Envelopes for multivariate linear regression with linearly constrained coefficients

R. Dennis Cook | Liliana Forzani | Lan Liu

1 School of Statistics, University of Minnesota, Minneapolis, Minnesota, USA
2 Facultad de Ingeniería Química, UNL, CONICET, Santa Fe, Argentina

Correspondence
Lan Liu, School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.
Email: liulan1815@outlook.com

Funding information
Foundation for the National Institutes of Health, Grant/Award Number: R01HL155417-01; National Science Foundation, Grant/Award Number: DMS 1916013; CONICET, Grant/Award Number: 11220200101595CO; ANPCYT, Grant/Award Number: PICT-2018-03005

Abstract
A constrained multivariate linear model is a multivariate linear model with the columns of its coefficient matrix constrained to lie in a known subspace. This class of models includes those typically used to study growth curves and longitudinal data. Envelope methods have been proposed to improve the estimation efficiency in unconstrained multivariate linear models, but have not yet been developed for constrained models. We pursue that development in this article. We first compare the standard envelope estimator with the standard estimator arising from a constrained multivariate model in terms of bias and efficiency. To further improve efficiency, we propose a novel envelope estimator based on a constrained multivariate model. We show the advantage of our proposals by simulations and by studying the probiotic capacity to reduce Salmonella infection.

KEYWORDS
envelope models, growth curves, repeated measures

1 | INTRODUCTION

Consider the multivariate linear regression model

\[ Y_i = \beta_0 + \beta X_i + \epsilon_i, \quad i = 1, \ldots, n, \]
with stochastic response \( Y_i \in \mathbb{R}^r \), nonstochastic predictors \( X_i \in \mathbb{R}^p \), \( \beta_0 \in \mathbb{R}^r \), \( \beta \in \mathbb{R}^{p \times p} \) and error vectors \( \epsilon_i \) independent copies of \( \epsilon \sim N(0, \Sigma) \). The predictors are naturally nonstochastic when they are selected by design. When the predictors are sampled, we condition on them at the outset and so treating them as nonstochastic because they are ancillary under model (1). Model (1) is un constrained in the sense that each response is allowed a separate linear regression: the maximum likelihood estimator (MLE) of the \( j \)th row of \( (\beta_0, \beta) \) is the same as the estimator of the coefficients from the linear regression of the \( j \)th response on \( X \). In many applications, particularly analyses of growth curves and longitudinal data, we may have information that span \( (\beta_0, \beta) \) is contained in a known subspace \( U' \) with basis matrix \( U \in \mathbb{R}^{r \times k} \). The classic dental data (Lee, 1988; Lee & Geisser, 1975; Potthoff & Roy, 1964; Rao, 1987) is an example of such a case.

**Example 1.** A study of dental growth measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure were obtained on 11 girls and 16 boys at ages 8, 10, 12, and 14. The goal was to study the growth measurement as a function of time and sex. We revisited this example using the methodology presented in this article in Supplement Section 5.

Let \( Y_{ik} \) denote the continuous measure of distance for child \( i \) at age \( t_k \), for \( t_k = 8, 10, 12, 14 \), and let \( X_i \) denote the gender indicator for child \( i \) (1 for boy and 0 for girl). After graphical inspection, many researchers treated the population means for distance as linear in time for each gender. Following that, a mixed-effects repeated measure model is \( Y_{ik} = \alpha_{00} + b_{0i} + \alpha_{01} X_i + (\alpha_{10} + b_{1i} + \alpha_{11} X_i) t_k + \epsilon_{ik}^* \), where \( \epsilon_{ik}^* = (\epsilon_{i1}^*, \ldots, \epsilon_{ip}^*)^T \), \( (\alpha_{00}, \alpha_{10}) \), \( \sigma_i = (\alpha_{01}, \alpha_{11}) \), \( \epsilon_i = \epsilon_{i1}^* + b_{0i} 1_{r \times 1} + b_{1i} t \) and \( \epsilon_i \sim N(0, \Sigma) \). Applying the same ideas to just \( \beta \) in (1), so span(\( \beta \)) \( \subseteq U' \) without requiring that span(\( \beta_0 \)) \( \subseteq U' \), leads to the model

\[
Y_i = \beta_0 + U \alpha X_i + \epsilon_i, \quad i = 1, \ldots, n. \tag{3}
\]

Let \( B = \text{span}(\beta) \). If we set \( U = 1_r \), so in model (3) \( \alpha \) is a row vector of length \( p \), then the mean functions for the individual responses are parallel. Although motivated in the context of the dental data, we use models (2) and (3) as general forms that can be adapted to different applications by varying the choice of \( U \), referring to them as constrained multivariate linear models. Cooper and Evans (2002) used a version of model (2) with \( U \) reflecting charge balance constraints on chemical constituents of water samples.

Constrained models occur in various areas including growth curve and longitudinal studies where the elements of \( Y_i \) are repeated observations on the \( i \)th experimental unit over time. It is common in such settings to model the rows of \( U \) as a user-specified vector-valued function \( u(t) \in \mathbb{R}^k \) of time \( t \), the \( i \)th row of \( U \) then being \( u^T(t_i) \). Polynomial bases \( u^T(t) = (1, t, t^2, \ldots, t^{k-1}) \) are prevalent, particularly in the foundational work of Potthoff and Roy (1964), Rao (1965), Grizzle and Allen (1969) and others, but splines (Nummi & Koskela, 2008) or other basis constructions (Izenman & Williams, 1989) could be used as well. In longitudinal studies, model (2) might be used when modeling profiles, while model (3) could be used when modeling just profile differences. For instance, if \( X = 0, 1 \) is a population indicator then under model (2) the
mean profiles are modeled as $\mathbf{U}\alpha_0$ and $\mathbf{U}(\alpha_0 + \alpha)$, while under model (3) the profile means are $\beta_0$ and $\beta_0 + \mathbf{U}\alpha$. It is known in the literature that constrained models gain efficiency in the estimators compared with model (1), provided that $\mathbf{U}$ is correctly specified. However, it may be very difficult to correctly specify $\mathbf{U}$ in some applications, as in the following study:

**Example 2.** Kenward (1987). An experiment was carried out to compare two treatments for the control of gut worm in cattle. Each treatment was randomly assigned to 30 cows whose weights were measured at 2, 4, 6,..., 18, and 19 weeks after treatment. The goal of the experiment was to see whether a differential treatment effect could be detected and, if so, the time point when the difference was first manifested.

The constrained models (2) and (3) require that we select $U$. Lacking prior knowledge, it is natural to inspect plots of the average weight by time, as shown in Figure 1. It seems clear from the figure that it would be difficult to model the treatment profiles, particularly their two crossing points, without running into problems of over fitting. Envelopes provide a way to model data like this without specifying a subspace $U$.

Envelope methodology is based on a relatively new paradigm for dimension reduction that, when applied in the context of model (1), has some similarity with constrained multivariate models. Briefly, envelopes produce a re-parameterization of model (1) in terms of a basis $\Gamma \in \mathbb{R}^{rxu}$ for the smallest reducing subspace of $\Sigma$ that contains $B$. Like the constrained model, envelopes produce an upper bound for $B$, $B \subseteq \text{span}(\Gamma)$, but unlike the constrained model, the bound is unknown and must be estimated. Also, unlike the constrained model, $\Gamma^T\mathbf{Y}$ contains the totality of $\mathbf{Y}$ that is affected by changing $\mathbf{X}$. Since $B \subseteq \text{span}(\Gamma)$, we have $\beta = \Gamma\eta$ for some $\eta \in \mathbb{R}^{uxp}$. Model (1) can be re-parameterized to give its envelope counterpart. For $i = 1,...,n$,

$$Y_i = \beta_0 + \Gamma\eta X_i + \epsilon_i, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T,$$

where $(\Gamma, \Gamma_0) \in \mathbb{R}^{rxr}$, orthogonal, $\Omega = \Gamma^T\Sigma\Gamma > 0$ and $\Omega_0 = \Gamma_0^T\Sigma\Gamma_0 > 0$. Envelopes are reviewed in more detail in Section 2.2.

Comparing (2)–(3) with (4), both express $\beta$ as a basis times a coordinate matrix: $\beta = \mathbf{U}\alpha$ in (2)–(3) and $\beta = \Gamma\eta$ in (4). However, as mentioned previously, $\Gamma$ is estimated but $\mathbf{U}$ is assumed known. Envelopes were first proposed by Cook et al. (2007) to facilitate dimension reduction and later were shown by Cook et al. (2010) to have the potential massive efficiency gains relative to the constrained models.
to the standard MLE of $\beta$, and for these gains to be passed on to other tasks such as prediction. There are now a number of extensions and applications of this basic envelope methodology, each demonstrating the potential for substantial efficiency gains (Cook & Zhang, 2015a,b; Forzani & Su, 2021; Li & Zhang, 2017; Rekabdarkolae et al., 2020; Su et al., 2016; Su & Cook, 2011). Studies over the past several years have demonstrated repeatedly that sometimes the efficiency gains of the envelope methods relative to standard methods amount to increasing the sample size many times over. See Cook (2018) for a review and additional extensions of envelope methodology.

