamma|\}$, and its minimization is the $r \times u$ Grassmannian (Edelman, Arias and Smith (1998); Adragni, Cook and Wu (2012)).

Su and Cook (2011) extended the envelope model to the partial envelope model. The partial envelopes center on the coefficients corresponding to the predictors of interest. Partition X into two sets of predictors, $X_1 \in \mathbb{R}^{p_1}$ and $X_2 \in \mathbb{R}^{p_2}$, where $p_1 + p_2 = p$ and $p_1 < r$, and partition the columns of β into β_1 and β_2. Then, model (1.1) can be rewritten as $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where β_1 corresponds to the coefficients of interest. The Σ-envelope for $\mathcal{B}_1 = \text{span}(\beta_1)$

is mainly taken into account, leaving β_2 as an unconstrained parameter. This results in the parametric structure $\mathcal{B}_1 \subseteq \mathcal{E}_\Sigma(\mathcal{B}_1)$ and $\Sigma = P_{\mathcal{E},1}\Sigma P_{\mathcal{E},1} + Q_{\mathcal{E},1}\Sigma Q_{\mathcal{E},1}$, where $P_{\mathcal{E},1}$ denotes the projection onto $\mathcal{E}_\Sigma(\mathcal{B}_1)$, called the partial envelope for $\mathcal{B}_1$. This is identical to the envelope structure, except the partial envelope is correlated with $\mathcal{B}_1$ instead of the larger space $\mathcal{B}$. In order to focus on the partial envelope, $\mathcal{E}_\Sigma(\mathcal{B})$ is considered as the full envelope. Because $\mathcal{B}_1 \subseteq \mathcal{B}$, the partial envelope is included in the full envelope, $\mathcal{E}_\Sigma(\mathcal{B}_1) \subseteq \mathcal{E}_\Sigma(\mathcal{B})$, which permits the partial envelope to provide gains that may be impossible with the full envelope.

Let $\Gamma \in \mathbb{R}^{r \times u_1}$ be a semi-orthogonal matrix with $\Gamma^T \Gamma = I_{u_1}$, the columns of which form a basis for $\mathcal{E}_\Sigma(\mathcal{B}_1)$. Let $(\Gamma, \Gamma_0) \in \mathbb{R}^{r \times r}$ be an orthogonal matrix and $\eta \in \mathbb{R}^{u_1 \times p_1}$ be the coordinates of β_1, which is related to the basis matrix Γ. Then, the coordinate version of the partial envelope model can be expressed as follows:

$$Y = \alpha + \Gamma\eta X_1 + \beta_2 X_2 + \varepsilon, \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{2.4}$$

where $\Omega \in \mathbb{R}^{u_1 \times u_1}$ and $\Omega_0 \in \mathbb{R}^{(r-u_1) \times (r-u_1)}$ are both positive-definite matrices that serve as coordinates of $\Sigma_{\mathcal{E}_1}$ and $\Sigma_{\mathcal{E}_1^\perp}$, respectively, which are related to the basis matrices Γ for $\mathcal{E}_\Sigma(\mathcal{B}_1)$ and its orthogonal complement.

To explain it better, let $R_{1|2}$ denote the population residuals from the multivariate linear regression of X_1 on X_2. The predictor X is required to be centered. Then, the linear model can be re-parameterized as $Y = \alpha + \beta_1 R_{1|2} + \beta_2^* X_2 + \varepsilon$, where β_2^* is a linear combination of β_1 and β_2. Next, let $R_{Y|2} = Y - \alpha - \beta_2^* X_2$, the population residuals from the regression of Y on X_2 alone. A linear model that contains β_1 alone is written as $R_{Y|2} = \beta_1 R_{1|2} + \varepsilon$. The partial envelope $\mathcal{E}_\Sigma(\mathcal{B}_1)$ is identical to the full envelope for $\mathcal{B}_1$ in the regression of $R_{Y|2}$ on $R_{1|2}$. That is, the partial envelope can be interpreted in light of $Q_\mathcal{S}Y|X \sim Q_\mathcal{S}Y, Q_\mathcal{S}Y \amalg P_\mathcal{S}Y \mid X$, which is applied to the regression of $R_{Y|2}$ on $R_{1|2}$. The predictors have been centered, so the maximum likelihood estimator of α is simply $\widehat{\alpha} = \overline{Y}$. The estimators of the remaining parameters require the estimator of $\mathcal{E}_\Sigma(\mathcal{B}_1)$.

2.2. Scaled envelope model

The envelope model (2.3) is not invariant or equivariant under linear transformations of the response. Assume that Y is rescaled via multiplication by a nonsingular diagonal matrix A. Let $Y_N = AY$ denote the new response, let $\widehat{\beta}$ and $\widehat{\Sigma}$ denote the estimators of β and Σ, respectively, based on the envelope model for Y on X, and let $\widehat{\beta}_N$ and $\widehat{\Sigma}_N$ denote the estimators of β and Σ, respectively, based on the envelope model for Y_N on X. Accordingly, there is usually no invariance, for example, $\widehat{\beta}_N = \widehat{\beta}$, $\widehat{\Sigma}_N = \widehat{\Sigma}$. Or equivalently, for instance, $\widehat{\beta}_N = A\widehat{\beta}$,

$\widehat{\Sigma}_N = A\widehat{\Sigma}A$. In fact, the dimension of the envelope subspace may alter on account of the transformation. Based on the above motivation and thinking, Cook and Su (2013) extended the envelope model to the scaled envelope model.

For the sake of exhibiting a rescaling, a diagonal matrix $\Lambda = \text{diag}\{1, \lambda_2, \ldots, \lambda_r\} \in \mathbb{R}^{r \times r}$ is introduced, with $\lambda_i > 0$ for $i = 2, \ldots, r$, such that $Y_N = \Lambda^{-1}Y$ follows an envelope model, where the dimension of the envelope subspace $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}$ $(\Lambda^{-1}\mathcal{B})$ is equivalent to u. As a result, $\Lambda^{-1}\mathcal{B} \subseteq \text{span}(\Gamma)$ and $\Lambda^{-1}\Sigma\Lambda^{-1} = P_\Gamma\Lambda^{-1}$ $\Sigma\Lambda^{-1}P_\Gamma + Q_\Gamma\Lambda^{-1}\Sigma\Lambda^{-1}Q_\Gamma$, where $\Gamma \in \mathbb{R}^{r \times u}$ is an orthogonal basis of $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}$ $(\Lambda^{-1}\mathcal{B})$ and $\Gamma_0 \in \mathbb{R}^{r \times (r-u)}$ is a completion of Γ.

The coordinate form of the scaled envelope model is as follows:

$$Y = \alpha + \Lambda\Gamma\eta X + \varepsilon, \Sigma = \Sigma_{\mathcal{E}} + \Sigma_{\mathcal{E}^\perp} = \Lambda\Gamma\Omega\Gamma^T\Lambda + \Lambda\Gamma_0\Omega_0\Gamma_0^T\Lambda. \qquad (2.5)$$

Here, $\beta = \Lambda\Gamma\eta$, with $\eta = \Gamma^T\Lambda^{-1}\beta \in \mathbb{R}^{u \times p}$, is the coefficient matrix, and $\Omega = \text{var}(\Gamma^T\Lambda^{-1}Y) = \Gamma^T\Lambda^{-1}\Sigma\Lambda^{-1}\Gamma \in \mathbb{R}^{u \times u}$ and $\Omega_0 = \text{var}(\Gamma_0^T\Lambda^{-1}Y) = \Gamma_0^T\Lambda^{-1}\Sigma\Lambda^{-1}\Gamma_0$ $\in \mathbb{R}^{(r-u) \times (r-u)}$ are both positive-definite matrices.

