Envelopes and reduced-rank regression

R. Dennis Cook [*], Liliana Forzani [†] and Xin Zhang [‡]

October 1, 2013

Abstract

We incorporate the nascent idea of envelopes (Cook, Li and Chiaromonte 2010) into reduced-rank regression, which is a popular techniques for dimension reduction and estimation in the multivariate linear model. We propose a reduced-rank envelope model, which is a hybrid of reduced-rank regression and envelope models. The resulting estimator is at least as efficient as both existing estimators and has total number of parameters no more than either of them. The methodology of this paper can be adapted easily to other envelope models such as partial envelopes (Su and Cook 2011) and envelopes in predictor space (Cook, Helland and Su 2013).

Key Words: Envelope model; Grassmannians; reduced-rank regression.

1 Introduction

The multivariate linear regression model for $p\times 1$ non-stochastic predictor $\mathbf{X}$ and $r\times 1$ stochastic response $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}, \tag{1.1}$$

where the error vector $\boldsymbol{\epsilon}$ has mean zero and covariance matrix $\boldsymbol{\Sigma} > 0$ and is independent of $\mathbf{X}$. This model is a foundation of multivariate statistics where interest lies in prediction

[*]R. Dennis Cook is Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: dennis@stat.umn.edu).

[†]Liliana Forzani is Facultad de Ingeniera Quimica and Instituto de Matematica Aplicada (UNL-CONICET) Guemes 3450 - 3000 Santa Fe, Argentina (E-mail: liliana.forzani@gmail.com).

[‡]Xin Zhang is Ph.D student, School of Statistics, University of Minnesota, Minneapolis, MN, 55455 (Email: zhan0648@umn.edu).

and in studying the interrelation between $\mathbf{X}$ and $\mathbf{Y}$ through the regression coefficient matrix $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$. There is a general awareness that the estimation of $\boldsymbol{\beta}$ may often be improved by reducing the dimensionalities of $\mathbf{X}$ and $\mathbf{Y}$, and reduced-rank regression is popular method for doing so. We propose a reduced-rank envelope model that extends the nascent idea of envelopes to reduced-rank regression. The purpose of this paper is to integrate reduced-rank regression and envelopes, resulting in an overarching method that can choose the better of the two methods when appropriate and that has the potential to perform better than either of them.

Reduced-rank regression (Anderson 1951; Izenman 1975; Reinsel and Velu 1998) arises frequently in multivariate statistical analysis, and has been applied widely across the applied sciences. By restricting the rank of the regression coefficient matrix $\mathrm{rank}(\boldsymbol{\beta}) = d < \min(r, p)$, the total number of parameters is reduced and efficiency in estimation is improved. The analysis of reduced-rank regression (Izenman 1975; Tso 1981; Reinsel and Velu 1998; Anderson 2002) connects with many important multivariate methods such as principal components analysis, canonical correlation analysis and multiple time series modeling. The asymptotic advantages of the reduced-rank regression estimator over the standard ordinary least squares estimator were studied by Stoica and Viberg (1996) and Anderson (1999). Chen et al. (2012) and Chen and Huang (2012) extended reduced-rank regression to high-dimensional settings and demonstrated the advantages of parsimoniously reducing model parameters and interrelating response variables.

Envelope regression, which was first proposed by Cook et al. (2010), is another way of parsimoniously reducing the total number of parameters from the standard model (1.1) and gaining both efficiency in estimation and accuracy in prediction. The key idea of envelopes is to identify and eliminate information in the responses and the predictors that is immaterial to the estimation of $\boldsymbol{\beta}$ but still introduces unnecessary variation into estimation. Envelope reduction can be effective even when $d = \min(p, r)$, which is the case where reduced-rank regression has no reduction.

Envelope and reduced-rank regressions have different perspectives on dimension reduction.

It may take considerable effort to find which method is more efficient for a problem in practice. The proposed reduced-rank envelope model combines the strengths of envelopes and reduced-rank regression, which mitigates the burden of selecting among the two methods. When one of the two methods behaves poorly, the reduced-rank envelope model automatically degenerates towards the other one; when both methods show efficiency gains, the reduced-rank envelope estimator will enjoy a synergy from combining the two approaches and may improve over both estimators.

The rest of this paper is organized as follows. In Section 2, we review and summarize some fundamental results for reduced-rank regression and envelopes that are relevant to our development. We set up our reduced-rank envelope model in Section 2.3, where we also give intuitive connections to reduced-rank regression and envelope models. In Section 3.1, we summarize parameterizations for each model and show that the total number of parameters in the reduced-rank envelope model is fewer than that of the other models. Likelihood-based estimators for the reduced-rank envelope model are derived in Section 3.2. Asymptotic properties are studied in Section 4. We show that the reduced-rank envelope estimator is asymptotically more efficient than ordinary least squares, reduced-rank regression and envelope estimators under normal errors, and is still $\sqrt{n}$-consistent without the normality assumption. Section 5 discusses procedures for selecting the rank of the coefficient matrix and the dimension of the envelope. Encouraging simulation results and real data examples are presented in Section 6 and 7. Proofs and other technical details are included in a Supplement to this article.

The following notations and definitions will be used in our exposition. Let $\mathbb{R}^{m\times n}$ be the set of all real $m \times n$ matrices. The Grassmannian consisting of the set of all u dimensional subspaces of $\mathbb{R}^r$, $u \leq r$, is denoted as $\mathcal{G}_{r,u}$. If $\mathbf{M} \in \mathbb{R}^{m\times n}$, then $\mathrm{span}(\mathbf{M}) \subseteq \mathbb{R}^m$ is the subspace spanned by columns of $\mathbf{M}$. If $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges to a normal random vector with mean 0 and covariance matrix $\mathbf{V}$ we write its asymptotic covariance matrix as $\mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\theta}}) = \mathbf{V}$. We use $\mathbf{P}_{\mathbf{A}(\mathbf{V})} = \mathbf{A}(\mathbf{A}^T\mathbf{V}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{V}$ to denote the projection onto $\mathrm{span}(\mathbf{A})$ with the $\mathbf{V}$ inner product and use $\mathbf{P}_{\mathbf{A}}$ to denote projection onto $\mathrm{span}(\mathbf{A})$ with the identity inner product. Let

$\mathbf{Q}_{\mathbf{A}(\mathbf{V})} = \mathbf{I} - \mathbf{P}_{\mathbf{A}(\mathbf{V})}$. We will use operators $\mathrm{vec}: \mathbb{R}^{a\times b} \to \mathbb{R}^{ab}$, which vectorizes an arbitrary matrix by stacking its columns, and $\mathrm{vech}: \mathbb{R}^{a\times a} \to \mathbb{R}^{a(a+1)/2}$, which vectorizes a symmetric matrix by extracting its columns of elements below or on the diagonal. Let $\mathbf{A} \otimes \mathbf{B}$ denote the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$. We use $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\xi}}$ to denote estimator of $\boldsymbol{\theta}$ with known true parameter value of $\boldsymbol{\xi}$. For a common parameter $\boldsymbol{\theta}$ in different models, we will use subscripts to distinguish the estimators according to different models: $\widehat{\boldsymbol{\theta}}_{\mathrm{RR}}$ for the reduced-rank regression estimator, $\widehat{\boldsymbol{\theta}}_{\mathrm{Env}}$ for the envelope estimator, $\widehat{\boldsymbol{\theta}}_{\mathrm{RE}}$ for the reduced-rank envelope estimator and $\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}$ for the ordinary least square estimator.

2 Reduced-rank envelope model

2.1 Reduced-rank regression

Reduced-rank regression allows that $\mathrm{rank}(\boldsymbol{\beta}) = d < \min(p, r)$ so that we can write the model parameterization as

$$\boldsymbol{\beta} = \mathbf{AB},\ \mathbf{A} \in \mathbb{R}^{r\times d},\ \mathbf{B} \in \mathbb{R}^{d\times p},\ \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}) = d, \tag{2.1}$$

where no additional constraints are imposed on $\mathbf{A}$ or $\mathbf{B}$. The maximum likelihood estimators for the reduced-rank regression parameters were derived by Anderson (1999), Reinsel and Velu (1998) and Stoica and Viberg (1996), under various constraints on $\mathbf{A}$ and $\mathbf{B}$ for identifiability, such as $\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T = \mathbf{I}_d$ or $\mathbf{A}^T\mathbf{A} = \mathbf{I}_d$. The decomposition $\boldsymbol{\beta} = \mathbf{AB}$ is still non-unique even with those identifiable constraints: for any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{d\times d}$, $\mathbf{A}_1 = \mathbf{AO}$ and $\mathbf{B}_1 = \mathbf{O}^T\mathbf{B}$ offer another valid decomposition that satisfies the constraints. The parameters of interests, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, are nevertheless identifiable as well as $\mathrm{span}(\mathbf{A}) = \mathrm{span}(\boldsymbol{\beta})$ and $\mathrm{span}(\mathbf{B}^T) = \mathrm{span}(\boldsymbol{\beta}^T)$. We present this article in an apparently novel unified framework so that every statement involving $\mathbf{A}$ or $\mathbf{B}$ holds universally for any decomposition $\boldsymbol{\beta} = \mathbf{AB}$ satisfying (2.1).

The log-likelihood of model (1.1) under normality of $\boldsymbol{\epsilon}$ can be written as,

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \simeq -\frac{n}{2}\left\{\log|\boldsymbol{\Sigma}| + \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mathbf{X}_i)^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mathbf{X}_i)\right\}, \qquad (2.2)$$

which is to be maximized under the constraint that $\mathrm{rank}(\boldsymbol{\beta}) = d$, or equivalently under the parameterization $\boldsymbol{\beta} = \mathbf{AB}$. The symbol $\simeq$ denotes an equality from which any unimportant additive constant has been eliminated. We treat $L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ as a general purpose objective function, which will be maximized under (2.1). The following lemma summarizes the reduced-rank regression estimator that maximizes (2.2). Rigorous derivation can be found in Anderson (1999).

Sample covariance matrices in this article are represented as $\mathbf{S}_{(\cdot)}$ and defined with the divisor n. For instance, $\mathbf{S}_{\mathbf{X}} = \sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T/n$, $\mathbf{S}_{\mathbf{XY}} = \sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{Y}_i - \overline{\mathbf{Y}})^T/n$, $\mathbf{S}_{\mathbf{Y}|\mathbf{X}} = \mathbf{S}_{\mathbf{Y}} - \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{XY}}$ denotes the sample covariance matrix of the residuals from the linear fit of $\mathbf{Y}$ on $\mathbf{X}$, and $\mathbf{S}_{\mathbf{Y}\circ\mathbf{X}} = \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{XY}}$ denotes the sample covariance matrix of the fitted vectors from the linear fit of $\mathbf{Y}$ on $\mathbf{X}$. We define the sample canonical correlation matrix between $\mathbf{Y}$ and $\mathbf{X}$ as $\mathbf{C}_{\mathbf{YX}} = \mathbf{S}_{\mathbf{Y}}^{-1/2}\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1/2}$ and $\mathbf{C}_{\mathbf{XY}} = \mathbf{C}_{\mathbf{YX}}^T$. Truncated matrices are represented with superscripts. For example, $\mathbf{C}_{\mathbf{YX}}^{(d)}$ and $\mathbf{S}_{\mathbf{YX}}^{(d)}$ are constructed by truncated singular value decompositions of $\mathbf{C}_{\mathbf{YX}}$ and $\mathbf{S}_{\mathbf{YX}}$ with only the largest d singular values being kept.

Lemma 1. *Under the reduced-rank regression parameterization (2.1), the likelihood-based objective function from (2.2) is maximized at* $\widehat{\boldsymbol{\alpha}}_{\mathrm{RR}} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RR}}\overline{\mathbf{X}}$ *and*

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\mathrm{RR}} &= \mathbf{S}_{\mathbf{Y}}^{1/2}\mathbf{C}_{\mathbf{YX}}^{(d)}\mathbf{S}_{\mathbf{X}}^{-1/2} \\ \widehat{\boldsymbol{\Sigma}}_{\mathrm{RR}} &= \mathbf{S}_{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RR}}\mathbf{S}_{\mathbf{XY}} = \mathbf{S}_{\mathbf{Y}}^{1/2}\left\{\mathbf{I}_r - \mathbf{C}_{\mathbf{YX}}^{(d)}\mathbf{C}_{\mathbf{XY}}^{(d)}\right\}\mathbf{S}_{\mathbf{Y}}^{1/2}.\end{aligned}$$

There are a variety forms of maximizers $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ in the literature under different constraints on $\mathbf{A}$ and $\mathbf{B}$. They could all be reproduced by decomposing the rank-d estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ in Lemma 1. The ordinary least squares estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ can be written as $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \mathbf{S}_{\mathbf{Y}}^{1/2}\mathbf{C}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1/2}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{OLS}} = \mathbf{S}_{\mathbf{Y}|\mathbf{X}} = \mathbf{S}_{\mathbf{Y}}^{1/2}\left\{\mathbf{I}_r - \mathbf{C}_{\mathbf{YX}}\mathbf{C}_{\mathbf{XY}}\right\}\mathbf{S}_{\mathbf{Y}}^{1/2}$ by replacing the

truncated sample canonical correlation matrices $\mathbf{C}_{(\cdot)}^{(d)}$ with the untruncated ones $\mathbf{C}_{(\cdot)}$. This Lemma also reveals the scale equivariant property of both reduced-rank regression and ordinary least squares estimators since the truncated sample canonical correlation matrices are scale invariant.

2.2 Review of Envelopes

The envelope model (Cook et al. 2010) seeks the smallest subspace $\mathcal{E} \subseteq \mathbb{R}^r$ such that

$$\mathbf{Q}_{\mathcal{E}}\mathbf{Y}|\mathbf{X} \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y} \text{ and } \operatorname{cov}(\mathbf{Q}_{\mathcal{E}}\mathbf{Y}, \mathbf{P}_{\mathcal{E}}\mathbf{Y}|\mathbf{X}) = 0. \tag{2.3}$$

For any $\mathcal{E}$ with those properties, $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ carries only information that is irrelevant to the linear regression. The projected response $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ is linearly immaterial to the estimation of $\boldsymbol{\beta}$ in the sense that it responds to neither the predictor nor the rest of response $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$, which presents material information in the response. When the conditional distribution of $\mathbf{Y}|\mathbf{X}$ is normal, the second statement in (2.3) implies that $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ is independent of $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ given $\mathbf{X}$. The smallest subspace satisfying (2.3) always uniquely exists and is denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ as defined formally in the following definitions.

Definition 1. *A subspace $\mathcal{R} \subseteq \mathbb{R}^r$ is said to be a reducing subspace of $\mathbf{M} \in \mathbb{R}^{r\times r}$, or equivalently saying $\mathcal{R}$ reduces $\mathbf{M}$, if and only if $\mathcal{R}$ decomposes $\mathbf{M}$ as $\mathbf{M} = \mathbf{P}_{\mathcal{R}}\mathbf{M}\mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}}\mathbf{M}\mathbf{Q}_{\mathcal{R}}$.*

The definition of a reducing subspace is basic in functional analysis (Conway 1990) but the notion of reduction is different from the common statistical meaning. Reducing subspaces are central to the study of envelope models and methods.

Definition 2. *Let $\mathbf{M} \in \mathbb{R}^{r\times r}$ and let $\mathcal{S} \subseteq \operatorname{span}(\mathbf{M})$. Then the $\mathbf{M}$-envelope of $\mathcal{S}$, denoted by $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$, is the intersection of all reducing subspaces of $\mathbf{M}$ that contain $\mathcal{S}$.*

Definition 2 guarantees the existence and the uniqueness of envelopes by noticing that the intersection of any two reducing subspaces of $\mathbf{M}$ is still a reducing subspace of $\mathbf{M}$. To avoid proliferation of notation, we may use a matrix in the argument of an envelope as $\mathcal{E}_{\mathbf{M}}(\mathbf{B}) :=$

$\mathcal{E}_{\mathbf{M}}(\mathrm{span}(\mathbf{B}))$. Under the reduced-rank regression model (2.1), $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\mathbf{A})$ and the dimension of the envelope denoted by u is always no less than d since $\dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})) \geq \dim(\mathrm{span}(\boldsymbol{\beta})) = \mathrm{rank}(\boldsymbol{\beta}) = d$. The following proposition from Cook, et al. (2010) gives a characterization of envelopes.