The choice between a constrained model, (2) or (3), and the envelope model (4) hinges on the ability to correctly specify an upper bound $U$ for span($\beta_0, \beta$) or $B$. In Section 2 we obtain the MLE estimators and the asymptotic variances for the parameters of model (2) and show that if we have a correct parsimonious basis $U$ then the constrained models are more efficient. But if bias is present or if we use a correct but excessive $U$, then the envelope model (4) can be much more efficient. To the best of our knowledge, no such a comparison has been investigated theoretically or empirically in the literature despite the similarity between both models. Although considerable methodology has been developed for the envelope version (4) of the unconstrained model (1), there are apparently no envelope counterparts available for the class of models represented by (2) and (3) when a correct parsimonious $U$ is available. In Section 3 we adapt the present envelope paradigm to model versions for (2) and (3) to achieve efficiency gains over those models. Simulations to support our finding are given in Section 4, and in Section 5 we compare our methodology with others in an example. We conclude the article with a discussion section. Proofs for all propositions, two extra data set comparison and discussions of related issues are available in a Supplement to this article.

1.1 Notational conventions

Given a sample $(a_i, b_i), i = 1, \ldots, n$, let $T_{a,b} = n^{-1}\sum_{i=1}^{n} a_i b_i^T$ denote the matrix of raw second moments, and let $T_a = n^{-1}\sum_{i=1}^{n} a_i a_i^T$. For raw second moments involving $Y_S$ and $Y_D$ defined below we use $S$ and $D$ as subscripts. We use a subscript 1 in residuals computed from a model containing a vector of intercepts. The absence of a 1 indicates no intercept was included. For instance, $R_{a1b}$ means the residuals from the regression of $a$ on $b$ without an intercept vector, $a_i = \beta b_i + e_i$, while $R_{a11(b)}$ means those with an intercept vector, $a_i = \beta_0 + \beta b_i + e_i$. Similarly, $R_{D1S}$ means a residual from the regression of $Y_D$ and $Y_S$ without an intercept, and $R_{D11,(S)}$ with an intercept.

Sample variances are written as $S_a = n^{-1}\sum_{i=1}^{n} (a_i - \bar{a})(a_i - \bar{a})^T$ and sample covariance matrices are written as $S_{a,b} = n^{-1}\sum_{i=1}^{n} (a_i - \bar{a})(b_i - \bar{b})^T$. For variances and covariances involving $Y_D$ and $Y_S$, we again use $D$ and $S$ as subscripts, for example, $S_D = n^{-1}\sum_{i=1}^{n} (Y_{Di} - \bar{Y}_D)(Y_{Di} - \bar{Y}_D)^T$. $S_{a1b}$ denotes the covariance matrix of the residuals from fit of the model $a_i = \beta_0 + \beta b_i + e_i$, which always includes an intercept. That is, $S_{a1b} = n^{-1}\sum_{i=1}^{n} R_{a11(b)}R_{a11(b)}^T$. Similarly, $S_{D1S} = \sum_{i=1}^{n} R_{D11,(S)}R_{D11,(S)}^T$.

We use $\text{span}(A)$ to denote the subspace spanned by the columns of the matrix $A$. The projection onto $S = \text{span}(A)$ will be denoted using either the subspace itself $P_S$ or its basis $P_A$. Projections onto an orthogonal complement will be denoted similarly using $Q_i = I - P_{(i)}$. For a subspace $S$ and conformable matrix $B$, $BS = \{BS|S \in S \}$. If an estimator $a \in \mathbb{R}^r$ of $\alpha \in \mathbb{R}^r$ has the property that $\sqrt{n}(a - \alpha)$ is asymptotically normal with mean 0 and variance $A$, we write $\text{var}(\sqrt{n}a) = A$ to denote its asymptotic variance.
2 | COMPARISON OF THE ENVELOPE AND CONSTRAINED ESTIMATORS

Models (2)–(3) and (4) are similar in the sense that $\beta$ is represented as a basis times a coordinate matrix, $\beta = Ua$ in (2)–(3) and $\beta = \Gamma \eta$ in (4). It might be thought that (2) and (3) would yield better estimators because $U$ is known while $\Gamma$ is not, but that turns out not to be true in general. This is in part because we may have $B \not\subseteq U^\perp$, which raises the issue of bias as discussed in Section 2.3, and in part because the envelope model capitalizes automatically on the structure in $\Sigma$, which can improve efficiency as discussed in Section 2.4. Our general conclusion is that, in practice, it may be necessary to compare their fits before selecting an estimator and that the envelope estimator may have a clear advantage when there is uncertainty in the choice of $U^\perp$, as illustrated in Figure 1.

Developments under models (2) and (3) are very similar because they differ only on how the intercept is handled. In the remainder of this article, we focus on model (2) and comment from time to time on modifications necessary for model (3).

2.1 | Maximum likelihood estimators for constrained models

Our treatment of maximum likelihood estimation from (2) is based on linearly transforming $Y$. Let $U_0$ be a semi-orthogonal basis matrix for $U^\perp$, and let $W = (U(U^T U)^{-1}, U_0) := (W_1, W_2)$. Then the transformed model becomes

$$W^T Y_i := \begin{pmatrix} Y_{Di} \\ Y_{Sl} \end{pmatrix} = \begin{pmatrix} (U^T U)^{-1} U^T Y_i \\ U_0^T Y_i \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha X_i \\ 0 \end{pmatrix} + W^T \epsilon_i, \; i = 1, \ldots, n, \tag{5}$$

where $Y_{Di} \in \mathbb{R}^k$ and $Y_{Sl} \in \mathbb{R}^{r-k}$ with $k$ the number of columns of $U$. The transformed variance can be represented block-wise as $\Sigma_W := \text{var}(W^T \epsilon) = (W_1^T \Sigma W_1), \; i, j = 1, 2$, where $\Sigma$ is as defined for model (2). The mean $E(Y_D|X)$ depends nontrivially on $X$, and thus, as indicated by the subscript $D$, we think of $Y_D$ as providing direct information about the regression. On the other hand, $E(Y_S|X) = 0$ and thus $Y_S$ provides no direct information but may provide useful subordinate information by virtue of its association with $Y_D$.

To find the MLEs from model (5), we write the full log likelihood as the sum of the log likelihoods for the marginal model for $Y_S|X$ and the conditional model for $Y_D|(X, Y_S)$:

$$Y_{Sl}|X = e_{Sl}, \tag{6}$$

$$Y_{Di}|(X_i, Y_{Sl}) = \alpha_0 + \alpha X_i + \phi_{D|S} Y_{Sl} + e_{D|S}, \tag{7}$$

where $\phi_{D|S} = (U^T U)^{-1} U^T \Sigma U_0(U_0^T \Sigma U_0)^{-1} \in \mathbb{R}^{k \times (r-k)}$, $e_{D|S} = W_1^T \epsilon$, $e_S = W_2^T \epsilon$. The variances of the errors are $\Sigma_S := \text{var}(e_S) = U_0^T \Sigma U_0$ and $\Sigma_{D|S} := \text{var}(e_{D|S}) = (U^T \Sigma U)^{-1}$. The number of free real parameters in this conditional model is $N_{cm}(k) = k(p + 1) + r(r + 1)/2$. The subscript ‘cm’ is used to indicate estimators arising from the conditional model (7). The MLE and its asymptotic variance for (2) are

$$\hat{\alpha}_{cm} = S_{D,RX|1,Sl} S^{-1}_{X|1} = (S_{D,X} - S_{D,S} S^{-1}_{S,X} S_{X|1}) \tag{8}$$

$$\hat{\beta}_{cm} = U \hat{\alpha}_{cm} = US_{D,RX|1,Sl} S^{-1}_{X|1} = U(S_{D,X} - S_{D,S} S^{-1}_{S,X} S_{X|1}) \tag{9}$$
The estimation for model (3) requires just a few modifications of the procedure for model (2). All modifications stem from the presence of an intercept vector in model (6), which becomes \( Y_S = W^T_2 \beta_0 + e_S \). The variance \( \Sigma_S \) is estimated as \( \hat{\Sigma}_S = S_S \) with corresponding changes in the estimator of \( \Sigma \), and the estimator of the intercept \( W^T_2 \beta_0 \) is just \( \overline{Y}_S \). The intercept in (7) is redefined as \( \alpha_0 = W^T_1 \beta_0 - \phi_{D|S} W^T_2 \beta_0 \). The MLE of \( \beta_0 \) in model (3) can be constructed in a straightforward way from the estimators of \( \alpha_0, W^T_2 \beta_0 \) and \( \phi_{D|S} \). The number of real parameters in (6) becomes \( N_{em} + r - k \). The estimators of the parameters in (7) are unchanged. In particular, \( \hat{\alpha}_{cm} \) and \( \hat{\beta}_{cm} \) along with their asymptotic variances are the same under models (2) and (3), although different \( U \)s might be used in their construction.