2.3. Scaled partial envelope model

Now, we discuss the case $n > \max(r, p)$. A diagonal matrix $\Lambda = \text{diag}\{1, \lambda_2, \ldots, \lambda_r\} \in \mathbb{R}^{r \times r}$, with $\lambda_i > 0 (i = 2, \ldots, r)$, is introduced, so that $(Y)_N = \Lambda^{-1}Y$ follows a partial envelope model, where the dimension of the partial envelope subspace $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}(\Lambda^{-1}\mathcal{B}_1)$ is equivalent to u_1. Hence, $\Lambda^{-1}\mathcal{B}_1 \subseteq \text{span}(\Gamma)$ and $\Lambda^{-1}\Sigma\Lambda^{-1} = P_\Gamma\Lambda^{-1}\Sigma\Lambda^{-1}P_\Gamma + Q_\Gamma\Lambda^{-1}\Sigma\Lambda^{-1}Q_\Gamma$, where $\Gamma \in \mathbb{R}^{r \times u_1}$ is an orthogonal basis of $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}(\Lambda^{-1}\mathcal{B}_1)$ and $\Gamma_0 \in \mathbb{R}^{r \times (r-u_1)}$ is a completion of Γ.

The coordinate form of the scaled partial envelope model is as follows:

$$Y = \alpha + \Lambda\Gamma\eta X_1 + \beta_2 X_2 + \varepsilon, \Sigma = \Sigma_{\mathcal{E}_1} + \Sigma_{\mathcal{E}_1^\perp} = \Lambda\Gamma\Omega\Gamma^T\Lambda + \Lambda\Gamma_0\Omega_0\Gamma_0^T\Lambda. \quad (2.6)$$

Here, $\beta_1 = \Lambda\Gamma\eta$ with $\eta = \Gamma^T\Lambda^{-1}\beta_1 \in \mathbb{R}^{u_1 \times p_1}$ is the coefficient matrix, and $\Omega = \text{var}(\Gamma^T\Lambda^{-1}Y) = \Gamma^T\Lambda^{-1}\Sigma\Lambda^{-1}\Gamma \in \mathbb{R}^{u_1 \times u_1}$ and $\Omega_0 = \text{var}(\Gamma_0^T\Lambda^{-1}Y) = \Gamma_0^T\Lambda^{-1}\Sigma\Lambda^{-1}\Gamma_0$ $\in \mathbb{R}^{(r-u_1) \times (r-u_1)}$ are both positive-definite matrices. A linear model that involves β_1 alone is written as $R_{Y|2} = \Lambda\Gamma\eta R_{1|2} + \varepsilon$. In this model, $R_{1|2}$ denotes the population residuals from the multivariate linear regression of X_1 on X_2, and $R_{Y|2} = Y - \alpha - \beta_2^* X_2$ denotes the population residuals from the regression of Y on X_2 alone, where β_2^* is a linear combination of β_1 and β_2. To aid the computation, we set the first element of Λ to one for the scaling parameters. Otherwise, we can multiply Λ by an arbitrary constant c and multiply η by its reciprocal $1/c$.

For a fixed dimension u_1, the number of parameters in the scaled partial

envelope model (2.6) is $\mathcal{N}_{\mathrm{SPE}} = 2r - 1 + p_1 u_1 + p_2 r + r(r+1)/2$. This is because we need r parameters for α, $r - 1$ parameters for Λ, $p_1 u_1$ parameters for η, $p_2 r$ parameters for β_2, $u_1(u_1 + 1)/2$ parameters for Ω, and $(r - u_1)(r - u_1 + 1)/2$ parameters for Ω_0. The envelope subspace $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}(\Lambda^{-1}\mathcal{B}_1)$ is on an $r \times u_1$ Grassmann manifold, which is the set of all u_1-dimensional subspaces in an r-dimensional space, with $u_1(r - u_1)$ free parameters. Then, we have the following summary for each model:

(i) standard linear model: $\mathcal{N}_{\mathrm{OLS}} = r + pr + r(r+1)/2$ (Cook, Li and Chiaromonte (2010));

(ii) envelope model: $\mathcal{N}_{\mathrm{E}} = r + pu + r(r+1)/2$ (Cook, Li and Chiaromonte (2010));

(iii) partial envelope model: $\mathcal{N}_{\mathrm{PE}} = r + p_1 u_1 + p_2 r + r(r+1)/2$ (Su and Cook (2011));

(iv) scaled envelope model: $\mathcal{N}_{\mathrm{SE}} = 2r - 1 + pu + r(r+1)/2$ (Cook and Su (2013));

(v) scaled partial envelope model: $\mathcal{N}_{\mathrm{SPE}} = 2r - 1 + p_1 u_1 + p_2 r + r(r+1)/2$.

3. Maximum Likelihood Estimators and Identifiability for Scaled Partial Envelope Model Parameters

3.1. Maximum likelihood estimators when Λ is known

Here, we discuss maximum likelihood estimation when Λ is known. When $\Lambda = I_r$, the subscript "o" is employed to indicate quantities in this model to differentiate it from the ordinary partial envelope model (2.4). The response Y in (2.6) is converted to $\Lambda^{-1}Y$, and then we express the generating partial envelope model as

$$\begin{cases} \Lambda^{-1}Y = \alpha_o + \Gamma\eta X_1 + \beta_2 X_2 + \varepsilon_o, \\ \mathrm{var}(\varepsilon_o) = \Sigma_o = \Sigma_{\mathcal{E}_1} + \Sigma_{\mathcal{E}_1^\perp} = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T. \end{cases} \tag{3.1}$$

Using the above equations, we can obtain the scaled partial envelope estimators $\widehat{\beta}_{1,\Lambda}$ of β_1 and $\widehat{\Sigma}_\Lambda$ of Σ. When Λ is known, we convert Y to $\Lambda^{-1}Y$, and then estimate $\beta_{1,o} = \Gamma\eta$ and Σ_o in model (3.1), which follows Su and Cook (2011). Then, $\widehat{\beta}_{1,\Lambda} = \Lambda\widehat{\beta}_{1,o}$ and $\widehat{\Sigma}_\Lambda = \Lambda\widehat{\Sigma}_o\Lambda$. Model (3.1), which has response $\Lambda^{-1}Y$, is simply an ordinary partial envelope model.

3.2. Maximum likelihood estimators when Λ is unknown

Supposing that the errors ε in (2.6) follow the normal distribution, we can obtain maximum likelihood estimators of β_1 and Σ. The definition of scaled

partial envelopes does not require the assumption of normality. In Section 6, we discuss different nonnormal error distributions when normality does not hold.

We assume that the data $((R_{1|2})_i, (R_{Y|2})_i)(i = 1, \ldots, n)$ are independent, and n is the sample size. Let $\overline{R_{Y|2}}$ denote the sample mean of $R_{Y|2}$. By minimizing the objective function, we can acquire the maximum likelihood estimators $\widehat{\Gamma}$ of Γ and $\widehat{\Lambda}$ of Λ. The objective function is as follows:

$$L(\Lambda, \Gamma) = \log|\Gamma^T \Lambda^{-1} \widetilde{\Sigma}_{\mathrm{res}} \Lambda^{-1} \Gamma| + \log|\Gamma^T \Lambda \widetilde{\Sigma}_{R_{Y|2}}^{-1} \Lambda \Gamma|. \qquad (3.2)$$

More details are provided in the Supplementary Material. Here, $\widetilde{\Sigma}_{\mathrm{res}}$ denotes the sample covariance matrix of the residuals from the least squares regression of $R_{Y|2}$ on $R_{1|2}$.