Proposition 1. *If $\mathbf{M} \in \mathbb{R}^{r\times r}$ has $s < r$ eigenspaces, then the $\mathbf{M}$-envelope of $\mathcal{S} \subseteq \mathrm{span}(\mathbf{M})$ can be constructed as $\mathcal{E}_{\mathbf{M}}(\mathcal{S}) = \sum_{i=1}^{s} \mathbf{P}_i\mathcal{S}$, where $\mathbf{P}_i$ is the projection onto the i-th eigenspace of $\mathbf{M}$.*

From this proposition, we see that the $\mathbf{M}$-envelope of $\mathcal{S}$ is the sum of the eigenspaces of $\mathbf{M}$ that are not orthogonal to $\mathcal{S}$. This implies that the envelope is the span of some subset of the eigenvectors of $\mathbf{M}$.

2.3 Reduced-rank envelope model

Let $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ be an orthogonal basis for $\mathbb{R}^r$ so that $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times u}$. Then $\dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})) = u$ and

$$\boldsymbol{\beta} = \mathbf{AB} = \boldsymbol{\Gamma}\boldsymbol{\xi} = \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{B},\ \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{2.4}$$

where $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are symmetric positive definite matrices in $\mathbb{R}^{u\times u}$ and $\mathbb{R}^{(r-u)\times(r-u)}$ respectively and $\boldsymbol{\eta} \in \mathbb{R}^{u\times d}$, $u \geq d$, are the coordinates of $\mathbf{A}$ with respect to $\boldsymbol{\Gamma}$. The parameterization $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\xi}$ with $\boldsymbol{\xi} \in \mathbb{R}^{u\times p}$ occurs in the envelope model of Cook et al. (2010). We still impose no additional constraint on $\mathbf{A}$, $\mathbf{B}$ or $\boldsymbol{\eta}$ other than requiring them all to have rank d. The decompositions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in (2.4) are not unique but $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are unique.

To see the connections between the reduced-rank envelope model and reduced-rank regression, we next consider the situation in which $\boldsymbol{\Gamma}$ is known. Notice that $\mathrm{span}(\boldsymbol{\Gamma})$ is uniquely defined while $\boldsymbol{\Gamma}$ is unique up to an orthogonal transformation in $\mathbb{R}^u$. Although expressions in Lemma 2 are given in terms of $\boldsymbol{\Gamma}$, the final estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}$ depend on $\boldsymbol{\Gamma}$ only via $\mathrm{span}(\boldsymbol{\Gamma})$: for any orthogonal transformation $\mathbf{O} \in \mathbb{R}^{u\times u}$, we have $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}\mathbf{O}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}\mathbf{O}}$.

Lemma 2. *Under the reduced-rank envelope model (2.4), the likelihood-based objective function from (2.2) with given* $\boldsymbol{\Gamma}$ *is maximized at* $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{\Gamma}} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}}\overline{\mathbf{X}}$ *and*

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}} &= \boldsymbol{\Gamma}\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma}}\widehat{\mathbf{B}}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\mathbf{S}^{1/2}_{\boldsymbol{\Gamma}^T\mathbf{Y}}\mathbf{C}^{(d)}_{\boldsymbol{\Gamma}^T\mathbf{Y},\mathbf{X}}\mathbf{S}^{-1/2}_{\mathbf{X}} \\ \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}} &= \boldsymbol{\Gamma}\mathbf{S}^{1/2}_{\boldsymbol{\Gamma}^T\mathbf{Y}}\left\{\mathbf{I}_u - \mathbf{C}^{(d)}_{\boldsymbol{\Gamma}^T\mathbf{Y},\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X},\boldsymbol{\Gamma}^T\mathbf{Y}}\right\}\mathbf{S}^{1/2}_{\boldsymbol{\Gamma}^T\mathbf{Y}}\boldsymbol{\Gamma}^T + \mathbf{Q}_{\boldsymbol{\Gamma}}\mathbf{S}_{\mathbf{Y}}\mathbf{Q}_{\boldsymbol{\Gamma}}.\end{aligned}$$

The implication of Lemma 2 is clear: once we know the envelope, we can focus our attention on the reduced response $\boldsymbol{\Gamma}^T\mathbf{Y}$ and find $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma}}\widehat{\mathbf{B}}_{\boldsymbol{\Gamma}}$, which is the rank-$d$ reduced-rank regression estimator of $\boldsymbol{\Gamma}^T\mathbf{Y}$ on $\mathbf{X}$. By Definition 1, the covariance estimator $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}$ is now reduced by $\mathrm{span}(\boldsymbol{\Gamma})$ since $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}} = \mathbf{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}\mathbf{P}_{\boldsymbol{\Gamma}} + \mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}\mathbf{Q}_{\boldsymbol{\Gamma}}$. Hence $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Gamma}}$ that also contains $\mathrm{span}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}})$, and the envelope structure is preserved by the construction of these estimators. In Section 3.2, we derive the likelihood-based estimator $\widehat{\boldsymbol{\Gamma}}$ and demonstrate that the reduced-rank envelope estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ coincide with the estimators in Lemma 2 by replacing $\boldsymbol{\Gamma}$ with $\widehat{\boldsymbol{\Gamma}}$.

When the envelope dimension $u = r$, there is no immaterial information to be reduced by the envelope method. Then the reduced-rank envelope model degenerates to the reduced-rank regression (2.1), $\boldsymbol{\Gamma} = \mathbf{I}_r$. When the regression coefficient matrix is full rank $\mathrm{rank}(\boldsymbol{\beta}) = p \leq r$, reduced-rank regression is equivalent to ordinary least squares and the reduced-rank envelope model degenerates to the ordinary envelope model. Two extreme situations are then: (a) if $p > r = 1$ then both methods degenerate to the standard method, which produces no reduction; (b) if $r > p = 1$ then reduced-rank regression can not provide any response reduction while reduced-rank envelopes can still gain efficiency by projecting the response onto the envelope $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. The reduced-rank envelope model can be extended to the predictor envelopes by Cook et al. (2013), so that it can resolve the problem in (a) and provide potential gain by enveloping in the predictor space.

3 Likelihood-based estimation for reduced-rank envelope

3.1 Parameters in different models

Following Cook, Li and Chiaromonte (2010), we define the following estimable functions $\mathbf{h}$ for the standard model (1.1), parameters $\boldsymbol{\psi}$ for the reduced-rank model, parameters $\boldsymbol{\delta}$ for the envelope model and parameter $\boldsymbol{\phi}$ for the reduced-rank envelope model. The common parameter $\boldsymbol{\alpha}$ is omitted because its estimator takes the following form for all methods: $\widehat{\boldsymbol{\alpha}} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}\overline{\mathbf{X}}$, while $\overline{\mathbf{Y}}$ and $\overline{\mathbf{X}}$ are asymptotically independent of the other estimators.

$$\mathbf{h} = \begin{pmatrix} \mathrm{vec}(\boldsymbol{\beta}) \\ \mathrm{vech}(\boldsymbol{\Sigma}) \end{pmatrix}, \ \boldsymbol{\psi} = \begin{pmatrix} \mathrm{vec}(\mathbf{A}) \\ \mathrm{vec}(\mathbf{B}) \\ \mathrm{vech}(\boldsymbol{\Sigma}) \end{pmatrix}, \ \boldsymbol{\delta} = \begin{pmatrix} \mathrm{vec}(\boldsymbol{\Gamma}) \\ \mathrm{vec}(\boldsymbol{\xi}) \\ \mathrm{vech}(\boldsymbol{\Omega}) \\ \mathrm{vech}(\boldsymbol{\Omega}_0) \end{pmatrix}, \ \boldsymbol{\phi} = \begin{pmatrix} \mathrm{vec}(\boldsymbol{\Gamma}) \\ \mathrm{vec}(\boldsymbol{\eta}) \\ \mathrm{vec}(\mathbf{B}) \\ \mathrm{vech}(\boldsymbol{\Omega}) \\ \mathrm{vech}(\boldsymbol{\Omega}_0) \end{pmatrix}, \tag{3.1}$$

where we define $\mathbf{h} = (\mathbf{h}_1^T, \mathbf{h}_2^T)^T$, $\boldsymbol{\psi} = (\boldsymbol{\psi}_1^T, \boldsymbol{\psi}_2^T, \boldsymbol{\psi}_3^T)^T$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_1^T, \ldots, \boldsymbol{\delta}_4^T)^T$ and $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_5)^T$ correspondingly. We have $\mathbf{h} = \mathbf{h}(\boldsymbol{\psi})$ under the reduced-rank model, $\mathbf{h} = \mathbf{h}(\boldsymbol{\delta})$ under the envelope model and $\mathbf{h} = \mathbf{h}(\boldsymbol{\phi})$ under the reduced-rank envelope model.

We use $\mathcal{N}(\cdot)$ to denote the total number of unique real parameters in a vector of model parameters. We have the following summary for each method:

(i) standard linear model, $\mathcal{N}_{\mathrm{OLS}} := \mathcal{N}(\mathbf{h}) = pr + r(r+1)/2$;

(ii) reduced-rank model, $\mathcal{N}_{\mathrm{RR}} := \mathcal{N}(\boldsymbol{\psi}) = (p + r - d)d + r(r+1)/2$;

(iii) envelope model, $\mathcal{N}_{\mathrm{Env}} := \mathcal{N}(\boldsymbol{\delta}) = pu + r(r+1)/2$;

(iv) reduced-rank envelope model, $\mathcal{N}_{\mathrm{RE}} := \mathcal{N}(\boldsymbol{\phi}) = (p + u - d)d + r(r+1)/2$.

By straightforward calculation we observe that the total number of unique parameters is reduced by $(p-d)(r-d) \geq 0$ from standard model to reduced-rank regression, and is further reduced by $(r-u)d \geq 0$ from reduced-rank regressionto reduced-rank envelopes. Similarly, the total number of unique parameters is reduced by $p(r-u) \geq 0$ from the standard model to envelopes,

and is further reduced by $(p-d)(u-d) \geq 0$ from the envelope model to the reduced-rank envelope model.

3.2 Estimators for the reduced-rank envelope model parameters

The goal of this section is to derive the reduced-rank envelope estimators for given d and u. Procedures for selecting d and u are discussed in Section 5. The likelihood-based reduced-rank envelope estimators is obtained by substituting $\mathbf{h} = \mathbf{h}(\boldsymbol{\phi})$ into (2.2) and maximizing $L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}(\boldsymbol{\phi}), \boldsymbol{\Sigma}(\boldsymbol{\phi})) \equiv L_n(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{B}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Gamma}|d, u)$ over all parameters except $\boldsymbol{\Gamma}$ because they live on a product space and the optimizing value of $\boldsymbol{\Gamma}$ cannot be found analytically. We then arrive at the estimator $\widehat{\boldsymbol{\Gamma}}$ from optimization over a Grassmannian as described in the following Proposition. For any semi-orthogonal $r \times u$ matrix $\mathbf{G}$, we define $\mathbf{Z}_{\mathbf{G}} = (\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G})^{-1/2}\mathbf{G}^T\mathbf{Y}$ to be the standardized version of $\mathbf{G}^T\mathbf{Y} \in \mathbb{R}^u$ with sample covariance $\mathbf{I}_u$, and let $\widehat{\omega}_i(\mathbf{G})$, $i = 1, \ldots, u$, be the i-th eigenvalue of $\mathbf{S}^{-1}_{\mathbf{Z}_{\mathbf{G}}|\mathbf{X}} = (\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2}(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G})(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2}$.

Proposition 2. *The estimator* $\widehat{\boldsymbol{\Gamma}} = \arg\min_{\mathbf{G}} \mathrm{F}_n(\mathbf{G}|d, u)$ *is the maximizer of* $L_n(\boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{B}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Gamma}|d, u)$, *where the optimization is over* $\mathcal{G}_{r,u}$ *and*

$$\begin{aligned} \mathrm{F}_n(\mathbf{G}|d,u) &= \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}| + \log|\mathbf{I}_u - \mathbf{S}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\circ\mathbf{X}}| \qquad (3.2) \\ &= \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}| + \sum_{i=d+1}^{u} \log[\widehat{\omega}_i(\mathbf{G})]. \qquad (3.3) \end{aligned}$$

We find in practice that the form of objective function (3.3) can be more easily and stably evaluated than (3.2). The analytical expression of $\partial \mathrm{F}_n(\mathbf{G}|d,u)/\partial\mathbf{G}$ based on (3.3) is used to facilitate the Newton-Raphson or conjugate gradient iterations. The formulation in (3.2) describes some operating characteristics of the reduced-rank envelope objective function. Lemma 1 and the relationship $\mathbf{S}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\circ\mathbf{X}} = \mathbf{C}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X}\mathbf{Z}_{\mathbf{G}}}$ implies that the term $\mathbf{I}_u - \mathbf{S}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\circ\mathbf{X}}$ equals the sample covariance of the residuals from reduced-rank regression fit of $\mathbf{Z}_{\mathbf{G}}$ on $\mathbf{X}$ with rank d. Let $\mathrm{F}_{n,1}(\mathbf{G}|u) = \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}|$ and $\mathrm{F}_{n,2}(\mathbf{G}|d,u) = \log|\mathbf{I}_u - \mathbf{S}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\circ\mathbf{X}}|$ so that $\mathrm{F}_n(\mathbf{G}|d,u) = \mathrm{F}_{n,1}(\mathbf{G}|u) + \mathrm{F}_{n,2}(\mathbf{G}|d,u)$. The first part $\mathrm{F}_{n,1}(\mathbf{G}|u) \geq 0$ for all $\mathbf{G} \in \mathcal{G}_{r,u}$ and equals zero when $\mathbf{G}$ is a u-dimensional reducing subspace of $\mathbf{S}_{\mathbf{Y}}$. The effect of $\mathrm{F}_{n,1}(\mathbf{G}|u)$ is

then to pull the solution towards eigenvectors of $\mathbf{S_Y}$. The second part $\mathrm{F}_{n,2}(\mathbf{G}|d,u)$ represents the magnitude of the sample covariance of the residual from reduced-rank regression fit of the standardized variable $\mathbf{Z_G}$ on $\mathbf{X}$ with given rank d. Simply put, this part is a scale-invariant measure for the lack-of-fit of the rank-d reduced-rank regression of $\mathbf{G}^T\mathbf{Y}$ on $\mathbf{X}$.

Our formulation and decomposition based on (3.2) offer a generic way of interpreting the likelihood-based objective functions for envelope methods. For example, the objective function for the standard envelope model in Cook et al. (2010) can be expressed as

$$\log|\mathbf{G}^T\mathbf{S_Y}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_\mathbf{Y}^{-1}\mathbf{G}| + \log|\mathbf{I}_u - \mathbf{S}_{\mathbf{Z_G}\circ\mathbf{X}}|, \tag{3.4}$$

which can be interpreted similar to (3.2) except the lack-of-fit term is now based on ordinary least squares fit rather than reduced-rank regression fit. The above objective function is the same as (3.2) when $d = p$ or $d = u$.

Additional properties of the objective function are given in the following Proposition.

Proposition 3. *The objective function* $\mathrm{F}_n(\mathbf{G}|d,u)$ *in (3.3) converges in probability as* $n \to \infty$ *to the population objective function* $\mathrm{F}(\mathbf{G}|u) = \log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{\Sigma}_\mathbf{Y}^{-1}\mathbf{G}|$ *uniformly in* $\mathbf{G}$. *The estimator* $\widehat{\mathbf{\Gamma}} = \arg\min_\mathbf{G} \mathrm{F}_n(\mathbf{G}|d,u)$ *is Fisher consistent,* $\mathcal{E}_\mathbf{\Sigma}(\boldsymbol{\beta}) = \mathrm{span}\{\arg\min_\mathbf{G} \mathrm{F}(\mathbf{G}|u)\}$.