### 2.2 Envelope estimator stemming from Model (1)

Consider a subspace \( S \subseteq \mathbb{R}^r \) that satisfies the two conditions (i) \( X \perp Q_S Y \) and (ii) \( P_S Y \perp Q_S Y | X \). Condition (i) insures that the marginal distribution of \( Q_S Y \) does not depend on \( X \), while statement (ii) insures that, given \( X, Q_S Y \) cannot provide material information via an association with \( P_S Y \). Together these conditions imply that the impact of \( X \) on the distribution of \( Y \) is concentrated solely in \( P_S Y \). One motivation underlying envelopes is then to characterize linear combinations \( Q_S Y \) that are unaffected by changes in \( X \) and that produce gains in estimative and predictive efficiency.

In terms of model (1), condition (i) holds if and only if \( B \subseteq S \) and condition (ii) holds if and only if \( S \) is a reducing subspace of \( \Sigma \); that is, \( S \) must decompose \( \Sigma = P_S \Sigma P_S + Q_S \Sigma Q_S \). The intersection of all subspaces with these properties is by construction the smallest reducing subspace of \( \Sigma \) that contains \( B \), which is called the \( \Sigma \)-envelope of \( B \) and is represented as \( \mathcal{E}_\Sigma(B) \) (Cook et al., 2010). These consequences of conditions (i) and (ii) can be incorporated into model (1) by using a basis, leading to model (4). Let \( u \in \{0, 1, \ldots, r\} \) denote the dimension of \( \mathcal{E}_\Sigma(B) \). The number of free real parameters is \( N_{em} = r + pu + r(r + 1)/2 \). The subscript ‘em’ is used to indicate selected quantities arising from this envelope model. The goal here is still to estimate \( \beta = \Gamma \eta \) and \( \Sigma \). Cook et al. (2010) derived the maximum likelihood envelope estimators of \( \beta \) and \( \Sigma \) along with their asymptotic variances. They showed that substantial efficiency gains in estimation of \( \beta \) are possible under this model, particularly when a norm of \( \text{var}(\Gamma_0^T Y) = \Omega_0 \) is considerably larger than the same norm of \( \text{var}(\Gamma Y) = \Omega \).

Given the envelope dimension \( u \), Cook et al. (2010) proved that the maximum likelihood estimator \( \hat{\beta}_{em} \) of \( \beta = \Gamma \eta \) from envelope model (4) has asymptotic variance given by

\[
\text{avar}\left( \sqrt{n} \text{vec} \left( \hat{\beta}_{cm} \right) \right) = \Sigma_X^{-1} \otimes \Gamma \Omega \Gamma^T + (\eta^T \otimes \Gamma_0) M^T(\Sigma_X)(\eta \otimes \Gamma_0^T),
\]

where for a \( C \in \mathbb{R}^{p \times p} \), \( M(C) := \eta C \eta^T \otimes \Omega_0^{-1} + \Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2 I \) and \( \dagger \) denotes the Moore-Penrose inverse. Cook et al. (2010) showed that \( \text{avar}\left( \sqrt{n} \text{vec} \left( \hat{\beta}_{cm} \right) \right) \leq \text{avar}\left( \sqrt{n} \text{vec} \left( \hat{\beta}_{um} \right) \right) \), where \( \hat{\beta}_{um} \) is the MLE under the unconstrained model (1). In consequence, estimators from the envelope model (4) are always superior to those from the unconstrained multivariate model (1). Cook et al. (2010) also showed that the envelope estimator is
\(\sqrt{n}\)-consistent even when the normality assumption is violated as long as the data have finite fourth moments.

### 2.3 Potential bias in \(\hat{\beta}_{\text{cm}}\)

Assuming that \(B \subseteq U^\ast\), \(\hat{\alpha}_{\text{cm}}\) and \(\hat{\beta}_{\text{cm}}\) are unbiased estimators of \(\alpha\) and \(\beta\). However, if \(B \not\subseteq U^\ast\) then both \(\hat{\alpha}_{\text{cm}}\) and \(\hat{\beta}_{\text{cm}}\) are biased, which could materially affect the estimators: 

\[
E(\hat{\alpha}_{\text{cm}}) = (U^T U)^{-1} U^T \beta
\]

and 

\[
E(\hat{\beta}_{\text{cm}}) = P_U \beta.
\]

Consequently, the bias in \(\hat{\beta}_{\text{cm}}\) is 

\[
\beta - P_U \beta = Q_U \beta.
\]

A nonzero bias must necessarily dominate the mean squared error asymptotically and so could limit the utility of \(\hat{\beta}_{\text{cm}}\). Simulation results that illustrate the potential bias effects are discussed in Section 4.2. We assume that \(B \subseteq U\) for the remainder of this article except for where otherwise indicated.

### 2.4 Comparison of asymptotic variances of \(\hat{\beta}_{\text{em}}\) and \(\hat{\beta}_{\text{cm}}\)

We now compare the asymptotic variances of the envelope and constrained estimators of \(\beta\), (12) and (11). Depending on the dimensions involved, the relationship between \(U\) and the envelope \(E_{\Sigma}(B)\) and other factors, the difference between the asymptotic covariance matrices for the estimators \(\hat{\beta}_{\text{em}}\) and \(\hat{\beta}_{\text{cm}}\) from these two models can be positive definite, negative definite, or indefinite. Since all comparisons are in terms of \(\beta\)'s, we assume without loss of generality that \(U\) is a semi-orthogonal matrix. Also, since \(\hat{\beta}_{\text{cm}}\) is the same under models (2) and (3) we do not distinguish between these two models in this section.

#### 2.4.1 \(B \subseteq U \subseteq E_{\Sigma}(B)\)

Assuming that \(U\) is correct so that \(B \subseteq U\) and \(U \subseteq E_{\Sigma}(B)\) can simplify the variance comparison:

**Proposition 1.** If \(B \subseteq U \subseteq E_{\Sigma}(B)\), then \(\text{avar}(\sqrt{n} \text{vec}(\hat{\beta}_{\text{cm}})) \leq \text{avar}(\sqrt{n} \text{vec}(\hat{\beta}_{\text{em}}))\).

In consequence, under this hypothesis, the constrained estimator \(\hat{\beta}_{\text{cm}}\) is superior to the envelope estimator \(\hat{\beta}_{\text{em}}\). However, this comparison may be seen as loaded in favor of \(\hat{\beta}_{\text{cm}}\) since the constrained estimator uses the additional knowledge that \(B \subseteq U\) and the envelope estimator does not. Additionally, neither estimator makes use of the proposition's hypothesis. The next proposition provides help in assessing the impact of the hypothesis on the underlying structure by connecting it with \(E_{\Sigma}(U)\), the \(\Sigma\)-envelope of \(U\).

**Proposition 2.** Assume that \(B \subseteq U\). Then

1. \(E_{\Sigma}(B) \subseteq E_{\Sigma}(U)\),
2. \(U \subseteq E_{\Sigma}(B)\) if and only if \(E_{\Sigma}(B) = E_{\Sigma}(U)\),
3. If \(\text{rank}(\alpha) = k\) then \(B = U\) and \(E_{\Sigma}(B) = E_{\Sigma}(U)\).

This proposition says essentially that if \(B \subseteq U \subseteq E_{\Sigma}(B)\) we can start with model (1) and parameterize in terms of \(E_{\Sigma}(U)\) rather than \(E_{\Sigma}(B)\). A key distinction here is that \(U\) is known while \(B\) is not. In consequence, we expect less estimative variation when parameterizing (1) in terms of \(E_{\Sigma}(U)\) instead of \(E_{\Sigma}(B)\). Since \(U \subseteq E_{\Sigma}(U)\) we can construct a semi-orthogonal
basis for $\mathcal{E}_\Sigma(U')$ as $\Gamma = (U, \Gamma_2)$ with $U_0 = (\Gamma_2, \Gamma_0)$ and, recognizing that $\beta = U\alpha = \Gamma\eta$, we get a new model

$$
Y_i = U\alpha_0 + U\alpha X_i + \epsilon_i, \quad i = 1, \ldots, n, \\
\Sigma = \Gamma \Omega \Gamma^T + \Gamma_0 \Omega_0 \Gamma_0^T.
$$

Consider estimating $\alpha$ from this model using the steps sketched in Section 2.1, and partition $\Omega = (\Omega_{ij})$ to conform to the partition of $\Gamma = (U, \Gamma_2)$. The envelope structure of (13) induces a special structure on the reduced model that corresponds to (6)–(7): $\Sigma_S = bdiag(\Omega_{22}, \Omega_0)$ is block diagonal, $\Sigma_{DJS} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}$ and $\phi_{DJS} = (\Omega_{12} \Omega_{22}^{-1}, 0)$. It can now be shown that the estimators of $\alpha$ from the constrained model (6)–(7) and from (13) have the same asymptotic variance. In other words, if we neglect the hypothesized condition that $\mathcal{U'} \subseteq \mathcal{E}_\Sigma(B)$ then the constrained estimator is better, but if we formulate the envelope model making use of that condition then the constrained and envelope estimators are asymptotically equivalent.