Next, we give the maximum likelihood estimators of the remaining parameters, as follows: $\widehat{\Gamma}_0$ can be any orthogonal basis of the orthogonal complement of $\mathrm{span}(\widehat{\Gamma})$, $\widehat{\eta} = \widehat{\Gamma}^T \widehat{\Lambda}^{-1} \widetilde{\beta}_1$, $\widehat{\Omega} = \widehat{\Gamma}^T \widehat{\Lambda}^{-1} \widetilde{\Sigma}_{\mathrm{res}} \widehat{\Lambda}^{-1} \widehat{\Gamma}$, $\widehat{\Omega}_0 = \widehat{\Gamma}_0^T \widehat{\Lambda}^{-1} \widetilde{\Sigma}_{R_{Y|2}} \widehat{\Lambda}^{-1} \widehat{\Gamma}_0$, $\widehat{\beta}_1 = \widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1} \widetilde{\beta}_1$, and

$$\begin{aligned} \widehat{\Sigma} &= \widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1} \widetilde{\Sigma}_{\mathrm{res}} \widehat{\Lambda}^{-1} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda} + \widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}_0} \widehat{\Lambda}^{-1} \widetilde{\Sigma}_{R_{Y|2}} \widehat{\Lambda}^{-1} \widehat{P}_{\widehat{\Gamma}_0} \widehat{\Lambda}, \\ &= \widehat{\Lambda} \widehat{\Gamma} \widehat{\Omega} \widehat{\Gamma}^T \widehat{\Lambda} + \widehat{\Lambda} \widehat{\Gamma}_0 \widehat{\Omega}_0 \widehat{\Gamma}_0^T \widehat{\Lambda}. \end{aligned}$$

Using the structures of $\widehat{\beta}_1$ and $\widehat{\Sigma}$, we can uncover the course of estimation in the scaled partial envelope model. We consider $\widehat{\beta}_1 = \widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1} U^T F_1 (F_1^T F_1)^{-1}$, where U is an $n \times r$ matrix with ith row $((R_{Y|2})_i - \overline{R_{Y|2}})^T$, and F_1 is an $n \times p_1$ matrix with ith row $(R_{1|2})_i^T (i = 1, \ldots, n)$. First, the response is rescaled, $R_{Y|2} \to \widehat{\Lambda}^{-1} R_{Y|2}$, and centered to obtain $\widehat{\Lambda}^{-1} U^T$. Then, we perform the ordinary partial envelope estimation by employing the rescaled response to obtain $\widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1} U^T F_1 (F_1^T F_1)^{-1}$. Next, the estimator is converted back to the original scales to obtain $\widehat{\beta}_1$. The scaled partial envelope model transforms $R_{Y|2}$ to $\widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1} R_{Y|2}$, and the process $\widehat{\Lambda} \widehat{P}_{\widehat{\Gamma}} \widehat{\Lambda}^{-1}$ is identical, regarding $\widehat{\Lambda}^{-1}$ as an analogous transformation to the original scale of $(R_{Y|2})_N$.

3.3. Parameter identifiability

There is almost always a unique pair $\{\widehat{\Lambda}, \mathrm{span}(\widehat{\Gamma})\}$ to make the objective function (3.2) the global minimizer. When Λ and $\mathrm{span}(\Gamma)$ are not identified, the objective function will usually be flat along some directions, and can return any value in those directions. The parameters β_1 and Σ are our focus of research interest, so this potential nonuniqueness is not a concern. In reality, Proposition 1 guarantees that the maximizers of β_1 and Σ, with reference to the loglikelihood

function, are uniquely defined. Then, we obtain the identical estimators $\widehat{\beta}_1$ and $\widehat{\Sigma}$, regardless of whether or not the global minimizer $\{\widehat{\Lambda}, \text{span}(\widehat{\Gamma})\}$ is unique.

According to Henderson and Searle (1979), the operator vec implies that $\mathbb{R}^{a \times b} \to \mathbb{R}^{ab}$ stacks the columns of a matrix, and the operator vech implies that $\mathbb{R}^{a \times a} \to \mathbb{R}^{a(a+1)/2}$ stacks the lower triangular part of a symmetric matrix. Here, we associate the constituent parameters $\Lambda, \eta, \Gamma, \Omega$, and Ω_0 in the scaled partial envelope models (2.6) with the vector $\phi = \{\lambda^T, \text{vec}(\eta)^T, \text{vec}(\Gamma)^T, \text{vech}(\Omega)^T, \text{vech}(\Omega_0)^T\}^T = (\lambda^T, \phi_o^T)^T$, where $\phi_o = \{\text{vec}(\eta)^T, \text{vec}(\Gamma)^T, \text{vech}(\Omega)^T, \text{vech}(\Omega_0)^T\}^T$ involves the constituent parameters in the model (3.1), and $\lambda = (\lambda_2, \ldots, \lambda_r)^T$ is the vector of the second to the rth diagonal elements of Λ. Let $A \otimes B$ denote the Kronecker product of matrices A and B, and let L denote the $r^2 \times (r-1)$ matrix with columns $e_i \otimes e_i$, where $e_i \in \mathbb{R}^r$ contains a one in the ith position and zero elsewhere, for $i = 2, \ldots, r$. Then, $\lambda = L^T \text{vec}(\Lambda)$. Because $\beta_1 = \Lambda\Gamma\eta = \Lambda\beta_{1,o}$ and $\Sigma = \Lambda(\Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T)\Lambda = \Lambda\Sigma_o\Lambda$, β_1 and Σ are both functions of ϕ.

Proposition 1. *Suppose that the errors in the scaled partial envelope model (2.6) are independent, but not necessarily normal, and have finite second moments, and that $(1/n)\sum_{i=1}^{n}(R_{1|2})_i(R_{1|2})_i^T > 0$. Then, $\beta_1(\phi)$ and $\Sigma(\phi)$ are identifiable and $\widehat{\beta}_1$ and $\widehat{\Sigma}$ are uniquely defined.*

Remark 1. From Proposition 1, when ϕ is not identifiable, β_1 and Σ are identifiable. Moreover, we can obtain unique estimators $\widehat{\beta}_1 = \beta_1(\widehat{\phi})$ and $\widehat{\Sigma} = \Sigma(\widehat{\phi})$. This offers the basis for our discussion of the asymptotic distribution and consistency of $\widehat{\beta}_1$ and $\widehat{\Sigma}$ in Section 4. The proof of Proposition 1 is contained in the Supplementary Material.