The population objective function $\mathrm{F}(\mathbf{G}|u)$, which does not depend explicitly on the given rank d, is exactly the same one as in Cook et al. (2010) for estimating an u-dimensional envelope $\mathcal{E}_\mathbf{\Sigma}(\boldsymbol{\beta})$. In the proof of Proposition 3, we show that $\log[\widehat{\omega}_i(\mathbf{G})]$, for any $i > d$, converges in probability to zero uniformly in $\mathbf{G}$. Therefore, we could view $\mathrm{F}_n(\mathbf{G}|d,u)$ in (3.3) as a sample version of $\mathrm{F}(\mathbf{G}|u)$, $\mathrm{F}_n(\mathbf{G}|u) := \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_\mathbf{Y}^{-1}\mathbf{G}|$, plus a finite sample adjustment for the rank deficiency, $\sum_{i=d+1}^{u}\log[\widehat{\omega}_i(\mathbf{G})]$, which goes to zero as $n \to \infty$. Minimizing $\mathrm{F}_n(\mathbf{G}|u)$ leads to another $\sqrt{n}$-consistent envelope estimator but it will not be optimal since it does not account for the rank deficiency. The impact of the rank $d < p$ on the envelope estimation diminishes as sample size increases and reduced-rank envelope estimation moves towards a two-stage estimation procedure: first estimate the envelope from $\mathrm{F}_n(\mathbf{G}|u)$ ignoring the rank, then obtain a rank-d estimator within the estimated envelope. The effects of rank deficiency

and envelope interdigitate at finite samples and there is a noticeable synergy when sample size is not large.

Finally, we summarize estimators for the parameters in the reduced-rank envelope model as follows. The results come naturally from Lemma 2.

Proposition 4. *The estimators for the reduced-rank envelope model (2.4) that minimize (2.2) are* $\widehat{\boldsymbol{\alpha}}_{\mathrm{RE}} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}}\overline{\mathbf{X}}$, $\widehat{\boldsymbol{\Gamma}} = \arg\max_{\mathbf{G}\in\mathcal{G}_{r,u}} \mathrm{F}_n(\mathbf{G}|d,u)$, $\widehat{\boldsymbol{\Omega}}_0 = \widehat{\boldsymbol{\Gamma}}_0^T\mathbf{S}_{\mathbf{Y}}\widehat{\boldsymbol{\Gamma}}_0$ *and*

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}} &= \mathbf{S}^{1/2}_{\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y}}\left\{\mathbf{I}_u - \mathbf{C}^{(d)}_{\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y},\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X},\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y}}\right\}\mathbf{S}^{1/2}_{\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y}} \\
\widehat{\boldsymbol{\Sigma}}_{\mathrm{RE}} &= \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Gamma}}^T + \widehat{\boldsymbol{\Gamma}}_0\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Gamma}}_0^T \\
\widehat{\boldsymbol{\beta}}_{\mathrm{RE}} &= \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\eta}}\widehat{\mathbf{B}}_{\mathrm{RE}} = \widehat{\boldsymbol{\Gamma}}\mathbf{S}^{1/2}_{\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y}}\mathbf{C}^{(d)}_{\widehat{\boldsymbol{\Gamma}}^T\mathbf{Y},\mathbf{X}}\mathbf{S}^{-1/2}_{\mathbf{X}}.
\end{aligned}$$

The rank of $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ is d and the span of $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ is a subset of the entire u-dimensional envelope. In contrast to reduced-rank regression, the estimator for $\widehat{\boldsymbol{\Sigma}}_{\mathrm{RE}}$ now has an envelope structure:

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{RE}} = \mathbf{P}_{\widehat{\boldsymbol{\Gamma}}}\widehat{\boldsymbol{\Sigma}}_{\mathrm{RE}}\mathbf{P}_{\widehat{\boldsymbol{\Gamma}}} + \mathbf{Q}_{\widehat{\boldsymbol{\Gamma}}}\widehat{\boldsymbol{\Sigma}}_{\mathrm{RE}}\mathbf{Q}_{\widehat{\boldsymbol{\Gamma}}}.$$

If we let $u = r$, which is equivalent to setting $\boldsymbol{\Gamma} = \mathbf{I}_r$ in Proposition 4, then there is no envelope reduction and the estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ is the same as the estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ in Lemma 1. If we let $d = p$, then the estimators in Proposition 4 is the same as the envelope estimators in Cook et al. (2010). The estimators for the reduced-rank envelope model parameters coincide with those estimators in Lemma 2 by replacing $\boldsymbol{\Gamma}$ by its estimator $\widehat{\boldsymbol{\Gamma}}$.

4 Asymptotics

4.1 Asymptotic properties under normality

In this section, we present asymptotic results assuming that the error term is normal, $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$, so that the estimators derived in Section 3 are all maximum likelihood estimators. We focus attention on the comparison between $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ because (1) comparisons between $\widehat{\boldsymbol{\beta}}_{\mathrm{Env}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ can be found in Cook et al. (2010); and (2) the advantage of $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ over $\widehat{\boldsymbol{\beta}}_{\mathrm{Env}}$ is

similar to the advantage of $\widehat{\boldsymbol{\beta}}_{\mathrm{RR}}$ over $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, which is due to the rank reduction in the material response $\boldsymbol{\Gamma}^T\mathbf{Y}$. We then relax the normality assumption in Section 4.2 and show the $\sqrt{n}$-consistency of the reduced-rank envelope estimator and its asymptotic distribution.

From Cook et al. (2010) we know that the Fisher information for $\mathbf{h}$ is

$$\mathbf{J}_{\mathbf{h}} = \begin{pmatrix} \mathbf{J}_{\boldsymbol{\beta}} & 0 \\ 0 & \mathbf{J}_{\boldsymbol{\Sigma}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Sigma}^{-1} & 0 \\ 0 & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{E}_r \end{pmatrix}, \tag{4.1}$$

where $\boldsymbol{\Sigma}_{\mathbf{X}} = \lim_{n\to\infty} \mathbf{S}_{\mathbf{X}}$ and $\mathbf{E}_r$ is the expansion matrix, $\mathbf{E}_r\mathrm{vec}(\mathbf{S}) = \mathrm{vech}(\mathbf{S})$ for any $r \times r$ symmetric matrix $\mathbf{S}$. The asymptotic covariance for the ordinary least squares estimator $\widehat{\mathbf{h}}_{\mathrm{OLS}}$ is $\mathbf{J}_{\mathbf{h}}^{-1}$, which is the asymptotic covariance of the unrestricted maximum likelihood estimator.

Define the gradient matrices

$$\mathbf{H} = \frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \text{ and } \mathbf{R} = \frac{\partial \mathbf{h}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}. \tag{4.2}$$

Then the asymptotic covariance for the reduced-rank regression estimator $\widehat{\mathbf{h}}_{\mathrm{RR}} = \mathbf{h}(\widehat{\boldsymbol{\psi}})$ and for the reduced-rank envelope estimator $\widehat{\mathbf{h}}_{\mathrm{RE}} = \mathbf{h}(\widehat{\boldsymbol{\phi}})$ are summarized in the following Proposition.

Proposition 5. *Assuming that* $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$*, then* $\mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{OLS}}) = \mathbf{J}_{\mathbf{h}}^{-1}$, $\mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{RR}}) = \mathbf{H}(\mathbf{H}^T\mathbf{J}_{\mathbf{h}}\mathbf{H})^{\dagger}\mathbf{H}^T$ *and* $\mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{RE}}) = \mathbf{R}(\mathbf{R}^T\mathbf{J}_{\mathbf{h}}\mathbf{R})^{\dagger}\mathbf{R}^T$*. Moreover,*

$$\begin{aligned} \mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{OLS}}) - \mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{RR}}) &= \mathbf{J}_{\mathbf{h}}^{-1/2}\mathbf{Q}_{\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{H}}\mathbf{J}_{\mathbf{h}}^{-1/2} \geq 0, \\ \mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{RR}}) - \mathrm{avar}(\sqrt{n}\widehat{\mathbf{h}}_{\mathrm{RE}}) &= \mathbf{J}_{\mathbf{h}}^{-1/2}\left(\mathbf{P}_{\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{H}} - \mathbf{P}_{\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{R}}\right)\mathbf{J}_{\mathbf{h}}^{-1/2} \\ &= \mathbf{J}_{\mathbf{h}}^{-1/2}\mathbf{P}_{\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{H}}\mathbf{Q}_{\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{R}}\mathbf{J}_{\mathbf{h}}^{-1/2} \geq 0, \end{aligned}$$

where † *indicates the Moore-Penrose inverse. In particular,* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})] \geq \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})] \geq \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$.

Proposition 5 follows directly from $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{\phi})$. Therefore, we have $\mathbf{R} = \mathbf{H}\partial\boldsymbol{\psi}(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ and $\mathrm{span}(\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{R}) \subseteq \mathrm{span}(\mathbf{J}_{\mathbf{h}}^{1/2}\mathbf{H})$. Similarly, it can be shown that $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})] \geq \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{Env}})] \geq \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$.

Since we are particularly interested in the asymptotic covariance of $\widehat{\mathbf{h}}_1 = \mathrm{vec}(\widehat{\boldsymbol{\beta}})$ from different estimators, we summarize some of the results in the following Propositions.

Proposition 6. *Assume that* $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$ *and that* $\mathrm{rank}(\boldsymbol{\beta}) = d$*. Then* $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})$ *and* $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}} - \boldsymbol{\beta})$ *are both asymptotically normal with mean zero and covariances as follows.*

$$\begin{aligned} \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})] &= \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma} \\ \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})] &= (\mathbf{I}_{pr} - \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})})\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})] \qquad (4.3) \\ &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{A}}\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})})] + \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{B}})] \qquad (4.4) \end{aligned}$$

where $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{A}})] = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes (\mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\boldsymbol{\Sigma})$ *and* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{B}})] = (\mathbf{P}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}) \otimes \boldsymbol{\Sigma}$*.*

The asymptotic result in (4.3) follows from Anderson (1999; equation (3.20)). The results in Proposition 6 rely on $\mathbf{A}$ and $\mathbf{B}$ only through their projections $\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}$ and $\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}$, which serve to orthogonalize the parameters in the asymptotic variance decompositions. This implies that all the equalities in Proposition 6 hold for any decomposition $\boldsymbol{\beta} = \mathbf{AB}$, with $\mathbf{A} \in \mathbb{R}^{r\times d}$ and $\mathbf{B} \in \mathbb{R}^{d\times p}$. Hence, Proposition 6 is a unification for all the asymptotic studies of reduced-rank regression in the literature such as Anderson (1999), Reinsel and Velu (1998), Stoica and Viberg (1996) and so on.

For the reduced-rank envelope model (2.4), we have the following results on asymptotic distributions.

Proposition 7. *Under the reduced-rank envelope model with normal error* $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$*,* $\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}} - \boldsymbol{\beta})$ *is asymptotically normal with mean zero and covariance*

$$\begin{aligned} \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})] &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}})] + \mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})] \\ &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}}\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})})] + \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\mathbf{B}})] \\ &\quad + \mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})], \qquad (4.5) \end{aligned}$$

where $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}}\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})})] = \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{A}}\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})})]$ *from (4.4). Explicit expressions for* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$ *can be found in the Supplemental material (D.7). The above equalities hold for any decomposition* $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{B}$*, where* $\boldsymbol{\Gamma}$ *is semi-orthogonal and the dimensions of* $\boldsymbol{\Gamma}$*,* $\boldsymbol{\eta}$ *and* $\mathbf{B}$ *are* $r \times u$*,* $u \times d$ *and* $d \times p$*.*

We view the asymptotic advantages of reduced-rank envelopes over reduced-rank regression by contrasting (4.4) with (4.5). From Propositions 6 and 7, we can write $\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{RR}})] - \text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{RE}})]$ as

$$\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\mathbf{B}})] - \text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\mathbf{B}})] - \text{avar}[\sqrt{n}\text{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})] \geq 0, \tag{4.6}$$

where $\widehat{\boldsymbol{\beta}}_{\mathbf{B}} = \widehat{\mathbf{A}}_{\mathbf{B}}\mathbf{B}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\mathbf{B}} = \boldsymbol{\Gamma}\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma},\mathbf{B}}\mathbf{B} = \widehat{\mathbf{A}}_{\boldsymbol{\Gamma},\mathbf{B}}\mathbf{B}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}} = \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta},\mathbf{B}}\boldsymbol{\eta}\mathbf{B} = \widehat{\mathbf{A}}_{\boldsymbol{\eta},\mathbf{B}}\mathbf{B}$ are estimators with given $\mathbf{B}$. When $\mathbf{B}$ is known, the original regression problem simplifies to the regression of $\mathbf{Y}$ on $\mathbf{BX}$ and $\mathbf{A}$ is the new regression coefficient matrix. The estimator $\widehat{\mathbf{A}}_{\mathbf{B}}$ is the ordinary least squares estimator of $\mathbf{Y}$ on $\mathbf{BX}$ and the estimators $\widehat{\mathbf{A}}_{\boldsymbol{\Gamma},\mathbf{B}}$ and $\widehat{\mathbf{A}}_{\boldsymbol{\eta},\mathbf{B}}$ correspond to the usual envelope estimators for $\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\eta}$. The difference in asymptotic covariances $\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{RR}})] - \text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{RE}})]$ from (4.6) equals the asymptotic efficiency gain of envelope estimator over the ordinary least squares estimator for regression of $\mathbf{Y}$ on $\mathbf{BX}$ and is consistent with the results presented in Cook et al. (2010).

Two special situations where the inequality in (4.6) becomes equality are: $\boldsymbol{\Gamma} = \mathbf{I}_r$ and $\boldsymbol{\Sigma} = \boldsymbol{\sigma}^2\mathbf{I}_r$, while the envelope estimator is asymptotically equivalent to the ordinary least squares estimator in these two cases.

To see the potential gain of the reduced-rank envelope estimator, we have the following Corollary, where we have ignored the cost of estimating an envelope.

Corollary 1. *Under the reduced-rank envelope model with normal error $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$,*

$$\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}})] = \mathbf{F}_1\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{Env},\boldsymbol{\Gamma}})] = \mathbf{F}_2\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{RR}})] = \mathbf{F}_1\mathbf{F}_2\text{avar}[\sqrt{n}\text{vec}(\widehat{\boldsymbol{\beta}}_{\text{OLS}})],$$

where $\mathbf{F}_1 = \mathbf{I}_{pr} - \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}$ and $\mathbf{F}_2 = \mathbf{I}_p \otimes \mathbf{P}_{\boldsymbol{\Gamma}}$ are two positive semi-definite matrices with eigenvalues between 0 and 1.

The two matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ represent fractions of asymptotic covariance reduction from the ordinary least squares estimator to the reduced-rank regression estimator and to the envelope estimator with given $\boldsymbol{\Gamma}$. Then the efficiency gain of reduced-rank envelope with known $\boldsymbol{\Gamma}$ over ordinary least squares is the superimposition of the efficiency gain of the reduced-rank regression and the envelope regression with known $\boldsymbol{\Gamma}$.