Rao (1967) posited a simple structure for the analysis of balanced growth curve data (see also Geisser, 1970, 1981; Lee, 1988; Lee & Geisser, 1975; Pan & Fang, 2002). In our context, Rao’s structure is obtained by assuming that $\mathcal{E}_\Sigma(B)$ induces a special structure on the reduced model that corresponds to (6)–(7): $\Sigma_S = bdiag(\Omega_{22}, \Omega_0)$ is block diagonal, $\Sigma_{DJS} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}$ and $\phi_{DJS} = (\Omega_{12} \Omega_{22}^{-1}, 0)$. It can now be shown that the estimators of $\alpha$ from the constrained model (6)–(7) and from (13) have the same asymptotic variance. In other words, if we neglect the hypothesized condition that $\mathcal{U'} \subseteq \mathcal{E}_\Sigma(B)$ then the constrained estimator is better, but if we formulate the envelope model making use of that condition then the constrained and envelope estimators are asymptotically equivalent.

Rao (1967) posited a simple structure for the analysis of balanced growth curve data (see also Geisser, 1970, 1981; Lee, 1988; Lee & Geisser, 1975; Pan & Fang, 2002). In our context, Rao’s structure is obtained by assuming that $\mathcal{E}_\Sigma(B)$ induces a special structure on the reduced model that corresponds to (6)–(7): $\Sigma_S = bdiag(\Omega_{22}, \Omega_0)$ is block diagonal, $\Sigma_{DJS} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}$ and $\phi_{DJS} = (\Omega_{12} \Omega_{22}^{-1}, 0)$. It can now be shown that the estimators of $\alpha$ from the constrained model (6)–(7) and from (13) have the same asymptotic variance. In other words, if we neglect the hypothesized condition that $\mathcal{U'} \subseteq \mathcal{E}_\Sigma(B)$ then the constrained estimator is better, but if we formulate the envelope model making use of that condition then the constrained and envelope estimators are asymptotically equivalent.

2.4.2  \quad \mathcal{U'} \supseteq \mathcal{E}_\Sigma(B)

Assuming that $\mathcal{U'} \supseteq \mathcal{E}_\Sigma(B)$ is another way to simplify the variance comparison. Let $\Gamma \in \mathbb{R}^{nxu}$ be a semi-orthogonal basis matrix for $\mathcal{E}_\Sigma(B)$ and let $(\Gamma, \Gamma_0)$ be an orthogonal matrix. Since $\mathcal{U'} \supseteq \mathcal{E}_\Sigma(B)$, we can construct semi-orthogonal bases $U = (\Gamma, \Gamma_0)$ and $\Gamma_0 = (\Gamma_0, \Gamma_0)$. Partition $\Omega_0 = (\Omega_{0,ij})$ to correspond to the partitioning of $\Gamma_0$. Then

**Proposition 3.** Assume that $\mathcal{U'} \supseteq \mathcal{E}_\Sigma(B)$ and let $c \in \mathbb{R}^r$. Then

1. If $c \in \mathcal{E}_\Sigma(B)$ then $\mathrm{avar} \left( \sqrt{nc}^T \hat{\beta}_{cm} \right) = \mathrm{avar} \left( \sqrt{nc}^T \beta_{em} \right)$.
2. If $c \in \mathrm{span}(\Gamma_{02})$ then $\mathrm{avar} \left( \sqrt{nc}^T \hat{\beta}_{cm} \right) \leq \mathrm{avar} \left( \sqrt{nc}^T \beta_{em} \right)$.
3. If $c \in \mathrm{span}(\Gamma_{01})$, $\mathrm{rank} (M(\Sigma_X)) = \mathrm{rank} (\eta \Sigma_X \eta^T \otimes \Omega_0^{-1})$ and $\Omega_{12} = 0$ then $\mathrm{avar} \left( \sqrt{nc}^T \hat{\beta}_{cm} \right) \geq \mathrm{avar} \left( \sqrt{nc}^T \beta_{em} \right)$.

The main takeaway of this lemma is that the difference between the asymptotic covariance matrices for the estimators $\hat{\beta}_{cm}$ and $\beta_{em}$ can be positive semi-definite or negative semi-definite, depending on the characteristics of the problem.

Although the above derivation is under two simple cases where $\mathcal{U'}$ and the envelope space are nested, the conclusion actually holds for the general case: if we have a correct parsimoniously parameterized constrained model then the envelope model (4) is less efficient, but if the basis $U$ in the constrained model is incorrect or excessively parameterized, then envelopes can be much more efficient. This motivated us to incorporate envelopes into the constrained model so that we can further improve efficiency if constraints are reasonably well modeled for the data.
3 | ENVELOPES IN CONSTRAINED MODELS

In this section, we consider two different ways of imposing envelopes in a constrained model when \( B \subseteq U \). As previously done, we focus on envelope estimators in the constrained model (2) and later describe the modifications necessary for model (3). In Section 3.1, we describe the envelope estimation of \( \alpha \) when there is an application-grounded basis \( U \) that is key to interpretation and inference. In Section 3.2, we address envelope estimation of \( \beta = U\alpha \). Here the choice of basis \( U \) has no effect on the MLE of \( \beta \) under the constrained models (2), but it does affect the envelope estimator of \( \beta \). Basis selection is also addressed in Section 3.2.

3.1 | Enveloping \( \alpha \)

Estimation of \( \alpha \) will be of interest when it is desirable to interpret \( \beta = U\alpha \) in terms of its coordinates \( \alpha \) relative to the known application-grounded basis \( U \). Let \( A = \text{span}(\alpha) \). The envelope estimator of \( \alpha \) in model (5) can be found by first transforming (5) into (6)–(7) and then parameterizing (7) in terms of a semi-orthogonal basis matrix \( \phi \in \mathbb{R}^{k\times u} \) for \( \mathcal{E}_{\Sigma_{DS}}(A) \), the \( \Sigma_{DS}\)-envelope of \( A \) with dimension \( u \leq k \). Since \( \text{var} \left( \sqrt{n}\text{vec}(\mathcal{E}_{\Sigma_{DS}}(A)) \right) = \Sigma_X^{-1} \otimes \Sigma_{DS} \) is in the form of a Kronecker product that allows separation of row and column effects of \( \alpha \), this structure follows also from the theory of Cook and Zhang (2015a,b) for matrix-valued envelope estimators based on envelopes of the form \( \mathbb{R}^p \oplus \mathcal{E}_{\Sigma_{DS}}(A) \), where \( \oplus \) denotes the direct sum.

Let \( \eta \in \mathbb{R}^{uxp} \) be an unconstrained matrix giving the coordinates of \( \alpha \) in terms of a semi-orthogonal basis matrix \( \phi \), so \( \alpha = \phi \eta \), and let \( (\phi, \phi_0) \in \mathbb{R}^{k\times k} \) be an orthogonal matrix. Then the envelope version of model (6)–(7) is

\[
\begin{align*}
Y_{Si} &= e_{Si}, \\
Y_{Di}(X_i, Y_{Si}) &= \alpha_0 + \phi \eta X_i + \phi_{DS} Y_{Si} + e_{Di|Si}, \\
\Sigma_{DS} &= \phi \Omega \phi^T + \phi_0 \Omega_0 \phi_0^T,
\end{align*}
\]

where \( \Omega \in \mathbb{R}^{uxu} \) and \( \Omega_0 \in \mathbb{R}^{(k-u)\times(k-u)} \) are positive definite matrices. Part of this model can be seen as a version of the partial envelope model (Su & Cook, 2011)