4. Theoretical Properties of the Scaled Partial Envelope Estimator

Here, we investigate the asymptotic distribution and consistency of the scaled partial envelope model. Because the parameters β_1 and Σ are our primary focus, the asymptotic distribution of the estimator $\{\text{vec}(\widehat{\beta}_1)^T, \text{vech}(\widehat{\Sigma})^T\}^T$ under normality is provided. To better illustrate the results, we have prepared several definitions below. That is, the contraction matrix $C_r \in \mathbb{R}^{r(r+1)/2 \times r^2}$ and the expansion matrix $E_r \in \mathbb{R}^{r^2 \times r(r+1)/2}$ link the vec and vech operators. For any symmetric matrix $A \in \mathbb{R}^{r \times r}$, $\text{vec}(A) = E_r\text{vech}(A)$ and $\text{vech}(A) = C_r\text{vec}(A)$. Let $\Sigma_{R_{1|2}} = \lim_{n \to \infty}(1/n)\sum_{i=1}^{n}(R_{1|2})_i(R_{1|2})_i^T$, and let $p_{1,ii}$ denote the ith diagonal element of the projection matrix P_{1,F_1}, where F_1 is the $n \times p_1$ matrix with ith row $(R_{1|2})_i^T (i = 1, \ldots, n)$. Let $A^\dagger$ denote the Moore–Penrose inverse of A. The symbol $P_{A(S)}$ denotes the projection in the S inner product onto A or $\text{span}(A)$ if A is a subspace or a matrix, and $Q_{A(S)} = I - P_{A(S)}$. If $\sqrt{n}(\widehat{\theta} - \theta)$ converges to

a normal random vector with mean zero and covariance matrix V, we write its asymptotic covariance matrix as $\mathrm{avar}(\sqrt{n}\widehat{\theta}) = V$.

The asymptotic covariance matrix in model (3.1) is written with subscripts "o". Next, we represent several formulations. The gradient matrix $G_o = \partial\{\mathrm{vec}(\beta_{1,o})^T, \mathrm{vech}(\Sigma_o)^T\}^T/\partial\phi_o^T$ in model (3.1) has dimension $\{p_1 r + r(r+1)/2\} \times \{p_1 u_1 + r(r+1)/2\}$, and is expressed as follows:

$$
\begin{pmatrix}
I_{p_1} \otimes \Gamma & \eta^T \otimes I_r & 0 & 0 \\
0 & 2C_r(\Gamma\Omega \otimes I_r - \Gamma \otimes \Gamma_0\Omega_0\Gamma_0^T) & C_r(\Gamma \otimes \Gamma)E_{u_1} & C_r(\Gamma_0 \otimes \Gamma_0)E_{r-u_1}
\end{pmatrix}.
$$

The Fisher information for $\{\mathrm{vec}(\beta_{1,o})^T, \mathrm{vech}(\Sigma_o)^T\}^T$ in model (3.1) is the $\{p_1 r + r(r+1)/2\} \times \{p_1 r + r(r+1)/2\}$ block diagonal matrix

$$
J_o = \begin{pmatrix}
\Sigma_{R_{1|2}} \otimes \Sigma_o^{-1} & 0 \\
0 & 2^{-1}E_r^T(\Sigma_o^{-1} \otimes \Sigma_o^{-1})E_r
\end{pmatrix}.
$$

Let $h_o = \{L^T(\beta_{1,o} \otimes I_r), 2L^T(\Sigma_o \otimes I_r)C_r^T\}^T$. We can obtain h_o from $h_o = \partial\{\mathrm{vec}(\beta_1)^T, \mathrm{vech}(\Sigma)^T\}^T/\partial\lambda$ in the scaled partial envelope model (2.6) assessed at $\Lambda = I_r$. At the same time, we have $A_o = Q_{G_o(J_o)}h_o L$ and $D_\Lambda = \mathrm{bdiag}\{I_{p_1} \otimes \Lambda, C_r(\Lambda \otimes \Lambda)E_r\}$, which is a block diagonal matrix and has identical dimensions to those of J_o. Here, of all of the quantities defined above, only D_Λ depends on Λ.

The gradient matrix $H = \partial\{\mathrm{vec}(\beta_1)^T, \mathrm{vech}(\Sigma)^T\}^T/\partial\phi^T$ in the scaled partial envelope model (2.6) has dimension $\{p_1 r + r(r+1)/2\} \times \{r - 1 + p_1 u_1 + r(r+1)/2\}$, and we can express the gradient matrix as $H = \{D_\Lambda h_o(I_{p_1} \otimes \Lambda^{-1})L, D_\Lambda G_o\}$. The Fisher information J in the scaled partial envelope model can be acquired by substituting Σ for Σ_o in J_o, and then we obtain J as follows:

$$
J = \begin{pmatrix}
\Sigma_{R_{1|2}} \otimes \Sigma^{-1} & 0 \\
0 & 2^{-1}E_r^T(\Sigma^{-1} \otimes \Sigma^{-1})E_r
\end{pmatrix}.
$$

Proposition 2. *Suppose that $\max_{i \leq n} p_{1,ii} \to 0$ as $n \to \infty$. Then, under model (2.6), which has normal errors, $\sqrt{n}[\{\mathrm{vec}(\widehat{\beta}_1) - \mathrm{vec}(\beta_1)\}^T, \{\mathrm{vech}(\widehat{\Sigma}) - \mathrm{vech}(\Sigma)\}^T]^T$ converges in distribution to a normal random vector with mean zero and covariance matrix*

$$
\begin{aligned}
V &= H(H^T J H)^\dagger H^T, \\
&= D_\Lambda\{A_o(A_o^T J_o A_o)^\dagger A_o^T\}D_\Lambda + D_\Lambda\{G_o(G_o^T J_o G_o)^\dagger G_o^T\}D_\Lambda, \\
&= V_1 + V_2,
\end{aligned}
$$

where $V_1 = D_\Lambda \{ A_o (A_o^T J_o A_o)^\dagger A_o^T \} D_\Lambda$ and $V_2 = D_\Lambda \{ G_o (G_o^T J_o G_o)^\dagger G_o^T \} D_\Lambda$.

The proof of Proposition 2 is provided in the Supplementary Material. Because $J^{-1} - V = J^{-1} - H(H^T J H)^\dagger H^T = J^{-1/2} Q_{J^{1/2}H} J^{-1/2} \geq 0$, we obtain $V \leq J^{-1}$, where J^{-1} is the asymptotic covariance matrix of $\{ \text{vec}(\widetilde{\beta}_1)^T, \text{vech}(\widetilde{\Sigma}_{\text{res}})^T \}^T$.

We employ the normal likelihood as an objective function to get the scaled partial envelope estimators. When the normality assumption fails, a material question is on the consistency of these estimators. The next proposition provides conditions for the $\sqrt{n}$-consistency of $\widehat{\beta}_1$ and $\widehat{\Sigma}$.

Proposition 3. *Suppose that $\max_{i \leq n} p_{1,ii} \to 0$ as $n \to \infty$, and the scaled partial envelope model (2.6) has independent, but not necessarily normal errors, with mean zero and finite fourth moments. Then,*

$$\sqrt{n} \left[\left\{ \text{vec}(\widehat{\beta}_1)^T, \text{vech}(\widehat{\Sigma})^T \right\}^T - \left\{ \text{vec}(\beta_1)^T, \text{vech}(\Sigma)^T \right\}^T \right]$$

is asymptotically normally distributed, and $\widehat{\beta}_1$ and $\widehat{\Sigma}$ are $\sqrt{n}$-consistent estimators of β_1 and Σ, respectively.

Remark 2. Huber (1973) established the consistency of the standard model estimator $\text{vec}(\widetilde{\beta}_1)$, which requires primarily that the maximum leverage tends to zero as $n \to \infty$. The assumption about $p_{1,ii}$ has the identical condition to that of Huber (1973). Furthermore, the estimators are relatively robust to mild deviations from normality in the finite samples. The proof of Proposition 3 is provided in the Supplementary Material.

5. Selection of Partial Envelope Dimension u_1

In the scaled partial envelope model, we can employ standard techniques such as sequential likelihood-ratio tests (LRTs), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) to select u_1. Similarly, we can employ nonparametric methods such as cross-validation or permutation tests (Cook and Yin (2001)) to choose u_1.