4.2 Consistency without the normality assumption

Let $\widehat{\mathbf{h}}_{\rm OLS} = \left(\mathrm{vec}^T(\widehat{\boldsymbol{\beta}}_{\rm OLS}), \mathrm{vech}^T(\mathbf{S}_{\mathbf{Y}|\mathbf{X}})\right)^T$ denote the ordinary least squares estimator of $\mathbf{h}$ under the standard linear regression model, and let $\widehat{\mathbf{h}}_{\rm RE} = \mathbf{h}(\widehat{\boldsymbol{\phi}})$ denote the reduced-rank envelope estimator. The true values of $\mathbf{h}$ and $\boldsymbol{\phi}$ are denoted as $\mathbf{h}_0$ and $\boldsymbol{\phi}_0$. The objective function $L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ in (2.2) can be written as, after partially maximized over $\boldsymbol{\alpha}$,

$$L_n(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \simeq -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{n}{2}\mathrm{trace}\left\{\boldsymbol{\Sigma}^{-1}[\mathbf{S}_{\mathbf{Y}|\mathbf{X}} + (\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta})\mathbf{S}_{\mathbf{X}}(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta})^T]\right\}. \quad (4.7)$$

We treat the objective function $L_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ as a function of $\mathbf{h}$ and $\widehat{\mathbf{h}}_{\rm OLS}$ and define $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\rm OLS}) = 2/n\left\{L_n(\widehat{\boldsymbol{\beta}}_{\rm OLS}, \mathbf{S}_{\mathbf{Y}|\mathbf{X}}) - L_n(\boldsymbol{\beta}, \boldsymbol{\Sigma})\right\}$, which satisfies the conditions of Shapiro's (1986) minimum discrepancy function (see Supplement Section F). Hence $\mathbf{J}_{\mathbf{h}} = 1/2\times\partial^2\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\rm OLS})/\partial\mathbf{h}\partial\mathbf{h}^T$ evaluated at $\widehat{\mathbf{h}}_{\rm OLS} = \mathbf{h} = \mathbf{h}_0$ is the Fisher information matrix for $\mathbf{h}$ when $\boldsymbol{\epsilon}$ is normal. The following proposition formally states the asymptotic distribution of $\widehat{\mathbf{h}}_{\rm RE}$ without normality of $\boldsymbol{\epsilon}$.

Proposition 8. *Assume that the reduced-rank envelope model (2.4) holds and that $\boldsymbol{\epsilon}_i$'s are independent and identically distributed with finite fourth moments. Then $\sqrt{n}(\widehat{\mathbf{h}}_{\rm OLS} - \mathbf{h}_0) \to N(0, \mathbf{K})$, for some positive definite covariance matrix $\mathbf{K}$. And $\sqrt{n}(\widehat{\mathbf{h}}_{\rm RE} - \mathbf{h}_0)$ converges in distribution to a normal random variable with mean $\mathbf{0}$ and covariance matrix*

$$\mathbf{W} = \mathbf{R}\left(\mathbf{R}^T\mathbf{J}_{\mathbf{h}}\mathbf{R}\right)^{\dagger}\mathbf{R}^T\mathbf{J}_{\mathbf{h}}\mathbf{K}\mathbf{J}_{\mathbf{h}}\mathbf{R}\left(\mathbf{R}^T\mathbf{J}_{\mathbf{h}}\mathbf{R}\right)^{\dagger}\mathbf{R}^T.$$

In particular, $\sqrt{n}(\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\rm RE}) - \mathrm{vec}(\boldsymbol{\beta}))$ converges in distribution to a normal random variable with mean $\mathbf{0}$ and covariance $\mathbf{W}_{11}$, the upper-left $pr \times pr$ block of $\mathbf{W}$. The explicit expression for the gradient matrix $\mathbf{R} = \partial\mathbf{h}(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ is given in the Supplement equation (D.1).

The $\sqrt{n}$-consistency of the reduced-rank envelope estimator $\widehat{\boldsymbol{\beta}}_{\rm RE}$ is essentially because that $\widehat{\boldsymbol{\beta}}_{\rm OLS}$ and $\mathbf{S}_{\mathbf{Y}|\mathbf{X}}$ are $\sqrt{n}$-consistent regardless of normality assumption and also because of the properties of $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\rm OLS})$. The asymptotic covariance matrix $\mathbf{W}_{11}$ can be estimated straightforwardly using the plug-in method once $\mathbf{K}$ is estimated, but its accuracy for any fixed sample size will depend on the distribution of $\boldsymbol{\epsilon}$, which is usually unknown in practice. Fortunately, bootstrap methods can provide good estimates of $\mathbf{W}_{11}$, as illustrated in Section 6.3.

5 Selections of rank and envelope dimension

5.1 Rank: $d = \mathrm{rank}(\boldsymbol{\beta})$

Bura and Cook (2003) developed a chi-squared test for the rank d that requires only that the response variables have finite second moments. The test statistic is $\Lambda_d = n \sum_{j=d+1}^{\min(p,r)} \varphi_j^2$, where $\varphi_1 \geq \cdots \geq \varphi_{\min(p,r)}$ are eigenvalues of the $p \times r$ matrix

$$\widehat{\boldsymbol{\beta}}_{\mathrm{std}} = \{(n-p-1)\}/n\}^{1/2} \mathbf{S}_{\mathbf{X}}^{1/2} \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} \mathbf{S}_{\mathbf{Y}|\mathbf{X}}^{-1/2}. \tag{5.1}$$

Under the null hypothesis that $H_0 : d = d_0$, Bura and Cook (2003) showed that Λ_{d_0} is asymptotically distributed as a $\chi^2_{(p-d_0)(r-d_0)}$ random variable. The rank d is then determined by comparing a sequence of test statistics Λ_{d_0}, $d_0 = 0, \ldots, \min(p,r) - 1$, to the percentiles of their null distribution $\chi^2_{(p-d_0)(r-d_0)}$. The sequence of tests terminates at the first non-significant test of $H_0 : d = d_0$ and then d_0 serves as an estimate of the rank of $\boldsymbol{\beta}$.

5.2 Envelope dimension: $u = \dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta}))$

Since the envelope dimension satisfies $d \leq u \leq r$, standard techniques such as sequential likelihood-ratio tests, AIC and BIC can be applied to select u, as in Cook et al. (2010).

For any possible combination (d, u) with $0 \leq d \leq u \leq r$, let $\widehat{L}_{d,u}$ denote the maximized log-likelihood function (c.f. (A.3)), which is evaluated at the maximum likelihood estimators in Proposition 4. Assuming d is known, then $\Lambda_{d,u_0} = 2(\widehat{L}_{d,r} - \widehat{L}_{d,u_0})$ is asymptotically distributed as a $\chi^2_{(r-u_0)d}$ random variable under the null hypothesis $H_0 : u = u_0$. Thus, a sequence of likelihood ratio tests of $u_0 = d, \ldots, r-1$ can be used to determine u after d is determined by the method described in Section 5.1. The first non-significant value of u_0 will serve as the envelope dimension.

Information criteria such as AIC and BIC can be used to select (d, u) simultaneously. We write AIC as $\mathcal{A}_{d,u} = 2K_{d,u} - 2\widehat{L}_{d,u}$, where $K_{d,u} = (p+u-d)d + r(r+1)/2$ is the total number of parameters in the reduced-rank envelope model, and write BIC as $\mathcal{B}_{d,u} = \log(n)K_{d,u} - 2\widehat{L}_{d,u}$. We search (d, u) from $(0, 0)$ to (r, r) with constraint $d \leq u$ and choose the pair that has the

smallest AIC or BIC. Alternatively, we can first determine d from the asymptotic chi-squared tests in Section 5.1 and then search u from $d, \ldots, r$ with the smallest AIC or BIC, which could save a lot of computation. The computation cost for determining d by the sequential chi-squared tests in Section 5.1 is substantially cheaper than the computation cost in calculating AIC and BIC, which involves sequence of Grassmannian optimizations.

When sample size is not too small, our experience suggests that the most favorable procedure is BIC selection for $u = d, \ldots, r$ where d is guided by the sequential chi-squared tests. Since the true envelope dimension always exist, BIC is consistent in the sense that the probability of selecting the correct u approaches 1, given the correct d. There are many articles comparing AIC and BIC, from both theoretical and practical points of view, for example Shao (1997) and Yang (2005).

The rank d and envelope dimension u can also be determined by cross-validation or by using hold-out samples. These approaches are especially appropriate when prediction is the primary goal of the study rather than correctness of the selected model.

6 Simulations

6.1 Rank and dimension

In all the simulations, we first filled in $\mathbf{\Gamma}$, $\boldsymbol{\eta}$ and $\mathbf{B}$ with random uniform (0,1) numbers, and then $\mathbf{\Gamma}$ was standardized so that $\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{I}_u$ and $\boldsymbol{\beta} = \mathbf{\Gamma}\boldsymbol{\eta}\mathbf{B}$ was standardized so that $||\boldsymbol{\beta}||_F = 1$. Estimation errors were defined as $||\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}||_F$. Unless otherwise specified, the predictors and errors were simulated independently from $N(0, \mathbf{I}_p)$ and $N(0, \mathbf{\Sigma})$ distributions. All figures were generated based on averaging over 200 independent replicate data sets.

In this section, we present simulation results to demonstrate the behavior of the proposed method using various sample sizes, ranks d, dimensions u. We simulated data from model (2.4), where $[\mathbf{\Omega}]_{ij} = (-0.9)^{|i-j|}$ and $[\mathbf{\Omega}_0]_{ij} = 5 \cdot (-0.5)^{|i-j|}$. Figure 6.1 summarizes the effect of dimension and rank on the relative performances of each methods. In the left plot $(d, u, p, r) = (1, 10, 10, 20)$. Since the rank was only one but the envelope dimension was ten,

reduced-rank regression had a dramatic improvement over ordinary least squares, while the ordinary envelope method had a relatively modest gain over ordinary least squares. The reduced-rank envelope had a relatively small edge over reduced-rank regression. The second case was $(d, u, p, r) = (4, 5, 6, 20)$, where $\boldsymbol{\beta}$ had nearly full column rank and the envelope dimension was much smaller than the number of response variables. Not surprisingly, reduced-rank regression had modest gain over ordinary least squares while the envelope estimator and the reduced-rank envelope estimator had similar behavior and significantly improved over ordinary least squares and reduced-rank regression. The last case was chosen as $(d, u, p, r) = (5, 10, 15, 20)$ so that there was no particular favor towards either the envelope method or reduced-rank regression. We found good improvement over ordinary least squares by both reduced-rank regression and envelopes. However, reduced-rank envelopes combined both of their strengths and resulted in a bigger gain.

We found in practice that reduced-rank envelopes typically have improved performance over reduced-rank regression and envelope estimators, and it has similar behavior to one of the two estimators if the other one performed poorly. Even in the extreme cases where $d = p$ or $u = r$, reduced-rank envelopes can still gain drastically over ordinary least squares similar to the results in Figure 6.1.

We next illustrate the asymptotic chi-squared test for rank detection combined with BIC selection for envelope dimension, as discussed in Section 5. Using the same simulation model, we took $(d, u, p, r) = (3, 5, 8, 12)$, where the total number of parameters in the reduced-rank envelope model was 108. The percentages of correct detections for d an u were plotted in Figure 6.2 versus sample size. The BIC selection of u was based on the correct rank d. The significance level of the chi-squared tests was 0.05. As seen from the figure, the probability of selecting the correct d was about 0.9 at $n = 400$ samples and the probability of correct detection settled at 95% for larger n, as predicted by the hypothesis testing theory. BIC selection for the envelope dimension u seemed to be very accurate even with small samples. The likelihood-ratio tests and AIC selection for u were not nearly as effective as BIC and thus were omitted from the plot.

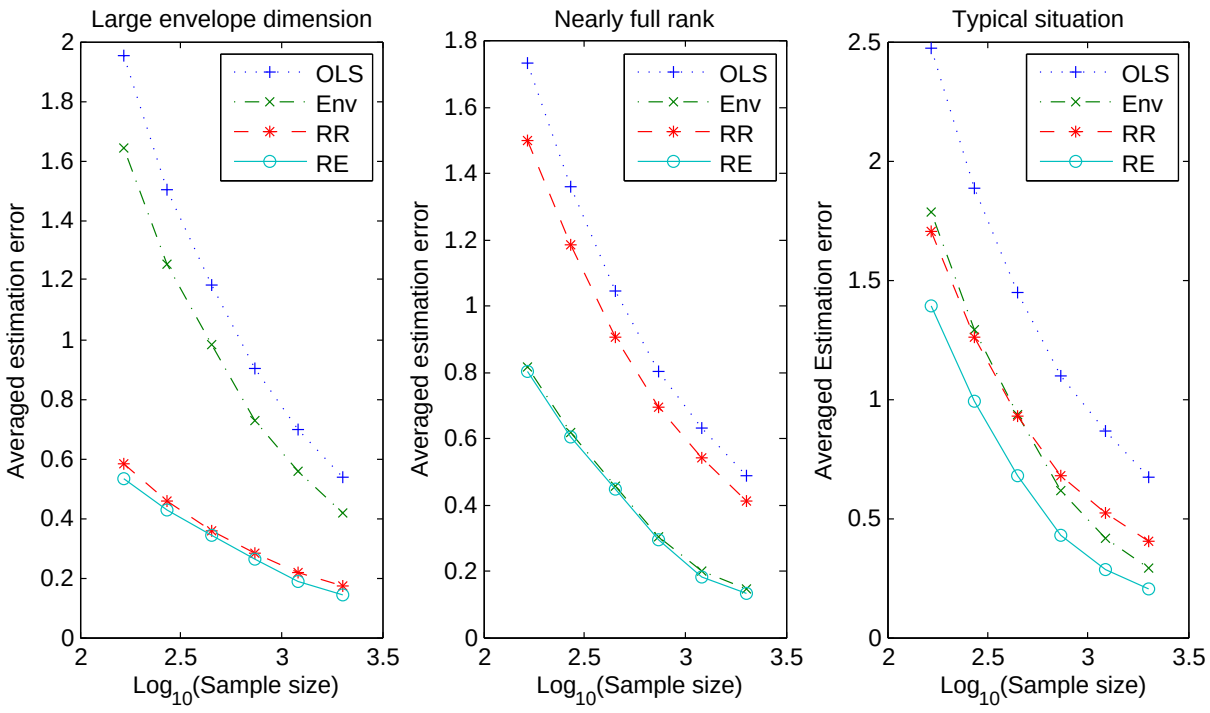

Figure 6.1: Effect of rank and dimension. Averaged estimation error on the vertical axis is defined as averaged $||\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}||_F$ over 200 independent data sets. The dimensions of the three plots were: (1) large envelope dimension case $(d, u, p, r) = (1, 10, 10, 20)$; (2) nearly full rank case $(d, u, p, r) = (4, 5, 6, 20)$ and (3) a typical situation $(d, u, p, r) = (5, 10, 15, 20)$. The sample sizes varied from 160 to 2000 and were shown in a logarithmic scale.

We also considered BIC selection for u and d simultaneously. The probability of simultaneous correctness was less than 70% for $n \leq 600$ but reached more than 95% correctness for $n \geq 900$. In our experience the best method for determining dimensions is to use the chi-squared test for d and BIC selection on u based on the selected d. Overestimation of d and u usually is not a serious issue but underestimation of d and u will certainly cause bias in estimation.

6.2 Signal-versus-noise and material-versus-immaterial

In this section, we describe the behavior of each method with varying signal-to-noise ratios and ratios of immaterial variation to material variation. We fixed the sample size at 400 and the dimensions were $(d, u, p, r) = (3, 7, 10, 20)$. The covariances had the forms of $[\boldsymbol{\Omega}]_{ij} = \sigma^2 \cdot (-0.9)^{|i-j|}$ and $[\boldsymbol{\Omega}_0]_{ij} = \sigma_0^2 \cdot (-0.9)^{|i-j|}$ with varying constants $\sigma^2,\ \sigma_0^2 > 0$.