The total real parameters in model (14) is \( N_{ecm}(u) = k + pu + r(r+1)/2 \), which reduces to that given previously for model (6)–(7) when \( u = k \). The subscript ecm is used to indicate selected key quantities that arise from enveloping \( A \) in the constrained model (2). A basis \( \hat{\phi} \) for the MLE \( \hat{\mathcal{E}}_{\Sigma_{DS}}(A) \) of \( \mathcal{E}_{\Sigma_{DS}}(A) \) is constructed as

\[
\hat{\phi} = \arg \min_{G} \log |G^T S_{DS}(X,S) G| + \log |G^T S_{DS}^{-1}(G)|,
\]

where the minimum is computed over all semi-orthogonal matrices \( G \in \mathbb{R}^{k\times u} \) with \( u \leq k \). The fully maximized log likelihood is

\[
\hat{L}_u = c - \frac{n}{2} \left\{ \log |T_S| + \log |S_{DS}| + \log |\hat{\phi}^T S_{DS}(X,S) \hat{\phi}| + \log |\hat{\phi}^T S_{DS}^{-1}(\hat{\phi})| \right\},
\]

where \( c = n \log |W| - \frac{(nr+2)(1 + \log(2\pi))}{2} \) with the log |\( W \)| term corresponding to the Jacobian transformation back to the scale of \( Y \).
Once \( \hat{\phi} \) is obtained, we get the following envelope estimators for constrained model (2). Specifically, we have

- \( \hat{\beta}_{\text{ecm}} = U\hat{\alpha}_{\text{ecm}}, \quad \hat{\alpha}_{\text{ecm}} = P_{\hat{\Phi}}\hat{\alpha}_{\text{cm}} = \hat{\phi}\hat{\eta}, \) and \( \hat{\alpha}_0 = \overline{Y}_D - \hat{\alpha}_{\text{ecm}}\overline{X} - \hat{\beta}_{\text{D|S}}\overline{S}, \)

- \( \hat{\eta} = \hat{\phi}^T\hat{\alpha}_{\text{cm}}, \quad \hat{\phi}_{\text{D|S}} = S_{\text{D|S}}S_S^{-1} - \hat{\alpha}_{\text{ecm}}S_XS_S^{-1}, \) and \( \hat{\beta}_{\text{ecm}} = U\hat{\alpha}_{\text{ecm}}, \)

- \( \hat{\Omega} = \hat{\phi}^T S_{\text{D|S}}(\hat{\Phi})\) and \( \hat{\Omega}_0 = \hat{\phi}_0^T S_{\text{D|S}}\hat{\phi}_0, \)

- \( \hat{\Sigma}_{\text{D|S}} = \hat{\phi}\hat{\Omega}\hat{\phi}^T + \hat{\phi}_0\hat{\Omega}_0\hat{\phi}_0^T \) and \( \hat{\Sigma}_S = T_S. \)

The variances \( \Sigma_W \) and \( \Sigma \) can be estimated as indicated in Section 2.1. The variances \( \Sigma_W \) and \( \Sigma \) can be estimated as indicated in Section 2.1. The asymptotic variances for \( \hat{\alpha}_{\text{ecm}} \) and \( \hat{\beta}_{\text{ecm}} \) can be deduced from recognizing that in our application \( Y_S \) is random, \( X \) is fixed, and the distribution of \( Y_S|X \) is the same as that of the marginal of \( Y_S \):

\[
\text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\alpha}_{\text{ecm}} \right) \right) = \Sigma_X^{-1} \otimes \phi\Omega\phi^T + (\eta^T \otimes \phi_0)M^T(\Sigma_X)(\eta \otimes \phi_0^T), \quad (17)
\]

\[
\text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\beta}_{\text{ecm}} \right) \right) = \Sigma_X^{-1} \otimes U[\phi\Omega\phi^T + (\eta^T \otimes \phi_0)M^T(\Sigma_X)(\eta \otimes \phi_0^T)]U^T. \quad (18)
\]

We have \( \text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\alpha}_{\text{ecm}} \right) \right) \leq \text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\alpha}_{\text{cm}} \right) \right) \) and \( \text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\beta}_{\text{ecm}} \right) \right) \leq \text{avar} \left( \sqrt{n} \text{vec} \left( \hat{\beta}_{\text{cm}} \right) \right) \) being equal when \( u = k \), so using an envelope in the constrained model always improves estimation asymptotically.

Because \( \mathcal{E}_{\Sigma_{\text{b|g}}}(A) \subseteq \mathbb{R}^k, \mathcal{E}_{\Sigma}(B) \subseteq \mathbb{R}^r \) and \( k \leq r \), it is reasonable to expect that \( \dim(\mathcal{E}_{\Sigma_{\text{b|g}}}(A)) \leq \dim(\mathcal{E}_{\Sigma}(B)) \), as we have estimated in many examples. However, this relationship between the envelope dimension is not guaranteed in general. The following proposition gives sufficient conditions to bound \( \dim(\mathcal{E}_{\Sigma_{\text{b|g}}}(A)) \).

**Proposition 4.** Assume that \( U = (\Gamma'G_0G_0') \), where the \( \Gamma \)'s are as defined for model (4), and that \( G \in \mathbb{R}^{uxu}, \) and \( G_0 \in \mathbb{R}^{(r-u)x(k-u)} \) both have full column rank, so that \( u_1 \leq u \). Then \( \dim(\mathcal{E}_{\Sigma_{\text{b|g}}}(A)) \leq u_1 \leq \dim(\mathcal{E}_{\Sigma}(B)). \)

We can assess the model fitting of (14) using BIC, assuming that the error terms follows a normal distribution. That is, we can compare the constrained envelope model with alternative models by inspecting whether \( -2\hat{\ell}_u + N_{\text{ecm}}(u)\log(n) \) is small. By comparing the BICs of the constrained model with different dimensions \( u \), we can also select the dimension that has the best fit. More about estimating the envelope dimension is given in Supplement Section 9.

### 3.2 Enveloping \( \beta \)

Estimation of \( \beta = U\alpha \) will be of interest in applications where prediction is important or where \( U \) is selected based on convenience, say, rather than on criteria that facilitate understanding and inference. For instance, if \( X \) serves to indicate different treatments then plots of the columns of \( \beta \) versus time give a visual comparisons of the treatment profiles. The choice of \( U^* \) is of course relevant to estimation of \( \beta \), but a basis \( U \) is not uniquely determined. While this flexibility has no effect on the MLE of \( \beta \) under the constrained model (2), it does affect the envelope estimator of \( \beta \). This raises the issue of selecting a good basis for the purpose of estimating \( \beta \) via envelopes.
Consider re-parameterizing $U$ as $UV^{-1}$ and $\alpha$ as $V\alpha$ for some positive definite matrix $V \in \mathbb{R}^{k \times k}$, giving $\beta = U\alpha = (UV^{-1})(V\alpha)$. We could use either $\mathcal{E}_{\mathcal{A}S_{DIS}}(A)$ or $\mathcal{E}_{V\Sigma_{DIS}}(V(A))$ to estimate $\beta$ as $\hat{\beta}_{ecm} = U\hat{\alpha}_{ecm}$ or, in terms of re-parameterized coordinates $V\alpha$, as $\hat{\beta}_{ecm,V} = UV^{-1}(V\alpha)_{ecm}$. In general $\hat{\beta}_{ecm} \neq \hat{\beta}_{ecm,V}$ and we cannot tell which estimator is better. In this section, we show that the envelope estimator of $\beta$ is invariant under orthogonal re-parameterization, so that we only need to consider diagonal re-parameterization: $\beta = U\alpha = (UL^{-1})(L\alpha)$, where $L$ is a diagonal matrix with positive diagonal elements. In growth curve or longitudinal analyses, the columns of $U$ may correspond to different powers of time, and then it seems natural to consider rescaling to bring the columns of $U$ closer to the same scale.

In Supplement Section 3.1, we provided technical tools for demonstrating that the maximum likelihood envelope estimator of $\beta = U\alpha$, when $U$ is semi-orthogonal, is simply $\hat{\beta}_{ecm} = U\hat{\alpha}_{ecm}$.

Thus, to consider the constrained model envelope under a linear transformation of $U$, it suffices to consider a re-scaling transformation. That is, we consider $\beta = U\alpha = (UL^{-1})(L\alpha)$, where $L = \text{diag}(1, \lambda_2, ..., \lambda_k)$. The first diagonal element of $L$ is 1 to ensure identifiability. We follow the general logic of Cook and Su (2013) in their development of a scaled version of model (2).