The BIC estimator of u_1 is $\arg\min -2\widehat{L}(u_1) + \log(n)\mathcal{N}(u_1)$, where the minimum ranges from zero to the integer r, and n is the sample size. In the above expression, $\mathcal{N}(u_1) = 2r - 1 + p_1 u_1 + p_2 r + r(r+1)/2$ is the number of parameters, and $\widehat{L}(u_1)$ is the maximized loglikelihood in the scaled partial envelope model with dimension u_1,

$$\widehat{L}(u_1) = -\frac{nr}{2}\log(2\pi) - \frac{n}{2}\log|\widetilde{\Sigma}_{R_{Y|2}}| - \frac{n}{2}|\widehat{\Gamma}^T \widehat{\Lambda}^{-1} \widetilde{\Sigma}_{\text{res}} \widehat{\Lambda}^{-1} \widehat{\Gamma}|$$

$$-\frac{n}{2}\log|\widehat{\Gamma}^T\widehat{\Lambda}\widetilde{\Sigma}_{R_{Y|2}}^{-1}\widehat{\Lambda}\widehat{\Gamma}|.$$

Here, $\mathrm{span}(\widehat{\Gamma})$ is the maximum likelihood estimator for $\mathcal{E}_{\Lambda^{-1}\Sigma\Lambda^{-1}}(\Lambda^{-1}\mathcal{B}_1)$ and $\widehat{\Lambda}$ is the maximum likelihood estimator of Λ in the scaled partial envelope model. The AIC estimator of u_1 is $\arg\min -2\widehat{L}(u_1) + 2\mathcal{N}(u_1)$, and the AIC operates analogously.

The properties of the AIC and BIC were researched by Cook and Su (2013, Proposition 4) in the context of response scaling. Similar results are established for the scaled partial envelope model. The candidate set is the set of scaled partial envelope models with dimensions varying from zero to r. Suppose that there are normal errors in the scaled partial envelope model (2.6). Then, if there is one and only one true model in the candidate set as $n \to \infty$, the BIC will choose the true model with probability approaching one, and the AIC will choose a model that at least includes the true model.

6. Simulation Study

In this section, we carry out a simulation study to compare the scaled partial envelope estimator with the standard model estimator, partial envelope estimator, and scaled envelope estimator in terms of the finite sample size. At the same time, we employ the algorithm proposed by Cook, Forzani and Su (2016) for the scaled partial envelope estimation, which does not require optimization over a Grassmannian and is shown to be much faster and typically more accurate than the best existing algorithm proposed by Cook and Zhang (2016). We generated data based on the scaled partial envelope model (2.6) following three cases. The first case was $r = 10$, $p = 12$, $p_1 = 8$, $u = 3$, and $u_1 = 2$, the second case was $r = 10$, $p = 10$, $p_1 = 6$, $u = 5$, and $u_1 = 3$, and the third case was $r = 10$, $p = 6$, $p_1 = 4$, $u = 6$, and $u_1 = 2$. The elements in X_1 were generated as independent $N(0, p_1)$ random variables, and the elements in X_2 were generated as independent $N(0, p - p_1)$ random variables. The matrix β_2 was generated as independent $N(r, p - p_1)$ random variables. We took $\Omega = \sigma^2 I_{u_1}$ and $\Omega_0 = \sigma_0^2 I_{r-u_1}$. The matrix η was generated as a $u_1 \times p_1$ matrix of independent $N(0, 2)$ random variables, and Γ was obtained by orthogonalizing an $r \times u_1$ matrix of independent $U(0, 1)$ random variables. The scale matrix Λ is a diagonal matrix with diagonal elements $2^0, 2^{0.5}, 2^1, 2^{1.5}, \ldots, 2^{4.5}$. We took σ^2 as 0.25 and σ_0^2 as 5. The sample sizes were 100, 300, 600, 1000, and 1500, and we produced 1,000 replicates for each sample size. For each sample size, we calculated the standard deviation of each element in $\widehat{\beta}_1$ over the replicates, which we refer to as the actual standard deviations of

the elements in $\widehat{\beta}_1$. In addition, we calculated the bootstrap standard deviations by bootstrapping the residuals 1,000 times.

Using the above parameter settings, we then fitted four models to the data: (1.1), (2.4), (2.5), and (2.6). Figure 1 plots the standard deviations of a chosen element in $\widehat{\beta}_1$ for three different situations, where the error is normally distributed when $u < r$. As the sample size increases, all standard deviations show a downward trend as a whole, and the efficiency gain increases effectively. In Figure 1, the partial envelope estimators improve on the ordinary least squares estimators, and the scaled envelope estimators improve on the partial envelope estimators. When $u < r$, in all three cases, the scaled partial envelope estimators show an obvious improvement over the ordinary least squares estimators and the partial envelope estimators. The performance between the scaled partial envelope estimators and the scaled envelope estimators is analogous in such a situation, where they have the same parameters σ^2 and σ_0^2, and the part of Y that is material to the predictors is much less variable than the immaterial part, regardless of whether it is for all predictors or part of predictors. In either case, the bootstrap standard deviation is a good approximate estimator of the actual standard deviation.

In the above three parameter settings, except for letting $u = r = 10$, the other parameter settings are unchanged. When $u = r$, the scaled envelope model reduces to the standard multivariate linear regression model, abbreviated as the standard model, so there is no possibility of efficiency gains in this setting. However, we can employ the scaled partial envelope model and still get efficiency gains, as long as $p_1 < r$. In this sense, the scaled partial envelope model is more flexible than the scaled envelope model. Figure 2 plots the standard deviations of a chosen element in $\widehat{\beta}_1$ for three different situations with a normal error when $u = r$. From Figure 2, we find that the standard model and the scaled envelope model have the same standard deviations, and the scaled partial envelope estimators show a remarkable efficiency gain over the partial envelope estimators. Furthermore, the partial envelope estimators have an obvious efficiency gain over the ordinary least squares estimators and the scaled envelope estimators.

Table 1 shows the mean and standard deviation of 1,000 estimated scales with $\sigma_0^2 = 5$ in the $r < p$ situation when $u < r$. The results in the $r \geq p$ situation are analogous to the $r < p$ situation. Our algorithm is shown to be relatively stable.