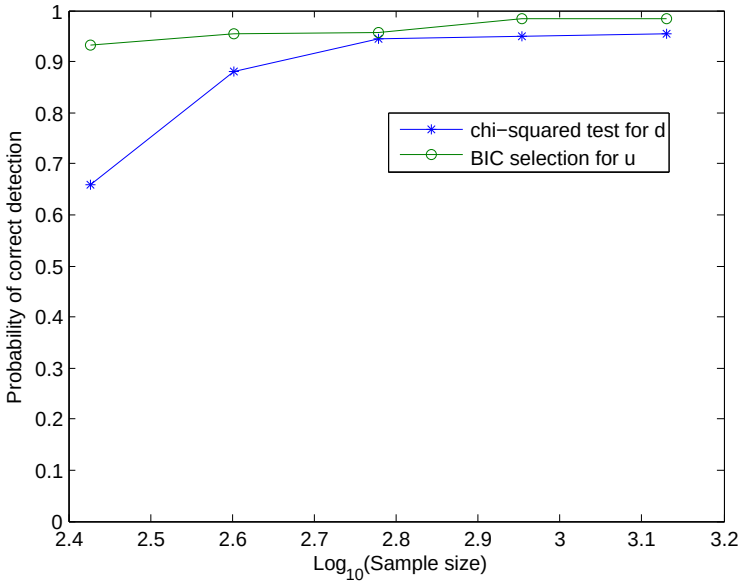

Figure 6.2: The empirical probability of correct detection versus sample sizes. BIC selection on u was based on true rank d.

In the study of varying signal-to-noise ratio, we kept $\sigma^2 = \sigma_0^2$. And because $||\boldsymbol{\beta}||_F = 1$, the signal-to-noise ratio was simply $1/\sigma^2$ which varied from 0.1 to 10. Figure 6.3 summarizes the results of two numerical experiments. All the four lines in this log-log scale signal-to-noise ratio plot are roughly parallel, which implies that the four methods are exponentially more distinguishable in weaker signal. Comparing reduced-rank regression to envelopes, the reduced-rank regression seemed to perform better in stronger signals (signal-to-noise ratio$\geq$ 1), but the envelope estimator was less vulnerable to weaker signals (signal-to-noise ratio$\leq$ 1). This was because the envelope method can gain information from the error term $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T$ while reduced-rank regression and the standard method cannot. Reduced-rank envelope estimators combined the strengths of reduced-rank regression and envelopes, and hence outperformed both estimators in strong and weak signals.

In the study of varying immaterial-to-material variance ratio, we kept $\sigma^2 = 1$ and changed σ_0^2. The ratio is then defined as σ_0^2 and the horizontal axis in the plot is $\log_{10}(\sigma_0^2)$, which varied from -0.5 to 2. Not surprisingly, reduced-rank regression and ordinary least squares behaved similarly because they did not gain information from the covariance structure of $\boldsymbol{\Sigma}$.

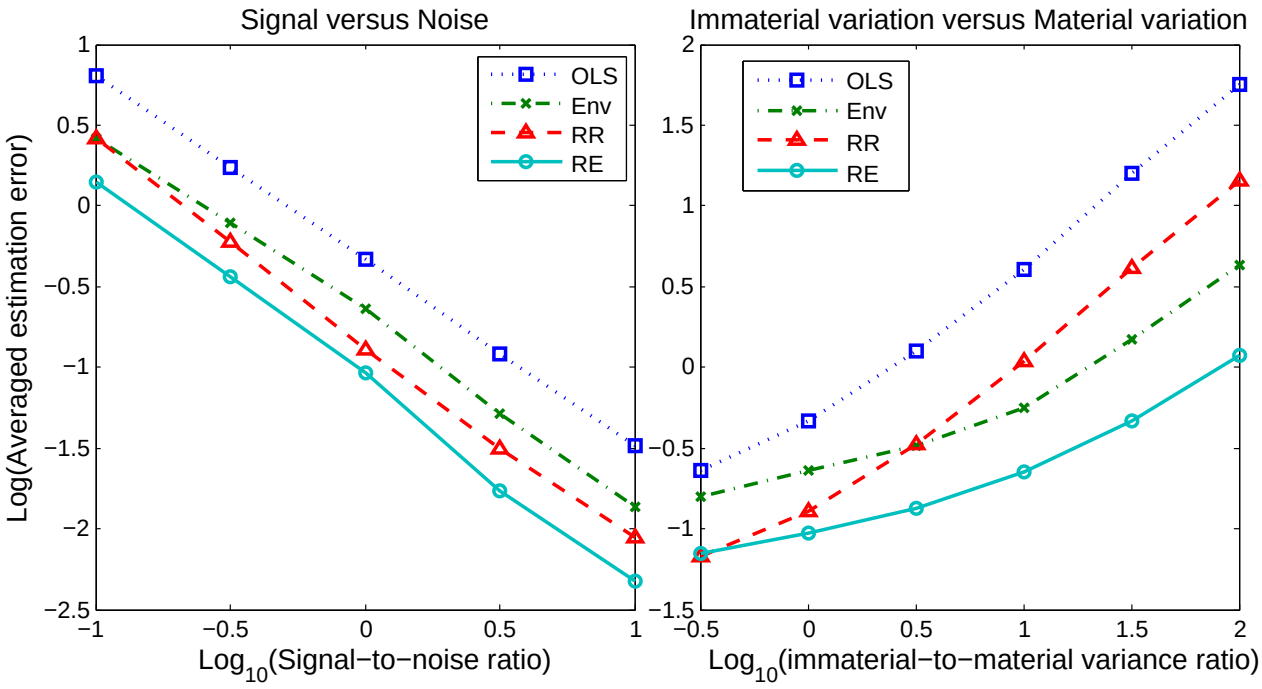

Figure 6.3: Varying the signal-to-noise ratio and the immaterial-to-material variance ratio.

The envelope estimator and the reduced-rank envelope estimator had similar behavior, and they had much better performances over ordinary least squares and reduced-rank regression when the immaterial variation was large. This is due to the fact that envelope methods can efficiently eliminate the immaterial information. In this example, the averaged estimation errors for ordinary least squares, reduced-rank regression and envelope were 7.2, 3.9 and 1.8 times of that of the envelope reduced-rank regression when $\sigma_0^2 = 100$.

6.3 Bootstrap standard errors

To illustrate the application of the bootstrap for estimating the standard errors of regression coefficients, we considered a model with $(d, u, p, r) = (2, 4, 6, 8)$. Residual bootstrap samples were used since we considered $\mathbf{X}$ as a non-stochastic predictor. Both $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ were randomly generated as $\mathbf{M}\mathbf{M}^T$, where $\mathbf{M} \in \mathbb{R}^{4\times 4}$ was filled with uniform (0,1) numbers. The error term $\boldsymbol{\epsilon}_i$ was simulated as $\boldsymbol{\epsilon}_i = \boldsymbol{\Sigma}^{1/2}\mathbf{U}_i$, where $\mathbf{U}_i$ was a vector of i.i.d. random variable with mean 0 standard deviation 1. We simulated both normal and uniform $\mathbf{U}_i$. The standard errors of a se-

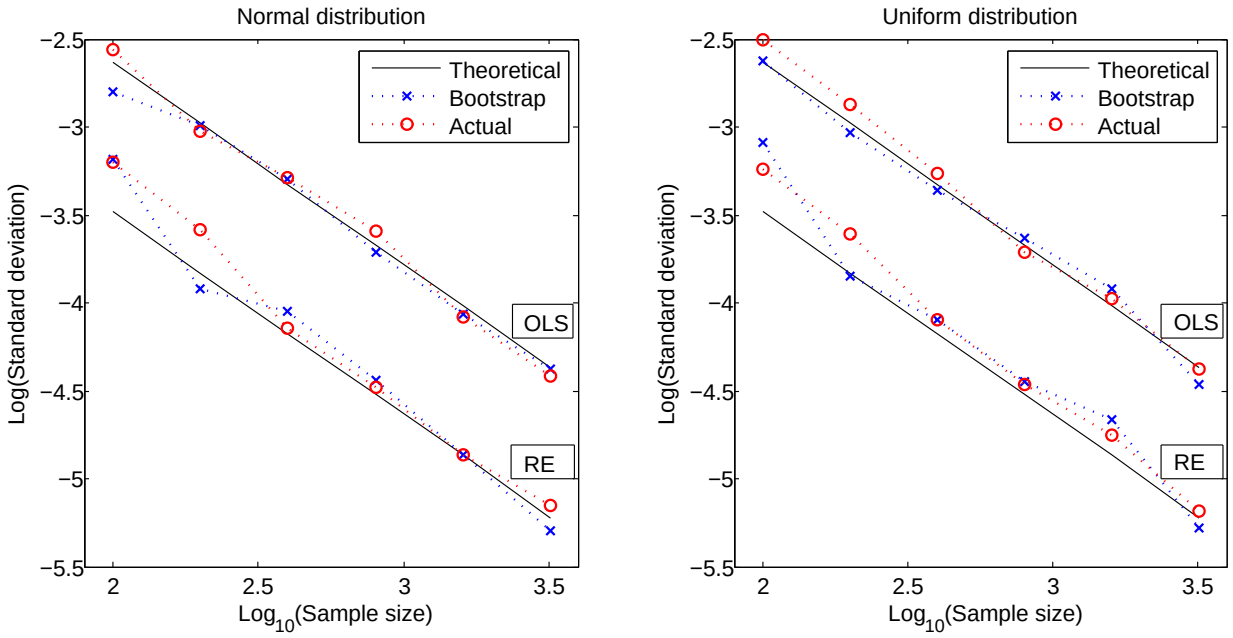

Figure 6.4: Theoretical, bootstrap and actual standard errors with normal and uniform errors $\boldsymbol{\epsilon}$. The sample sizes were 100, 200, 400,. . . , 3200. The standard errors for reduced-rank regression and envelope estimators were consistently between the ordinary least squares and the envelope reduced-rank regression standard errors, and were not included in these plots for better visualization.

lected element in $\widehat{\boldsymbol{\beta}}$ were plotted in Figure 6.4. For both normal and non-normal data, the three types of standard error estimates agreed well: the theoretical standard errors were the squared roots of the diagonal elements in the asymptotic covariances of each estimators divided by $\sqrt{n}$; the actual standard errors were based on 200 independent realizations; and the bootstrap standard errors were based on 200 bootstrap replicate data sets. Moreover, the bootstrap standard errors were close to the theoretical standard errors of the maximum likelihood estimators even when the normality assumption was violated. As expected, the reduced-rank envelope estimator had much smaller standard errors than those of the ordinary least squares estimator. We also simulated non-normal errors from t-distribution and χ^2-distribution, and obtained results similar to Figure 6.4.

7 Sales people test scores data

This data set consisted of 50 sales people from a firm. Three performances variables were used as predictors: growth of sales (X_1), profitability of sales (X_2) and new account sales (X_3). And four response variables were test scores on creativity (Y_1), mechanical reasoning (Y_2), abstract reasoning (Y_3) and mathematical ability (Y_4). This data set can be found in Johnson and Wichern (2007).

The chi-squared rank test in Section 5.1 suggested that $d = 2$ at level 0.01. Then based on BIC we selected the envelope dimension to be $u = 3$. We computed the fractions $f_{ij} := 1 - \widehat{\text{avar}}^{1/2}(\sqrt{n}\widehat{\boldsymbol{\beta}}_{ij,\text{RE}})/\widehat{\text{avar}}^{1/2}(\sqrt{n}\widetilde{\boldsymbol{\beta}}_{ij})$ for all i and j, where $\widetilde{\boldsymbol{\beta}}$ denotes one of the estimators to be compared: $\widehat{\boldsymbol{\beta}}_{\text{RR}}$, $\widehat{\boldsymbol{\beta}}_{\text{Env}}$ and $\widehat{\boldsymbol{\beta}}_{\text{OLS}}$. Comparing to ordinary least squares , the standard deviations of the elements in the reduced-rank envelope estimator were 5% to 60% smaller, $0.05 \leq f_{ij} \leq 0.60$. Hypothetically, a sample size of more than 300 observations, in contrast to the original 50 observations, would be needed to achieve a 60% smaller standard deviation in ordinary least squares. The fractions for comparing with the reduced-rank regression estimator were $0.01 \leq f_{ij} \leq 0.24$, where 24% smaller standard deviation than reduced-rank regression implies a doubling of the observations for reduced-rank regression to achieve the same performance as reduced-rank envelope estimator. At last, the reduced-rank envelope estimator compared to the ordinary envelope estimator, had 3% to 51% smaller standard deviations, where 51% smaller standard deviation meant four times the sample size, $n = 200$, for the ordinary envelope estimator.

Supplementary Materials

Proofs and Technical Details: Detailed proofs for all Lemmas and Propositions are provided in the online supplement to this article. (PDF file)

References

[1] ANDERSON, T. W. (1951), Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, **22**, 327–351.

[2] ANDERSON, T. W. (1999), Asymptotic distribution of the reduced rank regression estimator under general conditions. *The Annals of Statistics*, **27**, 1141–1154.

[3] ANDERSON, T. W. (2002), Canonical correlation analysis and reduced rank regression in autoregressive models. *The Annals of Statistics*, **30**, 1134–1154.

[4] ADRAGNI, K., COOK, R.D. AND WU, S. (2012), GrassmannOptim: An R Package for Grassmann Manifold Optimization. *Journal of Statistical Software*, **50**, 1–18.

[5] BURA, E. AND COOK, R.D. (2003), Rank estimation in reduced-rank regression. *Journal of Multivariate Analysis*, **87**, 159–176.

[6] CHEN, L. AND HUANG, J.Z. (2012), Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *Journal of the American Statistical Association*, **107**, 1533–1545.

[7] CHEN, K., CHAN, K.S. AND STENSETH, N.C. (2012), Reduced rank stochastic regression with a sparse singular value decomposition. *JRSS-B*, **74**, 203–221.

[8] CONWAY, J. (1990). A course in functional analysis. Second edition. *Springer, New York*.

[9] COOK, R. D., HELLAND, I. S. AND SU, Z. (2013), Envelopes and partial least squared regression. *To appear in JRSS-B*.

[10] COOK, R. D., LI, B. AND CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression (with discussion). *Statistica Sinica*, **20**,927–1010.

[11] EDELMAN, A., TOMAS, A.A. AND SMITH, S.T. (1998), The geometry of algorithms with orthogonality constraints, *SIAM Journal of Matrix Analysis and Applications*, **20**, 303–353.

[12] HENDERSON, H. V. AND SEARLE, S. R. (1979), Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canadian Journal of Statistics*, **7**, 65–81.

[13] IZENMAN, A. J. (1975), Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*. **5**, 248–264.

[14] REINSEL, G. C. AND VELU, R. P. (1998), Multivariate reduced-rank regression: theory and applications. *Springer, New York*.

[15] SHAO, J. (1997), An asymptotic theory for linear model selection (with discussion). *Statistica sinica*, **7**, 221–264.

[16] SHAPIRO, A. (1986), Asymptotic theory of overparameterized structural models, *Journal of the American Statistical Association*, **81**, 142–149.

[17] STOICA, P. AND VIBERG, M. (1996), Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Transactions on Signal Processing*, **44**, 3069–3079.

[18] SU, Z. AND COOK, R.D. (2011), Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, **98**, 133–146.

[19] TSO, M.K.S. (1981), Reduced-rank regression and canonical analysis. *JRSS-B*, **43**, 183–189.

[20] TYLER, D.E. (1981). Asymptotic inference for eigenvectors. *Annals of Statistics*, **9**, 725–736.

[21] YANG, Y. (2005), Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika*. **92**, 934–950.

Supplement: Proofs and Technical Details for "Reduced-rank Envelope Model"

R. Dennis Cook, Liliana Forzani and Xin Zhang

A Maximizing the likelihood-based objective function (2.2)

In this Section, we consider maximizing $L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ from (2.2) under different model parameterizations regarding standard, reduced-rank, envelope and reduced-rank envelope models. Maximizing L_n from (2.2) is equivalent to deriving maximum likelihood estimators with normally distributed error $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$ as follows. Lemmas 1 and 2 and Propositions 2 and 4 are proved directly in the derivation of estimators.