Without loss of generality, we cast our discussion of scaling in the context of the conditional model (7). We assume that there is a scaling of the response $Y_D$ so that the scaled response $LY_D$ follows an envelope model in $A\alpha$ with the envelope $\mathcal{E}_{A\Sigma_{DIS}}(A, A)$ having dimension $v$ and semi-orthogonal basis matrix $\Theta \in \mathbb{R}^{k \times v}$. Let $(\Theta, \Theta_0)$ denote an orthogonal matrix. Then we can parameterize $L\alpha = \Theta\eta$ and $L\Sigma_{DIS}L = \Theta\Omega\Theta^T + \Theta_0\Omega_0\Theta_0^T$; equivalently, this setup can also be viewed as a rescaling $U \mapsto UL^{-1}$ of $U$, since $LY_D = L(U^T U)^{-1}U^T Y = (L^{-1} U^T U L^{-1} U^T Y$. Since $LY_D$ is unobserved, we now transform back to the original scale for analysis, leading to the marginal model $Y_{SL}|X = e_{SI}$ and conditional model

$$
Y_{DIS}(X_i, Y_{SI}) = \alpha_0 + L^{-1}\Theta\eta X_i + \phi_{DJS}Y_{SI} + e_{DJS},
$$

$$
\Sigma_{DIS} = L^{-1}(\Theta\Omega\Theta^T + \Theta_0\Omega_0\Theta_0^T) L^{-1}.
$$

The total real parameters in this scaled envelope model is $N_{secm}(v) = 2k - 1 + pv + r(r + 1)/2$, where the subscript $secm$ is used to indicate quantities arising from the scaled envelope version of the conditional model. For identifiability, we typically need $N_{secm}(v) \leq N_{cm}$ or $p(k - v) \geq k - 1$.

The goal now is to estimate $\alpha_0$, the coefficient matrix $\beta = UL^{-1}\Theta\eta$ and $\Sigma_{DIS}$, which requires the estimation of several constituent parameters. We presented in Supplement Section 3.2 the maximum likelihood estimators under this model and prove that the asymptotic variance of the estimator $\hat{\beta}_{secm}$ of $\beta$ is $\text{ar}(\sqrt{n}{\text{vec}}(\hat{\beta}_{secm})) = (I_p \otimes U)V_{secm}(I_p \otimes U^T)$ and therefore it is never less efficient than $\hat{\beta}_{cm}$.

### 3.3 Estimation under model (3)

The modifications necessary to adapt the results in Sections 3.1–3.2 for model (3) all stem from the new model for the subordinate response, $Y_S = W^T_S \beta_0 + e_S$, and the new definitions of $\alpha_0 = W^T_S \beta_0 - \phi_{DJS}W^T_S \beta_0$ for models (14) and (19). This implies that $T_S$ is replaced by $S_S$ throughout, including log likelihoods (16) and (13) and that the estimator of $\beta_0$ can be constructed as indicated near the end of Section 2.1. There is no change in the objective functions (15) and (12), and consequently no change in the envelope estimators of $\alpha$ and $\beta$. 
4  |  SIMULATIONS

4.1  |  Efficiency comparison

We first evaluate the efficiency of the envelope estimator \( \hat{\beta}_{em} \), the constrained estimator \( \hat{\beta}_{cm} \) and the constrained envelope estimator \( \hat{\beta}_{ecm} \) using simulations in two scenarios. We also include the unconstrained estimator \( \hat{\beta}_{um} \) as a reference. In the first scenario considered, the eigenvalue corresponding to the material part is small relative to the immaterial part and the dimension of \( U \) is large; therefore, the envelope estimator \( \hat{\beta}_{em} \) is expected to have substantial efficiency gain. In the second scenario, the eigenvalue of the immaterial part is small relative to the one of the material part and the envelope estimator is not expected to have substantial efficiency gain.

4.1.1  |  Scenario 1

The simulation for Scenario 1 is carried out in the following steps:

Step 1. We first generated a sample of size \( n = 200 \). For each individual \( i \), we generated \( p = 8 \) predictors \( X_i \) from a multivariate normal distribution with mean 0 and variance \( CC^T \), where each element in \( C \) is identically and independently distributed with a standard normal distribution \( N(0, 1) \).

Step 2. Set \( r = 20 \), \( u = 6 \), \( q = 15 \), \( q_1 = 4 \) and \( q_2 = q - q_1 \). Set \( \Omega = \text{bdiag}(0.5I_{q-u}, 1.5I_{q_1}) \) and \( \Omega_0 = 50I_{r-u} \). Set \( (\Gamma, \Gamma_0) = \mathbf{O} \) and let \( \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T \), where \( \mathbf{O} \) is an orthogonal matrix obtained by singular value decomposition of a randomly generated matrix. Set \( \eta = K_1K_2 \), where \( K_1 \in \mathbb{R}^{u \times q_1} \), \( K_2 \in \mathbb{R}^{q_1 \times p} \) and each element in \( K_1 \) and \( K_2 \) is identically and independently generated from \( N(0, 1) \). Let \( U = (\Gamma, \Gamma_0)\phi \), where \( \phi = \text{bdiag}(M^U, M_0^U) \), \( M^U = K_1 \) and \( M_0^U = (I_{q_1}, 0_{q_1 \times (r-u-q_2)})^T \), given \( U \in \mathbb{R}^{p \times q} \). Set \( \beta = \Gamma\eta \); notice that it also satisfies \( \beta = U\alpha \) with \( \alpha = (K_2^T, 0_{p \times q_2})^T \).

Step 3. For each individual \( i \), generated \( Y_i \) identically and independently from a normal distribution \( N(\beta X_i, \Sigma) \).

Step 4. Calculate \( \hat{\beta}_{um}, \hat{\beta}_{cm}, \beta_{cm} \), and \( \hat{\beta}_{ecm} \), where \( U \) is correctly specified when calculating \( \hat{\beta}_{cm} \) and \( \hat{\beta}_{ecm} \).

Step 5. Repeat Steps 3–4 1000 times.

We also carried out simulations with a smaller sample size of \( n = 80 \), with results similar to those presented below. They are presented in Supplement Section 4.1.

From the choice of \( \eta \) in Step 2, we have \( \text{colrank}(\eta) = q_1 \), and \( \text{span}(\beta) \) is strictly contained in both \( \text{span}(\Gamma) \) and \( \text{span}(U) \) since the dimension of \( \text{span}(\beta) \) is \( q_1 = 4 \) which is smaller than \( \text{min}(u, q) = 6 \). Specifically, we have \( \text{span}(\beta) = \text{span}(\Gamma^U) = \text{span}(\Gamma) \cap \text{span}(U) \).

As mentioned before, \( \alpha = (K_2^T, 0_{p \times q_2})^T \) in this example. Therefore, we also have a nontrivial constrained envelope of dimension 4: Under the conditions of Scenario 1, we have that \( \Sigma^{-1} = \phi^T \text{bdiag}(\Omega^{-1}, \Omega_0^{-1}) \phi = \text{bdiag}(K_1^T\Omega^{-1}K_1, 50^{-1}I_{q_1}) \). As a consequence, the envelope of \( \alpha \) with respect to \( \Sigma_{DS} \) is \( (K_2^T, 0_{q_1 \times (r-u-q_2)})^T \) with \( K_2 \in \mathbb{R}^{q_1 \times p} \) such that \( K_2 = K_2J \) with \( J \in \mathbb{R}^{q_1 \times p} \) and \( K_2 \) orthogonal. Such a decomposition of \( K_2 \) is possible since \( K_2 \) as rank \( q_1 \). In the 1000 simulations repetitions, the envelope dimension was always correctly estimated as 6 using BIC. The dimension of the constrained envelope estimator was correctly estimated as 4 for 989 times and 5 for 11 times. The empirical results for \( \hat{\beta}_{um} - \beta, \hat{\beta}_{cm} - \beta, \beta_{cm} - \beta \) and \( \hat{\beta}_{ecm} - \beta \) are shown in Figure 2a,
where all the elements of $\beta$ are plotted in the same boxplot as if they are from the same population and the outliers are suppressed for a cleaner representation. Since $U$ is correctly specified, $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$ are asymptotically unbiased estimators, as are $\hat{\beta}_{um}$ and $\hat{\beta}_{em}$. Hence, the boxplots of the four estimators are all centered at 0. In Step 2, the larger eigenvalues of $\Sigma$ are contained in $\Omega_0$ rather than $\Omega$. That is, the variability of the immaterial part is bigger than that of the material part. Additionally, the column space of $U$ is very conservatively specified as $q = 15$, which is much bigger than the dimension of $\text{colrank}(\beta) = 4$, and the span($U$) contains 11 eigenvectors corresponding to large eigenvalues (i.e., 50 in this simulation). Hence, this scenario favors the envelope estimator in terms of the efficiency: the envelope estimator is the most efficient estimator among the three estimators, while $\hat{\beta}_{cm}$ is also more efficient than the saturated estimator $\hat{\beta}_{um}$.

The average estimated asymptotic variances were close to the theoretical asymptotic variances calculated using the true parameter values for all three estimators. The mean of the empirical asymptotic variances across all the elements in four estimators are 51.66 for $\sqrt{n} \hat{\beta}_{um}$ and 39.48 for $\sqrt{n} \hat{\beta}_{cm}$ but is only 1.12 for $\sqrt{n} \hat{\beta}_{em}$ and 0.37 for $\sqrt{n} \hat{\beta}_{ecm}$. That is, in this setting, the envelope estimator is about 40 times more efficient that the saturated estimator and the unconstrained estimator, and the constrained envelope estimator is about 100 times more efficient than those two. A comparison between the envelope and constrained envelope estimator demonstrates the advantage of leveraging prior information in terms of achieving better efficiency.