Figure 3 exhibits the asymptotic action of the scaled partial envelope estimators under nonnormal errors when $u < r$, including a comparison with the standard model estimators, partial envelope estimators, and scaled envelope esti-
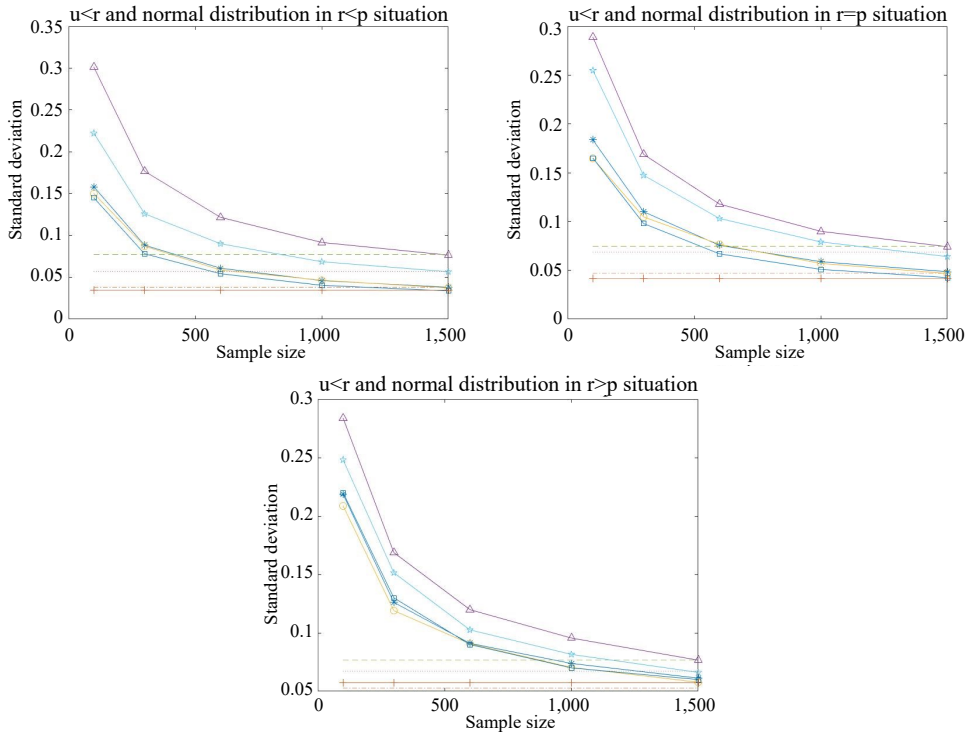
Figure 1. Comparison of the scaled partial envelope estimators, standard model estimators, partial envelope estimators, and scaled envelope estimators: the actual standard deviation of the scaled partial envelope estimators is denoted by "−∗"; the asymptotic standard deviation of the scaled partial envelope estimators is denoted by "−·"; the bootstrap standard deviation of the scaled partial envelope estimators is denoted by "−◯"; the actual standard deviation of the standard model estimators is denoted by "−△"; the asymptotic standard deviation of the standard model estimators is denoted by "−−"; the actual standard deviation of the partial envelope estimators is denoted by "−☆"; the asymptotic standard deviation of the partial envelope estimators is denoted by "..."; the actual standard deviation of the scaled envelope estimators is denoted by "−□", the asymptotic standard deviation of the scaled envelope estimators is denoted by "−+".

mators. We implemented the simulations using the same setup as that in the first case of Figure 1. However, we employed a centered t-distribution with six degrees of freedom, a centered uniform (0,1) distribution, and a chi-squared distribution with four degrees of freedom to represent the distributions with heavier tails, shorter tails, and skewness, respectively. Figure 3 does not show any obvious differences causing the different error distributions, so we conclude that moderate departures from normality do not materially influence the performance of the scaled partial envelope. When the errors obey a nonnormal distribution, the estimator is no longer a maximum likelihood estimator, but efficiency gains are
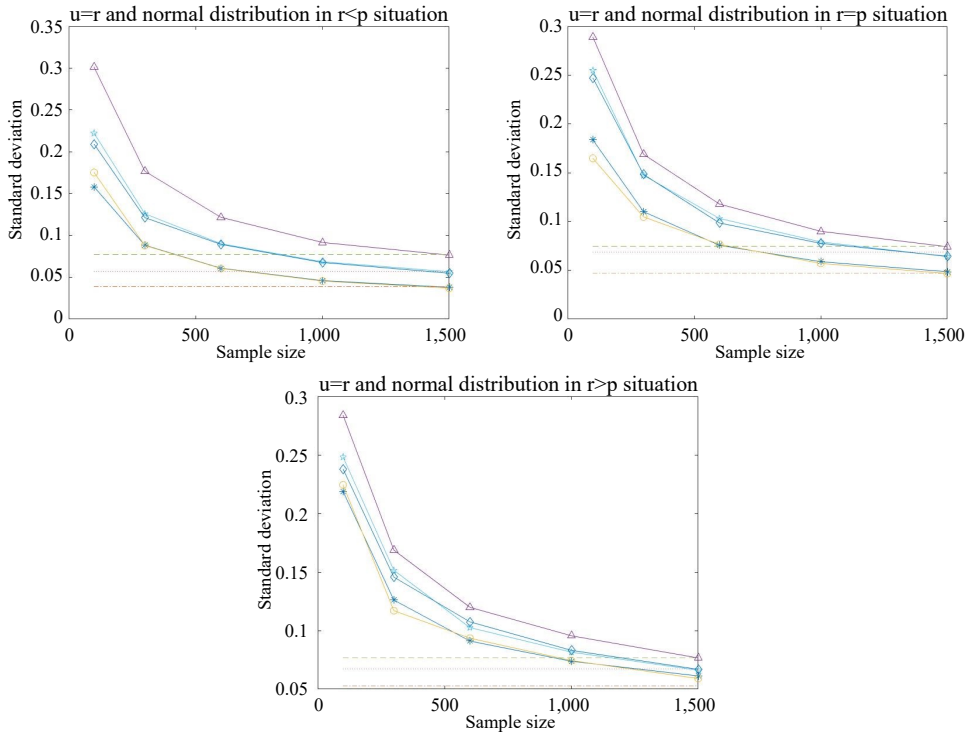
Figure 2. Comparison of the scaled partial envelope estimators, standard model estimators, and partial envelope estimators: the actual standard deviation of the scaled partial envelope estimators is denoted by "$-*$"; the asymptotic standard deviation of the scaled partial envelope estimators is denoted by "$-\cdot$"; the bootstrap standard deviation of the scaled partial envelope estimators is denoted by "$-\bigcirc$"; the actual standard deviation of the standard model estimators is denoted by "$-\triangle$"; the asymptotic standard deviation of the standard model estimators is denoted by "$--$"; the actual standard deviation of the partial envelope estimators is denoted by "$-\star$"; the asymptotic standard deviation of the partial envelope estimators is denoted by "$\ldots$"; the bootstrap standard deviation of the partial envelope estimators is denoted by "$-\diamond$".

still accomplished. The bootstrap standard deviation is still a good approximate estimator of the actual standard deviation. More importantly, the performance of the scaled partial envelope estimators is significantly better than that of the standard model estimators and the partial envelope estimators. The scaled partial envelope estimators and the scaled envelope estimators also have similar behavior when they have the same parameters σ^2 and σ_0^2 and the part of Y that is material to predictors is much less variable than the immaterial part, whether it is for all predictors or part of predictors.

Figure 4 displays the asymptotic action of the scaled partial envelope esti-

Table 1. Mean of base-2 logarithms of the diagonal elements in $\widehat{\Lambda}$. The numbers in parentheses are standard deviations.

n	100	600	1,500
$\log_2(\widehat{\lambda}_2)$	0.5012(0.0640)	0.5022(0.0241)	0.5015(0.0318)
$\log_2(\widehat{\lambda}_3)$	1.0002(0.0856)	1.0030(0.0260)	1.0003(0.0448)
$\log_2(\widehat{\lambda}_4)$	1.5012(0.1185)	1.5033(0.0385)	1.5003(0.0664)
$\log_2(\widehat{\lambda}_5)$	2.0017(0.2551)	2.0024(0.0766)	2.0005(0.0940)
$\log_2(\widehat{\lambda}_6)$	2.5065(0.4660)	2.5035(0.4651)	2.4993(0.1268)
$\log_2(\widehat{\lambda}_7)$	3.0028(0.2971)	3.0028(0.2312)	3.0039(0.1854)
$\log_2(\widehat{\lambda}_8)$	3.5031(0.4375)	3.5045(0.1948)	3.5000(0.2564)
$\log_2(\widehat{\lambda}_9)$	4.0006(0.8645)	4.0026(0.2501)	4.0000(0.3702)
$\log_2(\widehat{\lambda}_{10})$	4.5033(0.9594)	4.5018(0.3681)	4.5000(0.5038)

mators under nonnormal errors when $u = r$, including a comparison with the standard model estimators, partial envelope estimators, and scaled envelope estimators. We implemented the simulations using the same setup as that in the first case of Figure 2, but we employed a centered t-distribution with six degrees of freedom, a centered uniform (0,1) distribution, and a chi-squared distribution with four degrees of freedom to represent the distributions with heavier tails, shorter tails, and skewness, respectively. From Figure 4, we find that when $u = r$ and the errors follow a nonnormal distribution, the standard model and the scaled envelope model have the same standard deviations. In addition, the scaled partial envelope estimators show a remarkable efficiency gain over the partial envelope estimators, and the latter show a distinct efficiency gain over the ordinary least squares estimators and the scaled envelope estimators. This agrees with the performance under the normal error.