A.1 Standard regression and envelope regression

Maximum likelihood estimators for the standard regression model is the ordinary least squares estimator, $\widehat{\boldsymbol{\beta}}_{\rm OLS} = \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}_{\rm OLS} = \mathbf{S}_{\mathbf{Y}|\mathbf{X}}$. From Cook et al. (2010), we have the maximum likelihood estimators for the envelope model as

$$\begin{aligned}
\widehat{\boldsymbol{\Gamma}}_{\rm Env} &= \arg\min_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}|\right\} \\
\widehat{\boldsymbol{\beta}}_{\rm Env} &= \widehat{\boldsymbol{\Gamma}}_{\rm Env}\mathbf{S}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}^T\mathbf{Y},\mathbf{X}}\mathbf{S}_{\mathbf{X}}^{-1} = \mathbf{P}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}}\widehat{\boldsymbol{\beta}}_{\rm OLS} \\
\widehat{\boldsymbol{\Sigma}}_{\rm Env} &= \mathbf{P}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{P}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}} + \mathbf{Q}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}}\mathbf{S}_{\mathbf{Y}}\mathbf{Q}_{\widehat{\boldsymbol{\Gamma}}_{\rm Env}}.
\end{aligned}$$

A.2 Reduced-rank regression (proof of Lemma 1)

Following Anderson (1999) equation (2.13), we let $\widehat{\mathbf{L}} \in \mathbb{R}^{p\times d}$ denote $\mathbf{S}_{\mathbf{X}}^{-1/2}[\mathbf{v}_1, \ldots, \mathbf{v}_d]$, where $\mathbf{v}_i$ is the i-th eigenvector of $\mathbf{S}_{\mathbf{X}}^{-1/2}\mathbf{S}_{\mathbf{X}\circ\mathbf{Y}}\mathbf{S}_{\mathbf{X}}^{-1/2}$. Then the estimators can be written as $\widehat{\boldsymbol{\alpha}}_{\rm RR} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\rm RR}\overline{\mathbf{X}}$, $\widehat{\boldsymbol{\beta}}_{\rm RR} = \mathbf{S}_{\mathbf{YX}}\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T$ and $\widehat{\boldsymbol{\Sigma}}_{\rm RR} = \mathbf{S}_{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\rm RR}\mathbf{S}_{\mathbf{XY}}$. We then use the sample canonical

correlation matrix notion to get the results in Lemma 1: $\mathbf{S}_{\mathbf{X}}^{-1/2}\mathbf{S}_{\mathbf{X}\circ\mathbf{Y}}\mathbf{S}_{\mathbf{X}}^{-1/2} = \mathbf{C}_{\mathbf{XY}}\mathbf{C}_{\mathbf{YX}}$ and

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\mathrm{RR}} &= \mathbf{S}_{\mathbf{YX}}\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T = \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{X}}^{-1/2}\mathbf{P}_{\mathbf{C}_{\mathbf{XY}}^{(d)}}\mathbf{S}_{\mathbf{X}}^{-1/2} \\ &= \mathbf{S}_{\mathbf{Y}}^{1/2}\mathbf{C}_{\mathbf{YX}}\mathbf{P}_{\mathbf{C}_{\mathbf{XY}}^{(d)}}\mathbf{S}_{\mathbf{X}}^{-1/2} = \mathbf{S}_{\mathbf{Y}}^{1/2}\mathbf{C}_{\mathbf{YX}}^{(d)}\mathbf{S}_{\mathbf{X}}^{-1/2}.\end{aligned}$$

A.3 Reduced-rank envelope regression

A.3.1 Proof of Lemma 2

Estimation for the envelope model is facilitated by the following consideration which is straightforward from (2.4).

$$\boldsymbol{\Gamma}^T\mathbf{Y}_i = \boldsymbol{\Gamma}^T\boldsymbol{\alpha} + \boldsymbol{\eta}\mathbf{B}\mathbf{X}_i + \boldsymbol{\Gamma}^T\boldsymbol{\epsilon}_i, \tag{A.1}$$

$$\boldsymbol{\Gamma}_0^T\mathbf{Y}_i = \boldsymbol{\Gamma}_0^T\boldsymbol{\alpha} + \boldsymbol{\Gamma}_0^T\boldsymbol{\epsilon}_i, \tag{A.2}$$

where $\boldsymbol{\Gamma}^T\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Omega})$, $\boldsymbol{\Gamma}_0^T\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Omega}_0)$, $\boldsymbol{\Gamma}^T\boldsymbol{\epsilon} \perp\!\!\!\perp \boldsymbol{\Gamma}_0^T\boldsymbol{\epsilon}$.

The maximum likelihood estimator of $\boldsymbol{\alpha}$ is $\widehat{\boldsymbol{\alpha}}_{\mathrm{RE}} = \overline{\mathbf{Y}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}}\overline{\mathbf{X}}$ and effectively we could use centered response $\mathbf{Y}_{ci} := \mathbf{Y}_i - \overline{\mathbf{Y}}$ and centered predictors $\mathbf{X}_{ci} = \mathbf{X}_i - \overline{\mathbf{X}}$ to omit the analysis on $\boldsymbol{\alpha}$ and $\widehat{\boldsymbol{\alpha}}_{\mathrm{RE}}$. Then the partially maximized log-likelihood with known dimensions u and d can be decomposed into the following two additive parts since $\boldsymbol{\Gamma}^T\boldsymbol{\epsilon}$ is independent of $\boldsymbol{\Gamma}_0^T\boldsymbol{\epsilon}$.

$$L_n(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \mathbf{B}, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}|d, u) \simeq L_{1,n}(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \mathbf{B}, \boldsymbol{\Omega}|d, u) + L_{2,n}(\boldsymbol{\Gamma}_0, \boldsymbol{\Omega}_0|u) \tag{A.3}$$

where $L_{1,n}(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \mathbf{B}, \boldsymbol{\Omega}|d, u)$ corresponds to the likelihood from (A.1) and is given by

$$-\frac{n}{2}\left\{\log|\boldsymbol{\Omega}| + \mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\Gamma}^T\mathbf{Y}_{ci} - \boldsymbol{\eta}\mathbf{B}\mathbf{X}_{ci})(\boldsymbol{\Gamma}^T\mathbf{Y}_{ci} - \boldsymbol{\eta}\mathbf{B}\mathbf{X}_{ci})^T\right]\right\}, \tag{A.4}$$

and $L_{2,n}(\boldsymbol{\Gamma}_0, \boldsymbol{\Omega}_0|u)$ corresponds to the likelihood from (A.2) and is equal to

$$-\frac{n}{2}\left\{\log|\boldsymbol{\Omega}_0| + \mathrm{trace}\left[\boldsymbol{\Omega}_0^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Gamma}_0^T\mathbf{Y}_{ci}\mathbf{Y}_{ci}^T\boldsymbol{\Gamma}_0\right]\right\}.$$

It follows that $L_{2,n}$ is maximized over $\boldsymbol{\Omega}_0$ by $\sum_{i=1}^{n}\boldsymbol{\Gamma}_0^T\mathbf{Y}_{ci}\mathbf{Y}_{ci}^T\boldsymbol{\Gamma}_0/n = \boldsymbol{\Gamma}_0^T\mathbf{S}_{\mathbf{Y}}\boldsymbol{\Gamma}_0$. Substituting back, we find the following partially maximized form for $L_{2,n}$:

$$L_{2,n}(\boldsymbol{\Gamma}_0|u) \simeq -(n/2)\log|\boldsymbol{\Gamma}_0^T\mathbf{S}_{\mathbf{Y}}\boldsymbol{\Gamma}_0|. \tag{A.5}$$

Holding $\mathbf{\Gamma}$ fixed, the log-likelihood $L_{1,n}$ is same as the log-likelihood for reduced rank regression of $\mathbf{\Gamma}^T\mathbf{Y}$ on $\mathbf{X}$. Therefore, by replacing $r \to u$, $\mathbf{Y} \to \mathbf{\Gamma}^T\mathbf{Y}$, $\mathbf{A} \to \boldsymbol{\eta}$, $\mathbf{B} \to \mathbf{B}$ and $\mathbf{\Sigma} \to \mathbf{\Omega}$ in (2.2) and in Lemma 1, we partially maximize $L_{1,n}(\mathbf{\Gamma}, \boldsymbol{\eta}, \mathbf{B}, \mathbf{\Omega}|d, u)$ over $\boldsymbol{\eta}$, $\mathbf{B}$ and $\mathbf{\Omega}$ and obtain the maximum likelihood estimators as

$$\begin{aligned} \widehat{\boldsymbol{\eta}}_{\mathbf{\Gamma}}\widehat{\mathbf{B}}_{\mathbf{\Gamma}} &= \mathbf{S}^{1/2}_{\mathbf{\Gamma}^T\mathbf{Y}}\mathbf{C}^{(d)}_{\mathbf{\Gamma}^T\mathbf{Y},\mathbf{X}}\mathbf{S}^{-1/2}_{\mathbf{X}} \\ \widehat{\mathbf{\Omega}}_{\mathbf{\Gamma}} &= \mathbf{S}^{1/2}_{\mathbf{\Gamma}^T\mathbf{Y}}\left\{\mathbf{I}_u - \mathbf{C}^{(d)}_{\mathbf{\Gamma}^T\mathbf{Y},\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X},\mathbf{\Gamma}^T\mathbf{Y}}\right\}\mathbf{S}^{1/2}_{\mathbf{\Gamma}^T\mathbf{Y}}, \end{aligned}$$

from which Lemma 2 follows.

A.3.2 Proof of Proposition 2

The log-likelihood function in (A.3) after partial maximization becomes

$$L_n(\mathbf{\Gamma}|d, u) \simeq -(n/2)\left\{\log|\mathbf{\Gamma}_0^T\mathbf{S}_{\mathbf{Y}}\mathbf{\Gamma}_0| + \log|\widehat{\mathbf{\Omega}}_{\mathbf{\Gamma}}|\right\}, \tag{A.6}$$

which lead us to the objective function $\mathrm{F}_n(\mathbf{G}|d, u) := (-2/n)L_n(\mathbf{G}|d, u)$ for numerical optimization over $\mathrm{span}(\mathbf{G}) \in \mathcal{G}_{r,u}$. We next simplify the expression of $\log|\widehat{\mathbf{\Omega}}_{\mathbf{G}}|$ as

$$\begin{aligned} \log|\widehat{\mathbf{\Omega}}_{\mathbf{G}}| &= \log|\mathbf{S}^{1/2}_{\mathbf{G}^T\mathbf{Y}}\left\{\mathbf{I}_u - \mathbf{C}^{(d)}_{\mathbf{G}^T\mathbf{Y},\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X},\mathbf{G}^T\mathbf{Y}}\right\}\mathbf{S}^{1/2}_{\mathbf{G}^T\mathbf{Y}}| \\ &= 2\cdot\log|\mathbf{S}^{1/2}_{\mathbf{G}^T\mathbf{Y}}| + \log|\mathbf{I}_u - \mathbf{C}^{(d)}_{\mathbf{G}^T\mathbf{Y},\mathbf{X}}\mathbf{C}^{(d)}_{\mathbf{X},\mathbf{G}^T\mathbf{Y}}| \\ &= \log|\mathbf{S}_{\mathbf{G}^T\mathbf{Y}}| + \log|\mathbf{I}_u - \mathbf{S}^{(d)}_{\mathbf{Z}_{\mathbf{G}}\circ\mathbf{X}}|, \end{aligned}$$

where $\mathbf{S}_{\mathbf{G}^T\mathbf{Y}} = \mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G}^T$ and $\mathbf{Z}_{\mathbf{G}} = (\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{Y}$ is the standardized random vector in $\mathbb{R}^u$. Equation (3.2) is then obtained by noticing $\log|\mathbf{G}_0\mathbf{S}_{\mathbf{Y}}\mathbf{G}_0^T| = \log|\mathbf{S}_{\mathbf{Y}}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}|$ in the objective function (A.6).

We next prove the equality in (3.3). The first term in (3.3) can be re-expressed as $\log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G}| = \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G}| + \log|\mathbf{S}_{\mathbf{Z}_{\mathbf{G}}|\mathbf{X}}|$ according to the following.

$$\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G} = \mathbf{G}^T\mathbf{S}_{\mathbf{Y}\mathbf{X}}\mathbf{S}_{\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X}\mathbf{Y}}\mathbf{G} = \mathbf{S}_{\mathbf{G}^T\mathbf{Y},\mathbf{X}}\mathbf{S}_{\mathbf{X}}^{-1}\mathbf{S}_{\mathbf{X},\mathbf{G}^T\mathbf{Y}} = \mathbf{S}^{1/2}_{\mathbf{G}^T\mathbf{Y}}\mathbf{S}_{\mathbf{Z}_{\mathbf{G}}|\mathbf{X}}\mathbf{S}^{1/2}_{\mathbf{G}^T\mathbf{Y}}.$$

The objective function in (3.3) now become

$$\log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G}| + \log|\mathbf{S}_{\mathbf{Z}_{\mathbf{G}}|\mathbf{X}}| + \log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}^{-1}\mathbf{G}| + \sum_{i=d+1}^{u}\log[\widehat{\omega}_i(\mathbf{G})],$$

where $\widehat{\omega}_i(\mathbf{G})$ is the i-th eigenvalue of $\mathbf{S}_{\mathbf{Z_G}\circ\mathbf{X}}$. The equality connecting (3.2) and (3.3) is proved by noticing that $\mathbf{S}_{\mathbf{Z_G}|\mathbf{X}} = \mathbf{S}_{\mathbf{Z_G}} - \mathbf{S}_{\mathbf{Z_G}\circ\mathbf{X}} = \mathbf{I}_u - \mathbf{S}_{\mathbf{Z_G}\circ\mathbf{X}}$ and that the log-determinant of a positive definite matrix is the sum of the logarithms of its eigenvalues.

A.3.3 Proof of Proposition 4

The proof follows trivially by combining the results in Lemma 2 and Proposition 2.