4.1.2 Scenario 2

To carry out simulations in Scenario 2, we modify some of the parameters in Step 2. In the new Scenario 2, $q = 6$ and we redefine the eigenvalues of $\Sigma$ by setting $\Omega = \text{bdiag}(50I_{u-q}, 0.5I_q)$ and $\Omega_0 = 0.5I_{r-u}$. In this scenario, the larger eigenvalues of $\Sigma$ are associated with $\Omega$. Now the dimension of $U$ is 6 and therefore is only two dimension larger than the dimension of $\beta$.

Since the envelope is also of dimension 6 and needs to be estimated, the envelope method is at a disadvantage in terms of the efficiency as compared with $\hat{\beta}_{cm}$. We have $\alpha = (K_2^T \cdot 0_{p \times q'})^T$ and $\Sigma_{DIS} = \text{bdiag}(K_1^T \Omega^{-1}K_1, (0.5)^{-1}I_q)$ and the dimension of the constraint envelope is still 4. In the 1000 simulations repetitions, the envelope dimension and the constrained envelope dimension are always correctly estimated as 6 and 4, respectively. The empirical biases of the estimators are shown in Figure 2b. Again, all four estimators are centered around 0, indicating the asymptotic unbiasedness. As expected, the estimator $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$ are the most efficient among the four estimators, while the envelope estimator $\hat{\beta}_{em}$ is still more efficient than $\hat{\beta}_{um}$.

The average estimated asymptotic variances of the three estimators were all close to their theoretical values. The average empirical variances of all the elements in three estimators are
8.26 for $\sqrt{n\hat{\beta}_{um}}$, 7.87 for $\sqrt{n\hat{\beta}_{em}}$, 1.50 for $\sqrt{n\hat{\beta}_{cm}}$ and 1.55 for $\sqrt{n\hat{\beta}_{ecm}}$. That is, in this setting, the estimators using a correctly specified $U$ are on average about four times more efficient than the saturated estimator and the envelope estimator, but the envelope constraint estimator does not provide additional advantages over the constrained estimator. We also carried out simulations with a smaller sample size of $n = 80$, with results similar to those presented here. The details are presented in Supplement Section 4.2.

4.2 Potential bias of the constrained estimator

We conducted a small simulation generating data from envelope model (4), to further illustrate potential bias effects. The sample size is again $n = 200$, and the parameters to generate the data are chosen as in Scenario 1, only changing the definition of $U$, which now is $U = (\Gamma, \Gamma_0)A_k$ with $A_k = (I_k, 0)^T$, $k = 1, \ldots, r$. For $k < u$, $B \not\subseteq U^*$ and so both $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$ are biased. But for $k \geq u$, $B \subseteq U^*$ and there is no bias in $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$. Again, the dimension of the constrained envelope remains at 4 when $k \geq u$.

We generated response vectors according to model (4) using normal errors, and fitted the resulting data to obtain the envelope estimator $\hat{\beta}_{em}$. We used the same data to construct the unconstrained estimator $\hat{\beta}_{um}$ and the constrained estimator $\hat{\beta}_{cm}$ with different selections for $U = (\Gamma, \Gamma_0)A_k$ where $A_k = (I_k, 0)^T$, $k = 1, \ldots, r$. For $k < u$, $B \not\subseteq U^*$ and so both $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$ are biased, but for $k \geq u$, $B \subseteq U^*$ and there is no bias in $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$. Actually, the dimension of the constrained envelope remains at 4 when $k \geq u$. We summarized the bias by computing the mean squared error over all elements $\beta_{ij}$ of $\beta$: $\text{MSE} = (rp)^{-1} \sum_{i=1}^{r} \sum_{j=1}^{p} (\hat{\beta}_{(i,j)} - \beta_{ij})^2$ for the three four estimators $\hat{\beta}_{um}, \hat{\beta}_{cm}, \hat{\beta}_{em}$ and $\hat{\beta}_{ecm}$. Shown in Figure 3 are plots of the MSE averaged over 1000 replications of this scheme for Scenario 1, each replication starting with the generation of the response vectors. The constant MSE for $\hat{\beta}_{em}$ was $7.4 \times 10^{-2}$ and that for unconstrained model was about 36 times greater at 0.27. The MSE for both the constrained estimator and the constrained envelope estimator decreased monotonically from its maximum value of about 1.75 at $k = 1$ to around $2.5 \times 10^{-3}$, at $k = u = 6$, the constrained estimator increased monotonically to 0.25 at $k = 20$, while the constrained envelope estimator remains about the same. This is because the constraint envelope is adapted to the data and does not lose much efficiency even if $U$ is large, as shown in Section 4. This suggests that it may be a good practice to
specify a conservative $U$ and apply the constrained envelope to gain more efficiency so that we can enjoy the benefit of prior information but do not suffer from large bias. The corresponding plot for Scenario 2 is similar and therefore is not presented here. It seems clear that the bias in the constrained and constrained envelope estimators can be substantial until we achieve $B \subseteq U'$.

### 4.3 A more general case

The previous simulations were conducted so that both the constrained model and the envelope model can hold under the data generating mechanism. Here, we consider a general case where $U$ is arbitrarily generated but correctly specified. Because $U$ is correctly specified but the envelope model no longer holds, we only compare the constrained model and the constrained envelope model. We carried out the simulations similar to those in Section 4.1, replacing Steps 2–4 with:

Step 2*. Set $r = 20$, $u^* = 3$, $q = 15$. Set $\Omega^* = I$ and $\Omega_0 = 50 I_{q-r}$. Set $(\Gamma^*, \Gamma^*_0) = I$ and let $\var(\varepsilon D|S) = \Sigma_{D|S} = \Gamma^* \Omega^* \Gamma^* + \Gamma^*_0 \Omega_0 \Gamma^*_0$. Generate $\eta^* \in \mathbb{R}^{u \times p}$ and $U$, where each element in $\eta^*$ and $U$ is identically and independently generated from $N(0, 1)$. Set $\alpha^* = \Gamma^* \eta^*$ and $\beta^* = U \alpha^*$.

Step 3*. For each $i$, generate $Y_{Si}$ identically and independently from normal distribution $N(0, I_{r-q})$. Generate $\phi \in \mathbb{R}^{q \times (r-q)}$, where each element is generated identically and independently from standard normal. Generate $Y_{Di}$ from the distribution $N(\alpha^* Z_i + \phi Y_{Si}, \Sigma_{D|S})$

Step 4*. Calculate $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$, where $U$ is correctly specified for both estimators.

The average MSE of $\hat{\beta}_{cm}$ and $\hat{\beta}_{ecm}$ was 0.12 and 0.03. The Monte Carlo mean variances over all the elements were 24.86 and 6.81 for $\sqrt{n} \hat{\beta}_{cm}$ and $\sqrt{n} \hat{\beta}_{ecm}$, demonstrating the efficiency of the additional envelope structure over the $\hat{\beta}_{cm}$ estimator.

### 5 APPLICATION: POSTBIOTICS STUDY

The aim of the postbiotics study (Dunand et al., 2019) was to determine the protective capacity against Salmonella infection in mice of the cell-free fraction (postbiotic) of fermented milk, produced at laboratory and industrial level. The capacity of the postbiotics produced by pH-controlled fermentation to stimulate the production of secretory IgA in feces and to protect mice against Salmonella infection was evaluated. There were three study groups with seven mice per group: (i) a control group (C) where mice received the unfermented milk supernatant, (ii) an F36 group (F36) where mice received the cell-free supernatant obtained by DSM-100H fermentation in 10% (w/v) skim milk produced in the laboratory, and (iii) an F36D group (F36D) where mice received the product F36 diluted 1/10 in tap water. Feces samples (approximately 50 mg per mouse) were collected once a week for 6 weeks and the concentration of secretory IgA (S-IgA) was determined by ELISA. The response was the IgA measures over the 6 weeks period and the predictors the group indicators.

The research question was whether there were differences in the IgA measures among the treatment groups. We present the average response by group over the weeks in Figure 4. We set the control group as the baseline and therefore $\beta \in \mathbb{R}^{6 \times 2}$. We calculate all estimators based on various envelopes on model (3) because we were interested in profile contrasts rather than modeling profile. We use $U^T(t) = (1, t/6, (t/6)^2, \cos(2\pi t/6), \sin(2\pi t/6))$, where $t = 1, \ldots, 6$ are the
weeks where the measures were taking. The unconstrained estimator $\hat{\beta}_{um}$ was considered in Dunand et al. (2019) and it did not show a difference between treatment groups even when exploratory differences can be seen (Figure 4).