7. Real-Data Analysis

This section is devoted to an example that illustrates the advantages of the scaled partial envelope model. The data set is from Johnson and Wichern (2007) and contains information on the properties of pulp fibers and the paper made from them. The data have 62 measurements on four paper properties: breaking length, elastic modulus, stress at failure, and burst strength. The predictors are three properties of fiber: arithmetic fiber length, long fiber fraction, and fine fiber fraction. We consider how the pulp fiber properties X affect the paper properties Y, yielding $r = 4$, $p = 3$, and $p_1 = 1$. Fine fiber fraction is assigned to the main predictor. Arithmetic fiber length and long fiber fraction are assigned to
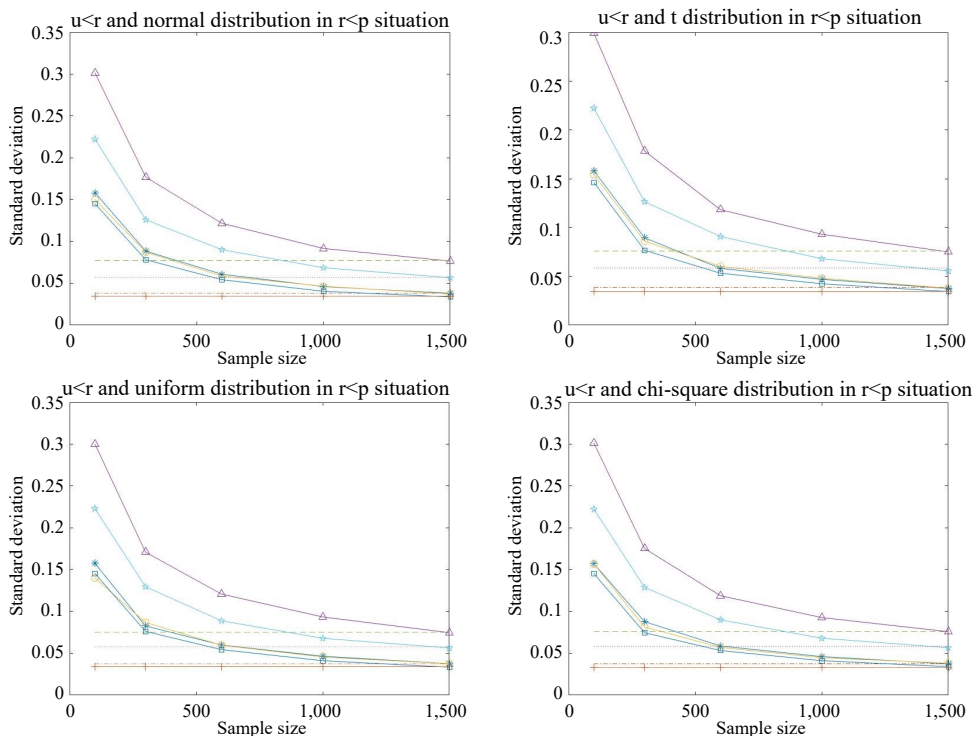
Figure 3. Comparison of the scaled partial envelope estimators with normal, t_6, $U(0,1)$, and χ_4^2 errors when $u < r$. The line marks are the same as those in Figure 1.

the covariates.

We compared the standard errors of the scaled partial envelope estimator $\widehat{\beta}_1$ with those of the ordinary least squares estimator $\widetilde{\beta}_1$ by employing the fractions $f_{1,ij} = 1 - \widehat{\mathrm{avar}}^{1/2}(\sqrt{n}\widehat{\beta}_{1,ij})/\widehat{\mathrm{avar}}^{1/2}(\sqrt{n}\widetilde{\beta}_{1,ij})$, where the subscripts i, j show the elements of the estimator of β_1. The standard errors of the ordinary partial envelope estimator and the ordinary least squares estimator were compared in the same manner. We also compared the standard errors of the scaled envelope estimator $\widehat{\beta}$ with those of the ordinary least squares estimator $\widetilde{\beta}$ by using the fractions $f_{ij} = 1 - \widehat{\mathrm{avar}}^{1/2}(\sqrt{n}\widehat{\beta}_{ij})/\widehat{\mathrm{avar}}^{1/2}(\sqrt{n}\widetilde{\beta}_{ij})$, where the subscripts i, j show the elements of the estimator of β. It is well known (Shao (1997); Yang (2005)) that the BIC behaves better than the AIC if the true model has a simple limited dimensional construction. If we are more concerned about the estimation bias of the associated envelope model itself over prediction, the AIC is more advantageous, because it is more conservative when choosing the dimensions. Here, we mainly consider u and u_1 from the BIC.

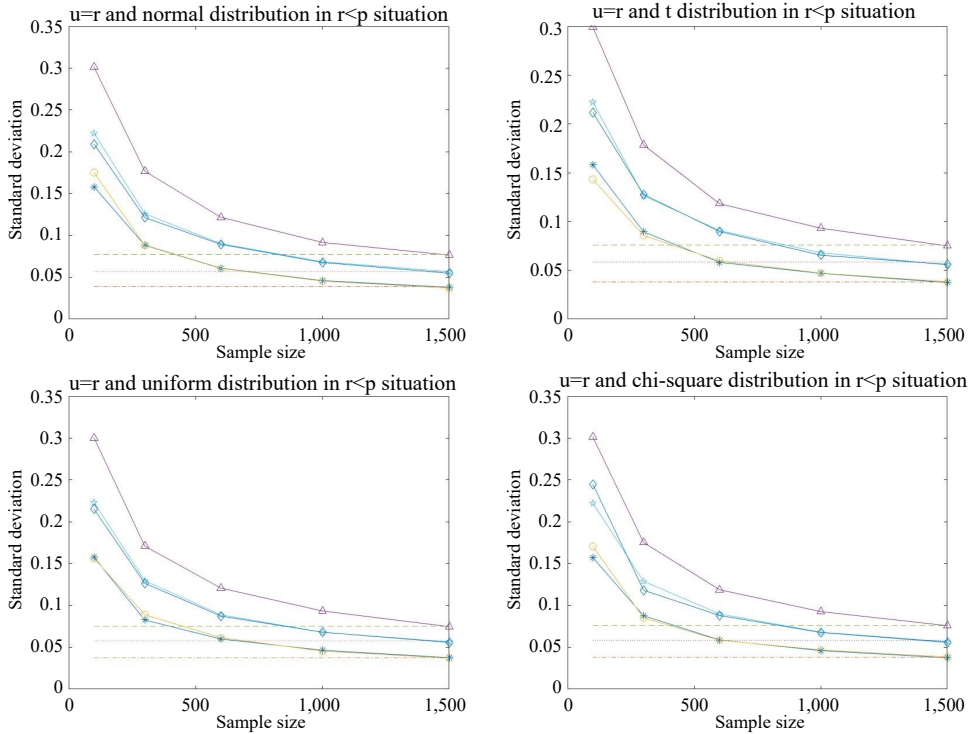We first fitted the scaled envelope model to all the predictors, and the BIC