B Proof for Proposition 3

Recall that in (3.3), $\widehat{\omega}_i(\mathbf{G})$ is the i-th eigenvalues of the following matrix.

$$\begin{aligned}\mathbf{S}^{-1}_{\mathbf{Z_G}|\mathbf{X}} &= (\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2}(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}}\mathbf{G})(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2} \\ &= \mathbf{I}_u + (\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2}(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}\circ\mathbf{X}}\mathbf{G})(\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G})^{-1/2},\end{aligned}$$

which relies on the two sample covariance matrices: $\mathbf{S}_{\mathbf{Y}|\mathbf{X}}$ and $\mathbf{S}_{\mathbf{Y}\circ\mathbf{X}}$. These two matrices are both positive semi-definite and converge to $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_{\mathbf{Y}\circ\mathbf{X}} = \mathbf{\Sigma}_{\mathbf{YX}}\mathbf{\Sigma}^{-1}_{\mathbf{X}}\mathbf{\Sigma}_{\mathbf{XY}}$ with probability one as $n \to \infty$. Since $\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{Y}\circ\mathbf{X}}) = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{YX}}\mathbf{\Sigma}^{-1}_{\mathbf{X}}\mathbf{\Sigma}_{\mathbf{XY}}) = \mathrm{rank}(\boldsymbol{\beta}\mathbf{\Sigma}_{\mathbf{X}}\boldsymbol{\beta}^T) = d$, the last $(u-d)$ eigenvalues $\widehat{\omega}_j(\mathbf{G})$, $j = d+1, \ldots, u$, will equal to one with probability one as $n \to \infty$ for any value of $\mathbf{G}$. Therefore, as $n \to \infty$,

$$\sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{ \sum_{i=d+1}^{u} \log[\widehat{\omega}_i(\mathbf{G})] \right\} \xrightarrow{p} 0. \tag{B.1}$$

We next show that $\log|\mathbf{G}^T\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{G}|$ converges in probability to $\log|\mathbf{G}^T\mathbf{\Sigma}^{-1}_{\mathbf{Y}}\mathbf{G}|$ uniformly in $\mathbf{G}$ by the following argument.

$$\begin{aligned}\delta(\mathbf{G}) &:= \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|\mathbf{G}^T\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{G}| - \log|\mathbf{G}^T\mathbf{\Sigma}^{-1}_{\mathbf{Y}}\mathbf{G}|\right\} \\ &= \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|(\mathbf{G}^T\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{G})(\mathbf{G}^T\mathbf{\Sigma}^{-1}_{\mathbf{Y}}\mathbf{G})^{-1}|\right\} \\ &= \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{G}(\mathbf{G}^T\mathbf{\Sigma}^{-1}_{\mathbf{Y}}\mathbf{G})^{-1}\mathbf{G}^T|_0\right\} \\ &= \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|\mathbf{\Sigma}^{1/2}_{\mathbf{Y}}\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{\Sigma}^{1/2}_{\mathbf{Y}} \cdot \mathbf{\Sigma}^{-1/2}\mathbf{G}(\mathbf{G}^T\mathbf{\Sigma}^{-1}_{\mathbf{Y}}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1/2}_{\mathbf{Y}}|_0\right\} \\ &= \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{\log|\mathbf{\Sigma}^{1/2}_{\mathbf{Y}}\mathbf{S}^{-1}_{\mathbf{Y}}\mathbf{\Sigma}^{1/2}_{\mathbf{Y}}\mathbf{P}_{\mathbf{\Sigma}^{-1/2}_{\mathbf{Y}}\mathbf{G}}|_0\right\},\end{aligned}$$

where we use $|\cdot|_0$ to denote the product of the non-zero eigenvalues of a positive semi-definite matrix. We then can derive that

$$\delta(\mathbf{G}) = \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{ \log |\mathbf{P}_{\mathbf{\Sigma}_\mathbf{Y}^{-1/2}\mathbf{G}} \mathbf{\Sigma}_\mathbf{Y}^{1/2} \mathbf{S}_\mathbf{Y}^{-1} \mathbf{\Sigma}_\mathbf{Y}^{1/2} \mathbf{P}_{\mathbf{\Sigma}_\mathbf{Y}^{-1/2}\mathbf{G}}|_0 \right\}, \tag{B.2}$$

where $\mathbf{\Sigma}_\mathbf{Y}^{1/2}\mathbf{S}_\mathbf{Y}^{-1}\mathbf{\Sigma}_\mathbf{Y}^{1/2}$ was projected onto an u-dimensional subspace $\mathrm{span}(\mathbf{\Sigma}_\mathbf{Y}^{-1/2}\mathbf{G})$. The quantity within $|\cdot|_0$ then has at most u nonzero eigenvalues. Because the projection matrix can not inflate the eigenvalues,

$$\delta(\mathbf{G}) \leq \sup_{\mathbf{G}\in\mathcal{G}_{r,u}} \left\{ \log |\mathbf{\Sigma}_\mathbf{Y}^{1/2} \mathbf{S}_\mathbf{Y}^{-1} \mathbf{\Sigma}_\mathbf{Y}^{1/2}|_0 \right\}, \tag{B.3}$$

which converges to zero in probability. Similarly, we can show that $\log|\mathbf{G}^T\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{G}|$ converges in probability to $\log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}|$ uniformly in $\mathbf{G}$. Hence we have proved that the objective function $\mathrm{F}_n(\mathbf{G}|d,u)$ in (3.3) converges in probability to $\mathrm{F}(\mathbf{G}|u)$ uniformly in $\mathbf{G}$. The rest of the proof is similar to the proof of Proposition 4.2 in Cook et al. (2013) that

$$\begin{aligned}
\log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}| + \log|\mathbf{G}^T\mathbf{\Sigma}_\mathbf{Y}^{-1}\mathbf{G}| &= \log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}| + \log|\mathbf{G}_0^T\mathbf{\Sigma}_\mathbf{Y}\mathbf{G}_0| \\
&= \log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}| + \log|\mathbf{G}_0^T(\mathbf{\Sigma} + \boldsymbol{\beta}\mathbf{\Sigma}_\mathbf{X}\boldsymbol{\beta}^T)\mathbf{G}_0| \\
&\geq \log|\mathbf{G}^T\mathbf{\Sigma}\mathbf{G}| + \log|\mathbf{G}_0^T\mathbf{\Sigma}\mathbf{G}_0| \\
&\geq \log|\mathbf{\Sigma}|,
\end{aligned}$$

where the first inequality achieves its lower bound if $\mathrm{span}(\boldsymbol{\beta}) \subseteq \mathrm{span}(\mathbf{G})$; and the second inequality achieves its lower bound if $\mathrm{span}(\mathbf{G})$ is a reducing subspace of $\mathbf{\Sigma}$. The uniqueness of the minimizer $\mathrm{span}(\widehat{\mathbf{\Gamma}}) = \mathrm{span}(\arg\min_\mathbf{G} \mathrm{F}(\mathbf{G}|u))$ is guaranteed by the uniqueness of the envelope, which has dimension u.

C Proof for Proposition 6

For notation convenience, we define two covariance matrices $\mathbf{M}_B := \mathbf{B}^T(\mathbf{B}\mathbf{\Sigma}_\mathbf{X}\mathbf{B}^T)^{-1}\mathbf{B} \leq \mathbf{\Sigma}_\mathbf{X}^{-1}$ and $\mathbf{M}_A := \mathbf{A}(\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T \leq \mathbf{\Sigma}$. For any full row rank transformation $\mathbf{O} \in \mathbb{R}^{d\times q}$ we could replace $\mathbf{A}$ by $\mathbf{AO}$ and replace $\mathbf{B}$ by $\mathbf{OB}$ without changing the value of $\mathbf{M}_A$ or $\mathbf{M}_B$. Also the projection matrices $\mathbf{P}_{\mathbf{A}(\mathbf{\Sigma}^{-1})} = \mathbf{M}_A\mathbf{\Sigma}^{-1}$ and $\mathbf{P}_{\mathbf{B}^T(\mathbf{\Sigma}_\mathbf{X})} = \mathbf{M}_B\mathbf{\Sigma}_\mathbf{X}$.

C.1 Obtaining equation (4.3)

This result can be found in Anderson (1999) using canonical variables. We replicate the computation in our framework with details. Recall that the Fisher information is

$$\mathbf{J_h} = \begin{pmatrix} \mathbf{J}_{\boldsymbol{\beta}} & 0 \\ 0 & \mathbf{J}_{\boldsymbol{\Sigma}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Sigma}^{-1} & 0 \\ 0 & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{E}_r \end{pmatrix}, \tag{C.1}$$

where $\mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}) = \mathbf{J}_{\boldsymbol{\beta}}^{-1} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}$.

By noticing $\mathbf{h}_1 = \mathrm{vec}(\boldsymbol{\beta}) = \mathrm{vec}(\mathbf{AB}) = (\mathbf{B}^T \otimes \mathbf{I}_r)\mathrm{vec}(\mathbf{A}) = (\mathbf{I}_p \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, we have

$$\mathbf{H} = \begin{pmatrix} \mathbf{B}^T \otimes \mathbf{I}_r & \mathbf{I}_p \otimes \mathbf{A} & 0 \\ 0 & 0 & \mathbf{I}_{r(r+1)/2} \end{pmatrix} := \begin{pmatrix} \mathbf{H}_1 & 0 \\ 0 & \mathbf{I}_{r(r+1)/2} \end{pmatrix}. \tag{C.2}$$

Because of the similar block-diagonal structure in $\mathbf{J_h} = \mathrm{diag}(\mathbf{J}_{\boldsymbol{\beta}}, \mathbf{J}_{\boldsymbol{\Sigma}})$, we can get

$$\mathbf{H}(\mathbf{H}^T\mathbf{J_h}\mathbf{H})^{\dagger}\mathbf{H}^T = \begin{pmatrix} \mathbf{H}_1(\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1)^{\dagger}\mathbf{H}_1^T & 0 \\ 0 & \mathbf{J}_{\boldsymbol{\Sigma}}^{-1} \end{pmatrix},$$

which means that $\boldsymbol{\beta} = \mathbf{AB}$ and $\boldsymbol{\Sigma}$ are orthogonal parameters in reduced-rank regression and the asymptotic covariance for $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})$ is $\mathbf{H}_1(\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1)^{\dagger}\mathbf{H}_1^T$. Because $\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1$ is not full rank under the reduced rank regression model, we can not use the block-matrix inversion formula. However, notice that asymptotic covariance $\mathbf{H}_1(\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1)^{\dagger}\mathbf{H}_1^T$ depends only on the column space of $\mathbf{H}_1$, we thus could use any full row rank matrix $\mathbf{T}_1$ to get

$$\mathbf{H}_1(\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1)^{\dagger}\mathbf{H}_1^T = \mathbf{H}_1\mathbf{T}_1(\mathbf{T}_1^T\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1\mathbf{T}_1)^{\dagger}\mathbf{T}_1^T\mathbf{H}_1. \tag{C.3}$$

More specifically, we have each part

$$\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1 = \begin{pmatrix} \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T \otimes \boldsymbol{\Sigma}^{-1} & \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{A} \\ \boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T \otimes \mathbf{A}^T\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Sigma}_{\mathbf{X}} \otimes \mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A} \end{pmatrix} \tag{C.4}$$

$$\mathbf{T}_1 = \begin{pmatrix} \mathbf{I}_{rd} & -(\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T)^{-1}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}} \otimes \mathbf{A} \\ 0 & \mathbf{I}_{pd} \end{pmatrix}$$

$$\mathbf{H}_1\mathbf{T}_1 = \begin{pmatrix} \mathbf{B}^T \otimes \mathbf{I}_r & (\mathbf{I}_p - \mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}}) \otimes \mathbf{A} \end{pmatrix}. \tag{C.5}$$

where we have used $\mathbf{M}_B = \mathbf{B}^T(\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T)^{-1}\mathbf{B}$ for notation convenience. Then,

$$\mathbf{T}_1^T\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1\mathbf{T}_1 \;=\; \begin{pmatrix} \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\otimes\boldsymbol{\Sigma}^{-1} & 0 \\ 0 & (\boldsymbol{\Sigma}_{\mathbf{X}}-\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})\otimes\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A} \end{pmatrix}.$$

To get the Moore-Penrose inverse of $\mathbf{T}_1^T\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1\mathbf{T}_1$, we first notice that it has rank $(p+r)d-d^2$ and the only non-invertable part is $(\boldsymbol{\Sigma}_{\mathbf{X}}-\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})$ which causes rank deficiency of d^2. The Moore-Penrose inverse of $\boldsymbol{\Sigma}_{\mathbf{X}}-\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}}$ is obtained as follows by noticing $\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{M}_B=\mathbf{M}_B$.

$$(\boldsymbol{\Sigma}_{\mathbf{X}}-\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})^{\dagger}=\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B. \tag{C.6}$$

Therefore,

$$\left(\mathbf{T}_1^T\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1\mathbf{T}_1\right)^{\dagger}=\begin{pmatrix} (\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T)^{-1}\otimes\boldsymbol{\Sigma} & 0 \\ 0 & (\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)\otimes(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1} \end{pmatrix}. \tag{C.7}$$

The asymptotic covariance $\mathrm{avar}(\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})]=\mathbf{H}_1\mathbf{T}_1(\mathbf{T}_1^T\mathbf{H}_1^T\mathbf{J}_{\boldsymbol{\beta}}\mathbf{H}_1\mathbf{T}_1)^{\dagger}\mathbf{T}_1^T\mathbf{H}_1$ is computed with (C.5),

$$\begin{aligned}
\mathrm{avar}(\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})] &= \mathbf{M}_B\otimes\boldsymbol{\Sigma} \\
&+ (\mathbf{I}_p-\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)(\mathbf{I}_p-\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})\otimes\mathbf{M}_A \\
&= \mathbf{M}_B\otimes\boldsymbol{\Sigma}+(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)\otimes\mathbf{M}_A \\
&= [\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)]\otimes\boldsymbol{\Sigma}+(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)\otimes\mathbf{M}_A \\
&= \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Sigma}-(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)\otimes(\boldsymbol{\Sigma}-\mathbf{M}_A),
\end{aligned} \tag{C.8}$$

then the equation (4.3) is derived from the following arguments.

$$\begin{aligned}
\mathrm{avar}(\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})] &= \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Sigma}-(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}-\mathbf{M}_B)\otimes(\boldsymbol{\Sigma}-\mathbf{M}_A) \\
&= \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Sigma}-\left[(\mathbf{I}_p-\mathbf{M}_B\boldsymbol{\Sigma}_{\mathbf{X}})\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\right]\otimes\left[(\mathbf{I}_r-\mathbf{M}_A\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}\right] \\
&= \boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Sigma}-\left[\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\right]\otimes\left[\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}\right] \\
&= (\mathbf{I}_{pr}-\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\otimes\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma})})\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\otimes\boldsymbol{\Sigma} \\
&= (\mathbf{I}_{pr}-\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\otimes\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma})})\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})].
\end{aligned}$$

C.2 Obtaining equation (4.4)

The Fisher information for $(\boldsymbol{\psi}_1^T, \boldsymbol{\psi}_2^T)^T = [\mathrm{vec}^T(\mathbf{A}), \mathrm{vec}^T(\mathbf{B})]^T$ is given in (C.4) as

$$\mathbf{H}_1^T \mathbf{J}_\beta \mathbf{H}_1 := \begin{pmatrix} \mathbf{J_A} & \mathbf{J_{AB}} \\ \mathbf{J_{BA}} & \mathbf{J_B} \end{pmatrix} = \begin{pmatrix} \mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T \otimes \boldsymbol{\Sigma}^{-1} & \mathbf{B}\boldsymbol{\Sigma}_\mathbf{X} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{A} \\ \boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T \otimes \mathbf{A}\boldsymbol{\Sigma}^{-1} & \boldsymbol{\Sigma}_\mathbf{X} \otimes \mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A} \end{pmatrix}. \tag{C.9}$$

If we known $\mathbf{A}$, then we could cross the first row and the first column, and hence

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\mathbf{B}}_\mathbf{A})] = \mathbf{J}_\mathbf{B}^{-1} = \boldsymbol{\Sigma}_\mathbf{X}^{-1} \otimes (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}. \tag{C.10}$$

Similarly,

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\mathbf{A}}_\mathbf{B})] = \mathbf{J}_\mathbf{A}^{-1} = (\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T)^{-1} \otimes \boldsymbol{\Sigma}. \tag{C.11}$$

Then by using the fact that $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_\mathbf{A}) = (\mathbf{I}_p \otimes \mathbf{A})\mathrm{vec}(\widehat{\mathbf{B}}_\mathbf{A})$ and that $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{I}_r)\mathrm{vec}(\widehat{\mathbf{A}}_\mathbf{B})$, we have

$$\begin{aligned} \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_\mathbf{A})] &= \boldsymbol{\Sigma}_\mathbf{X}^{-1} \otimes \mathbf{M}_A \\ \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_\mathbf{B})] &= \mathbf{M}_B \otimes \boldsymbol{\Sigma}. \end{aligned}$$

By noticing $\mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})} = \mathbf{M}_A\boldsymbol{\Sigma}^{-1}$ and $\mathbf{P}_{\mathbf{B}^T(\boldsymbol{\Sigma}_\mathbf{X})} = \mathbf{M}_B\boldsymbol{\Sigma}_\mathbf{X}$, we have

$$\begin{aligned} \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_\mathbf{A}\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_\mathbf{X})})] &= [\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_\mathbf{X})} \otimes \mathbf{I}_r](\boldsymbol{\Sigma}_\mathbf{X}^{-1} \otimes \mathbf{M}_A)[\mathbf{Q}^T_{\mathbf{B}^T(\boldsymbol{\Sigma}_\mathbf{X})} \otimes \mathbf{I}_r] \\ &= \left[(\mathbf{I}_p - \mathbf{M}_B\boldsymbol{\Sigma}_\mathbf{X})\boldsymbol{\Sigma}_\mathbf{X}^{-1}(\mathbf{I}_p - \boldsymbol{\Sigma}_\mathbf{X}^T\mathbf{M}_B)\right] \otimes \mathbf{M}_A \\ &= (\boldsymbol{\Sigma}_\mathbf{X}^{-1} - \mathbf{M}_B) \otimes \mathbf{M}_A \\ \mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\widehat{\boldsymbol{\beta}}_\mathbf{B})] &= [\mathbf{I}_p \otimes \mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}](\mathbf{M}_B \otimes \boldsymbol{\Sigma})[\mathbf{I}_p \otimes \mathbf{Q}^T_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}] \\ &= \mathbf{M}_B \otimes (\mathbf{I}_r - \mathbf{M}_A\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\mathbf{I}_r - \boldsymbol{\Sigma}^{-1}\mathbf{M}_A) \\ &= \mathbf{M}_B \otimes (\boldsymbol{\Sigma} - \mathbf{M}_A). \end{aligned}$$

The proof of Proposition 6 is then completed by compare the above quantities with (C.8).