Table 1 shows the BIC, envelope dimension, and MSE of the estimators. We listed the maximum envelope dimension for the two nonenvelope methods as their estimated envelope dimensions. The unconstrained estimator performs the worst and the scaled constrained envelope estimator performs the best in terms of both the BIC and the efficiency.

To answer the researcher question, we look the $p$-values of the $\hat{\beta}$ components. From Table 2, we can see that the unconstrained estimator does not reveal any difference, which aligns with the findings in Dunand et al. (2019). None of the estimators demonstrate any evidence of difference between F36D group and the control group at any time. On the other hand, $\hat{\beta}_{secm}$ reveals a

![Figure 4](image-url)  
*Figure 4* Average of IgA by group over time in the Posbiotics Study data.

| Table 1 | Envelope dimension, BIC, BIC order, and MSE for the Postbiotics Study. |
|---------|------------------|-----------|-------|-------|
| Estimator | Dimension | BIC      | BIC order | MSE  |
| $\hat{\beta}_{um}$ | 6          | $-133.90$ | 6      | 0.15  |
| $\hat{\beta}_{em}$ | 1          | $-163.52$ | 2      | 0.13  |
| $\hat{\beta}_{cm}$ | 2          | $-144.37$ | 5      | 0.15  |
| $\hat{\beta}_{secm}$ | 1         | $-160.76$ | 3      | 0.14  |
| $\hat{\beta}_{secm}$ | 1         | $-251.48$ | 1      | 0.13  |

| Table 2 | The $p$-values for coefficients for $\hat{\beta}_{um}$, $\hat{\beta}_{em}$ and $\hat{\beta}_{secm}$. |
|---------|---------------------------------|------------------|-----------|-------|-------|
| week   | F36 versus control | F36 D versus control |
|        | $\hat{\beta}_{um}$ | $\hat{\beta}_{em}$ | $\hat{\beta}_{secm}$ | $\hat{\beta}_{um}$ | $\hat{\beta}_{em}$ | $\hat{\beta}_{secm}$ |
| 1      | 0.91 | 0.07 | 0.13 | 0.77 | 0.27 | 0.30 |
| 2      | 0.09 | 0.10 | 0.01 | 0.83 | 0.28 | 0.21 |
| 3      | 0.83 | 0.01 | 0.01 | 0.48 | 0.20 | 0.22 |
| 4      | 0.26 | 0.06 | 0.02 | 0.90 | 0.23 | 0.22 |
| 5      | 0.55 | 0.05 | 0.00 | 0.16 | 0.20 | 0.20 |
| 6      | 0.57 | 0.63 | 0.01 | 0.59 | 0.64 | 0.21 |
significance difference between the control and F36 groups in all follow-up weeks. The \( p \)-values for such a comparison of \( \hat{\beta}_{em} \) are only significant in week 3. Other estimators also fail to find all follow-up weeks significant between F36 and control groups, for example, the scaled envelope is not significant in weeks 5 and 6, and constrained envelope is significant only in week 2.

The variance gains for the scale version of the constrained envelope model over the unconstrained model (and therefore the \( p \)-values) are reflected by the eigenvalue \( 1 \times 10^{-4} \) of \( \hat{\Omega} \) and the four eigenvalues of \( \hat{\Omega}_0 \) which are 23.06, 13.67, 0.41, and 0.22. The reason for the envelope estimator to be not as significant when comparing F36 and control groups is that there is not as big a discrepancy between the eigenvalues of \( \hat{\Omega} \) \( (2 \times 10^{-3}) \) and those of \( \hat{\Omega}_0 \) \( (0.02, 0.04, 0.03, 0.01, 4 \times 10^{-3}) \).

6 | DISCUSSION

In this article, we first compared the envelope model with the commonly used linear constraint model in terms of both the potential bias and efficiency. We then proposed a constrained envelope model for studying growth curve and longitudinal data when a well-grounded linear constraint is available. We recommend using the constrained envelope model with a relatively conservative \( U \) so that it is likely to contain the space of interest and to achieve efficiency gain. Extensions to unbalanced data and random effects models are designated for future research.

The primary computational step for all of the envelope methods described herein involves finding

\[
\hat{G} = \arg \min_{G \in \mathcal{G}} \log |G^T M_1 G| + \log |G^T M_2 G|
\]

over a class \( \mathcal{G} \) of semi-orthogonal matrices, where the inner product matrices \( M_1 \) and \( M_2 \) depend on the application. The R package Renvlp by M. Lee and Z. Su contains a routine for minimizing objective functions of this form. Computations are straightforward once \( \hat{G} \) has been found. Renvlp also implements specialized methodology for data analysis under envelope model (4) and the partial envelope model. The associated routines can be modified for the models described herein. The codes to reproduce the examples and simulations from this article can be found at https://github.com/lanliu1815/constrained_env.

ORCID

Liliana Forzani 🄩 https://orcid.org/0000-0002-5032-5132

REFERENCES

Cook, R. D. (2018). An introduction to envelopes. Wiley.
Cook, R. D., Li, B., & Chiaromonte, F. (2007). Dimension reduction in regression without matrix inversion. Biometrika, 94(3), 569–584.
Cook, R. D., Li, B., & Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. Statistica Sinica, 20(3), 927–960.
Cook, R. D., & Su, Z. (2013). Scaled envelopes: Scale-invariant and efficient estimation in multivariate linear regression. Biometrika, 100(4), 939–954.
Cook, R. D., & Zhang, X. (2015a). Foundations for envelope models and methods. Journal of the American Statistical Association, 110(510), 599–611.
Cook, R. D., & Zhang, X. (2015b). Simultaneous envelopes for multivariate linear regression. Technometrics, 57(1), 11–25.
Cooper, D. M., & Evans, C. D. (2002). Constrained multivariate trend analysis applied to water quality variables. *Environment, 13*, 42–53. https://doi.org/10.1002/env.504

Dunand, E., Burns, P., Binetti, A., Bergamini, C., Peralta, G. H., Forzani, L., Reinheimer, J., & Vinderola, G. (2019). Postbiotics produced at laboratory and industrial level as potential functional food ingredients with the capacity to protect mice against salmonella infection. *Journal of Applied Microbiology, 127*(1), 219–229.

Forzani, L., & Su, Z. (2021). Envelopes for elliptical multivariate linear regression. *Statistica Sinica, 31*, 301–332.

Geisser, S. (1970). Bayesian analysis of growth curves. *Sankhya, Ser A, 32*(1), 53–64. www.jstor.org/stable/25049636

Geisser, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. *Biometrika, 68*(1), 243–250. https://doi.org/10.1093/biomet/68.1.243

Grizzle, J. E., & Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics, 25*(2), 357–381. https://www.jstor.org/stable/2288859

Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association, 83*(402), 432–440. https://www.jstor.org/stable/2288859

Lee, J. C., & Geisser, S. (1975). Applications of growth curve prediction. *Sankhyā: The Indian Journal of Statistics, Series A, 37*(2), 239–256. https://www.jstor.org/stable/25049978

Li, L., & Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association, 112*(519), 1131–1146. https://doi.org/10.1080/01621459.2016.1193022

Nummi, T., & Koskela, L. (2008). Analysis of growth curve data by using cubic smoothing splines. *Journal of Applied Statistics, 35*(6), 681–691. https://doi.org/10.1080/02664760801923964

Pan, J. X., & Fang, K. T. (2002). Growth curve models and statistical diagnostics. Springer.

Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*(3), 313–326. https://www.jstor.org/stable/2334137

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52*(3), 447–458. https://www.jstor.org/stable/2333697

Rao, C. R. (1967). *Least squares theory using an estimated dispersion matrix and its application to measurement of signals*. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 355–372). University of California Press.

Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science, 2*(4), 434–441. https://projecteuclid.org/euclid.ss/1177013119

Rekabdarkolaeae, H. M., Wang, Q., Naji, Z., & Fluentes, M. (2020). New parsimonious multivariate spatial model: Spatial envelope. *Statistica Sinica, 30*, 1583–1604. www3.stat.sinica.edu.tw/statistica/j30n3/j30n320/j30n320.html

Su, Z., & Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika, 98*(1), 133–146. https://doi.org/10.1093/biomet/asq063

Su, Z., Zhu, G., Chen, X., & Yang, Y. (2016). Sparse envelope model: Estimation and response variable selection in multivariate linear regression. *Biometrika, 103*(3), 579–593.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cook, R. D., Forzani, L., & Liu, L. (2024). Envelopes for multivariate linear regression with linearly constrained coefficients. *Scandinavian Journal of Statistics, 51*(2), 429–446. https://doi.org/10.1111/sjos.12690