Figure 4. Comparison of the scaled partial envelope estimators with normal, t_6, $U(0,1)$, and χ_4^2 errors when $u = r$. The line marks are the same as those in Figure 2.

suggested that $u = 2$. Compared with $\widetilde{\beta}$, the standard deviations of the elements in the scaled envelope estimator were 0.15% to 1.97% smaller, $0.0015 \leq f_{ij} \leq 0.0197$. A sample size of about $n = 65$ observations would be needed to reduce the standard error of the ordinary least squares estimator by 1.97, so employing the scaled envelope estimator is roughly the same as using the sample size for inference on some elements of β with the ordinary least squares estimator. This implies that the scaled envelope model does not offer much of an efficiency gain over the standard model. The reason roots in the estimated structure of Σ: the eigenvalues of $\widehat{\Sigma}_{\mathcal{E}}$ are 6.6441 and 0.0176 and the eigenvalues for $\widehat{\Sigma}_{\mathcal{E}^\perp}$ are 0.1310 and 0.0109. Thus, the part of Y that is material to X is no less variable than the immaterial part, and we do not acquire much efficiency from the scaled envelope model.

Next, we fitted an ordinary partial envelope model to the data, and the BIC suggested that $u_1 = 1$. Compared with $\widetilde{\beta}_1$, the standard deviations of the elements in the ordinary partial envelope estimator were 61.62% to 87.46% smaller, $0.6162 \leq f_{1,ij} \leq 0.8746$. Therefore, employing the ordinary partial

envelope estimator is roughly equal to being 64 times the sample size for inference on some elements of β_1 with the ordinary least squares estimator.

When the scaled partial envelope model was fitted to the data, the BIC suggested that $u_1 = 1$. The scale transformation matrix Λ was estimated with diagonal elements $1, 2^{0.5}, 2^1, 2^{1.5}$. Compared with $\widetilde{\beta}_1$, the standard deviations of the elements in the scaled partial envelope estimator were 44.62% to 99.10% smaller, $0.4462 \leq f_{1,ij} \leq 0.9910$, which is a significant improvement over the gains provided by the scaled partial envelope model. In other words, a sample size of about $n = 12 \times 10^3$ observations would be needed to reduce the standard error of the ordinary least squares estimator by 99.10. Because the part of Y that is material to this predictor is much less variable than the immaterial part, there is substantial reduction achieved when we pay close attention to the fine fiber fraction. Focusing on the estimated structure of $\widehat{\Sigma}$, $\widehat{\Sigma}_{\mathcal{E}_1}$ has eigenvalue 0.00007, while $\widehat{\Sigma}_{\mathcal{E}_1^\perp}$ has eigenvalues 3.9521, 0.0145, and 0.00002. In general, when the part of Y that is material to all predictors is no less variable than the immaterial part, but the part of Y that is material to part of the predictors is much less variable than the immaterial part, the scaled partial envelope estimators have a significant advantage over the scaled envelope estimators.

8. Conclusion

We have extended the partial envelope model to the scaled partial envelope model to reduce the dimension efficiently and keep the scale invariable. Then, we gave the maximum likelihood estimation and parameter identifiability. We also showed the theoretical properties and selection of the partial envelope dimension. Simulation studies compared the proposed scaled partial envelope model with the standard model, partial envelope model, and scaled envelope model. A real-data example demonstrated the superiority of the scaled partial envelope model. By introducing a scaling parameter for each response variable, the scaled partial envelope estimator widens the effective range of partial envelope structures, and can bring efficiency gains that are not provided by the ordinary partial envelope estimator.

We consider the case where the predictor and the response variables in the scaled partial envelope model are vector valued. In future work, we will investigate the case where the predictor and the response variables in the model are extended to functional data or are matrix valued. The techniques described here can be applied to other settings, such as tensor regression and discriminant analysis. If prior information emerges, a Bayesian version of this model is possible.

The same idea and techniques can be extended to semiparametric settings, such as quantile regression and expectile regression.

Supplementary Material

The online Supplementary Material contains detailed proofs of Propositions 1–3 in the main manuscript.

Acknowledgments

References

Adragni, K. P., Cook, R. D. and Wu, S. (2012). Grassmannoptim: An R package for Grassmann manifold optimization. *Journal of Statistical Software* **50**, 1–18.

Cook, R. D., Forzani, L. and Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis* **150**, 42–54.

Cook, R. D., Helland, I. S. and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 851–877.

Cook, R. D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica* **20**, 927–1010.

Cook, R. D. and Su, Z. (2013). Scaled envelopes: Scale-invariant and efficient estimation in multivariate linear regression. *Biometrika* **100**, 939–954.

Cook, R. D. and Su, Z. (2016). Scaled predictor envelopes and partial least-squares regression. *Technometrics* **58**, 155–165.

Cook, R. D. and Yin, X. (2001). Dimension reduction and visualization in discriminant analysis (with discussion). *Australian and New Zealand Journal of Statistics* **43**, 147–199.

Cook, R. D. and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association* **110**, 599–611.

Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics* **25**, 284–300.

Cook, R. D. and Zhang, X. (2018). Fast envelope algorithms. *Statistica Sinica* **28**, 1179–1197.

Ding, S. and Cook, R. D. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 387–408.

Edelman, A., Arias, T. A. and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications* **20**, 303–353.

Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses

in Jacobians and multivariate statistics. *The Canadian Journal of Statistics* **7**, 65–81.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics* **1**, 799–821.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. 6th Edition. Prentice Hall, Upper Saddle River.

Khare, K., Pal, S. and Su, Z. (2017). A Bayesian approach for envelope models. *The Annals of Statistics* **45**, 196–222.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association* **112**, 1131–1146.

Pan, Y., Mai, Q. and Zhang, X. (2019). Covariate-adjusted tensor classification in high dimensions. *Journal of the American Statistical Association* **114**, 1305–1319.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica* **7**, 221–242.

Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98**, 133–146.

Su, Z. and Cook, R. D. (2012). Inner envelopes: Efficient estimation in multivariate linear regression. *Biometrika* **99**, 687–702.

Su, Z. and Cook, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica* **23**, 213–230.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* **92**, 937–950.

Zhang, X. and Li, L. (2017). Tensor envelope partial least-squares regression. *Technometrics* **59**, 426–436.

Zhu, G. and Su, Z. (2020). Envelope-based sparse partial least squares. *The Annals of Statistics* **48**, 161–182.

Jing Zhang

School of Science, Nanjing University of Science and Technology, Nanjing, 210094, Jiangsu, P. R. China.

School of Finance, Chuzhou University, Chuzhou, 239000, Anhui, P. R. China.

E-mail: s20082112095@163.com

Zhensheng Huang

School of Science, Nanjing University of Science and Technology, Nanjing, 210094, Jiangsu, P. R. China.

E-mail: stahzs@126.com

Lixing Zhu

Center for Statistics and Data Science, Beijing Normal University at Zhuhai.

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China.

E-mail: lzhu@bnu.edu.cn