D Proof for Proposition 7

The role of $\boldsymbol{\eta}$ is analogous to $\mathbf{A}$ given $\boldsymbol{\Gamma}$, thus we define $\mathbf{M}_\eta := \boldsymbol{\eta}(\boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^T \leq \boldsymbol{\Omega}$. Note that the projection matrices $\mathbf{P}_{\boldsymbol{\eta}(\boldsymbol{\Omega}^{-1})} = \mathbf{M}_\eta\boldsymbol{\Omega}^{-1}$.

D.1 Explicit expression for $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$

By noticing $\mathbf{h}_1 = \mathrm{vec}(\boldsymbol{\beta}) = \mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{B}) = (\boldsymbol{\eta}^T\mathbf{B}^T \otimes \mathbf{I}_r)\mathrm{vec}(\boldsymbol{\Gamma}) = (\mathbf{B}^T \otimes \boldsymbol{\Gamma})\mathrm{vec}(\boldsymbol{\eta}) = (\mathbf{I}_p \otimes \boldsymbol{\Gamma}\boldsymbol{\eta})\mathrm{vec}(\mathbf{B})$, we have

$$\mathbf{R} = \begin{pmatrix} \mathbf{B}^T\boldsymbol{\eta}^T \otimes \mathbf{I}_r & \mathbf{B}^T \otimes \boldsymbol{\Gamma} & \mathbf{I}_p \otimes \boldsymbol{\Gamma}\boldsymbol{\eta} & 0 & 0 \\ 2\mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_r - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) & 0 & 0 & \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_u & \mathbf{C}_r(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\mathbf{E}_{r-u} \end{pmatrix}. \tag{D.1}$$

The asymptotic covariance $\mathrm{avar}(\sqrt{n}\mathbf{h}(\widehat{\boldsymbol{\phi}})) = \mathbf{R}(\mathbf{R}^T\mathbf{J}_\mathbf{h}\mathbf{R})^\dagger\mathbf{R}^T = \widetilde{\mathbf{R}}(\widetilde{\mathbf{R}}^T\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}})^\dagger\widetilde{\mathbf{R}}^T$ for any $\widetilde{\mathbf{R}}$ such that $\mathbf{R} = \widetilde{\mathbf{R}}\mathbf{T}$ for a full row rank matrix $\mathbf{T}$. We choose $\widetilde{\mathbf{R}}$ to make $\widetilde{\mathbf{R}}^T\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}$ block-diagonal as follows.

$$\widetilde{\mathbf{R}} = \begin{pmatrix} \mathbf{B}^T\boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}_0 & \mathbf{B}^T \otimes \boldsymbol{\Gamma} & (\mathbf{I}_p - \mathbf{M}_B\boldsymbol{\Sigma}_\mathbf{X}) \otimes \boldsymbol{\Gamma}\boldsymbol{\eta} & 0 & 0 \\ 2\mathbf{C}_r(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0) & 0 & 0 & \mathbf{C}_r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_u & \mathbf{C}_r(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\mathbf{E}_{r-u} \end{pmatrix}, \tag{D.2}$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{I}_u \otimes \boldsymbol{\Gamma}_0^T & 0 & 0 & 0 & 0 \\ \boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}^T & \mathbf{I}_{ud} & (\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T)^{-1}\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X} \otimes \boldsymbol{\eta} & 0 & 0 \\ 0 & 0 & \mathbf{I}_{pd} & 0 & 0 \\ 2\mathbf{C}_u(\boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}^T) & 0 & 0 & \mathbf{I}_{\frac{1}{2}r(r+1)} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I}_{\frac{1}{2}(r-u)(r-u+1)} \end{pmatrix}. \tag{D.3}$$

Next, we calculate $\widetilde{\mathbf{R}}^T\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}$ and verify that it is block-diagonal. We decompose $\widetilde{\mathbf{G}}$ by it 2×5 blocks as $\widetilde{\mathbf{R}} := (\widetilde{\mathbf{G}}_1, \widetilde{\mathbf{G}}_2, \widetilde{\mathbf{G}}_3, \widetilde{\mathbf{G}}_4, \widetilde{\mathbf{G}}_5)$. We first calculate $\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}$ and write down the 2×5 blocks by column:

$$\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}_1 = \begin{pmatrix} \boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T\boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1} \\ \mathbf{E}_r^T(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1} - \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}_0) \end{pmatrix}, \tag{D.4}$$

$$\mathbf{J}_\mathbf{h}[\widetilde{\mathbf{R}}_2, \widetilde{\mathbf{R}}_3] = \begin{pmatrix} \boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1} & (\boldsymbol{\Sigma}_\mathbf{X} - \boldsymbol{\Sigma}_\mathbf{X}\mathbf{M}_B\boldsymbol{\Sigma}_\mathbf{X}) \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\boldsymbol{\eta} \\ 0 & 0 \end{pmatrix}, \tag{D.5}$$

$$\mathbf{J}_\mathbf{h}[\widetilde{\mathbf{G}}_4, \widetilde{\mathbf{G}}_5] = \begin{pmatrix} 0 & 0 \\ \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1})\mathbf{E}_u & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1})\mathbf{E}_{r-u} \end{pmatrix}. \tag{D.6}$$

Then $\widetilde{\mathbf{R}}^T\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}$ equals to a block-diagonal matrix with five blocks: $\widetilde{\mathbf{R}}_i^T\mathbf{J}_\mathbf{h}\widetilde{\mathbf{R}}_i$, $i = 1, \ldots, 5$.

The explicit expressions are given as follows.

$$\begin{aligned}
\widetilde{\mathbf{R}}_1^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_1 &= \boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Omega}_0^{-1} \\
&\quad +2(\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}_0^T-\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T\mathbf{E}_r^T(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}-\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Gamma}_0) \\
&= \boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Omega}_0^{-1}+\boldsymbol{\Omega}\otimes\boldsymbol{\Omega}_0^{-1}-2\mathbf{I}_{u(r-u)}+\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}_0, \\
\widetilde{\mathbf{R}}_2^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_2 &= \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\otimes\boldsymbol{\Omega}^{-1}, \\
\widetilde{\mathbf{R}}_3^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_3 &= \boldsymbol{\Sigma}_{\mathbf{X}}\otimes\boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}, \\
\widetilde{\mathbf{R}}_4^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_4 &= \mathbf{E}_u^T(\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\mathbf{C}_r^T\cdot\frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1})\mathbf{E}_u \\
&= \frac{1}{2}\mathbf{E}_u^T(\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)(\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1})\mathbf{E}_u \\
&= \frac{1}{2}\mathbf{E}_u^T(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\mathbf{E}_u.
\end{aligned}$$

$$\begin{aligned}
\widetilde{\mathbf{R}}_5^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_5 &= \mathbf{E}_{r-u}^T(\boldsymbol{\Gamma}_0^T\otimes\boldsymbol{\Gamma}_0^T)\mathbf{C}_r^T\cdot\frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1})\mathbf{E}_{r-u} \\
&= \frac{1}{2}\mathbf{E}_{r-u}^T(\boldsymbol{\Omega}_0^{-1}\otimes\boldsymbol{\Omega}_0^{-1})\mathbf{E}_{r-u}
\end{aligned}$$

Then the asymptotic covariance is

$$\mathrm{avar}[\sqrt{n}\mathbf{h}(\widehat{\boldsymbol{\phi}})] = \sum_{i=1}^{5}\widetilde{\mathbf{R}}_i(\widetilde{\mathbf{R}}_i^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_i)^{\dagger}\widetilde{\mathbf{R}}_i^T$$

We are only interested in the asymptotic covariance of $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$, which is the upper left block of $\mathrm{avar}[\sqrt{n}\mathbf{h}(\widehat{\boldsymbol{\phi}})]$. And $\widetilde{\mathbf{R}}_4(\widetilde{\mathbf{R}}_4^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_4)^{\dagger}\widetilde{\mathbf{R}}_4^T$ and $\widetilde{\mathbf{R}}_5(\widetilde{\mathbf{R}}_5^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_5)^{\dagger}\widetilde{\mathbf{R}}_5^T$ have no contribution to that. So we will focus our attention on the upper left block of $\widetilde{\mathbf{R}}_i(\widetilde{\mathbf{R}}_i^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_i)^{\dagger}\widetilde{\mathbf{R}}_i^T$, $i=1,2,3$. The upper left block of $\widetilde{\mathbf{R}}_1(\widetilde{\mathbf{R}}_1^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_1)^{\dagger}\widetilde{\mathbf{R}}_1^T$ is

$$\begin{aligned}
&(\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Gamma}_0)(\widetilde{\mathbf{R}}_1^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_1)^{\dagger}(\boldsymbol{\eta}\mathbf{B}\otimes\boldsymbol{\Gamma}_0^T) \\
=\ &(\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Gamma}_0)(\boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Omega}_0^{-1}+\boldsymbol{\Omega}\otimes\boldsymbol{\Omega}_0^{-1}-2\mathbf{I}_{u(r-u)}+\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}_0)^{\dagger}(\boldsymbol{\eta}\mathbf{B}\otimes\boldsymbol{\Gamma}_0^T).
\end{aligned}$$

The upper left block of $\widetilde{\mathbf{R}}_2(\widetilde{\mathbf{R}}_2^T\mathbf{J_h}\widetilde{\mathbf{R}}_2)^\dagger\widetilde{\mathbf{R}}_2^T$ is

$$\begin{aligned}
& (\mathbf{B}^T\otimes\boldsymbol{\Gamma})(\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T\otimes\boldsymbol{\Omega}^{-1})^\dagger(\mathbf{B}\otimes\boldsymbol{\Gamma}^T)\\
= \;& (\mathbf{B}^T\otimes\boldsymbol{\Gamma})[(\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T)^{-1}\otimes\boldsymbol{\Omega}](\mathbf{B}\otimes\boldsymbol{\Gamma}^T)\\
= \;& \mathbf{M}_B\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T.
\end{aligned}$$

The upper left block of $\widetilde{\mathbf{R}}_3(\widetilde{\mathbf{R}}_3^T\mathbf{J_h}\widetilde{\mathbf{R}}_3)^\dagger\widetilde{\mathbf{R}}_3^T$ is

$$\begin{aligned}
& [(\mathbf{I}_p-\mathbf{M}_B\boldsymbol{\Sigma}_\mathbf{X})\otimes\boldsymbol{\Gamma}\boldsymbol{\eta}](\boldsymbol{\Sigma}_\mathbf{X}\otimes\boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta})^\dagger[(\mathbf{I}_p-\boldsymbol{\Sigma}_\mathbf{X}\mathbf{M}_B)\otimes\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T]\\
= \;& (\boldsymbol{\Sigma}_\mathbf{X}^{-1}-\mathbf{M}_B)\otimes\boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T,
\end{aligned}$$

where $\mathbf{M}_\eta=\boldsymbol{\eta}(\boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^T$.

Hence, the asymptotic covariance $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})]$ equals to

$$\begin{aligned}
& (\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Gamma}_0)(\boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_\mathbf{X}\mathbf{B}^T\boldsymbol{\eta}^T\otimes\boldsymbol{\Omega}_0^{-1}+\boldsymbol{\Omega}\otimes\boldsymbol{\Omega}_0^{-1}-2\mathbf{I}_{u(r-u)}+\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}_0)^\dagger(\boldsymbol{\eta}\mathbf{B}\otimes\boldsymbol{\Gamma}_0^T)\\
+ \;& \mathbf{M}_B\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T+(\boldsymbol{\Sigma}_\mathbf{X}^{-1}-\mathbf{M}_B)\otimes\boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T. \qquad \text{(D.7)}
\end{aligned}$$

D.2 Interpretation

The Fisher information matrix for $\widehat{\boldsymbol{\phi}}$ is simply $\mathbf{R}^T\mathbf{J_h}\mathbf{R}$:

$$\mathbf{R}^T\mathbf{J}\mathbf{R}:=\begin{pmatrix}
\mathbf{J}_{\boldsymbol{\Gamma}} & \mathbf{J}_{\boldsymbol{\Gamma}\boldsymbol{\eta}} & \mathbf{J}_{\boldsymbol{\Gamma}\mathbf{B}} & \mathbf{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}} & 0\\
\mathbf{J}_{\boldsymbol{\eta}\boldsymbol{\Gamma}} & \mathbf{J}_{\boldsymbol{\eta}} & \mathbf{J}_{\boldsymbol{\eta}\mathbf{B}} & 0 & 0\\
\mathbf{J}_{\mathbf{B}\boldsymbol{\Gamma}} & \mathbf{J}_{\mathbf{B}\boldsymbol{\eta}} & \mathbf{J}_{\mathbf{B}} & 0 & 0\\
\mathbf{J}_{\boldsymbol{\Omega}\boldsymbol{\Gamma}} & 0 & 0 & \mathbf{J}_{\boldsymbol{\Omega}} & 0\\
0 & 0 & 0 & 0 & \mathbf{J}_{\boldsymbol{\Omega}_0}
\end{pmatrix}. \qquad \text{(D.8)}$$

Each nonzero block is

$$
\begin{aligned}
\mathbf{J}_{\boldsymbol{\Gamma}} &= \boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\boldsymbol{\eta}^T \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma})\mathbf{K}_{ru} \\
&\quad + \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T - 2\mathbf{I}_u \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T \\
\mathbf{J}_{\boldsymbol{\eta}} &= \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T \otimes \boldsymbol{\Omega}^{-1} \\
\mathbf{J}_{\mathbf{B}} &= \boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta} \\
\mathbf{J}_{\boldsymbol{\Omega}} &= \frac{1}{2}\mathbf{E}_u^T(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})\mathbf{E}_u \\
\mathbf{J}_{\boldsymbol{\Omega}_0} &= \frac{1}{2}\mathbf{E}_{r-u}^T(\boldsymbol{\Omega}_0^{-1} \otimes \boldsymbol{\Omega}_0^{-1})\mathbf{E}_{r-u}. \\
\mathbf{J}_{\boldsymbol{\Gamma}\boldsymbol{\eta}} &= \boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1} \\
\mathbf{J}_{\boldsymbol{\Gamma}\mathbf{B}} &= \boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\boldsymbol{\eta} \\
\mathbf{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}} &= (\mathbf{I}_u \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1})\mathbf{E}_u \\
\mathbf{J}_{\boldsymbol{\eta}\mathbf{B}} &= \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Omega}^{-1}\boldsymbol{\eta}
\end{aligned}
$$

D.2.1 Asymptotic covariance when $\boldsymbol{\eta}$ and $\mathbf{B}$ are known

The asymptotic covariance for $\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta},\mathbf{B}})$ is

$$
\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta},\mathbf{B}})] = \left(\mathbf{J}_{\boldsymbol{\Gamma}} - \mathbf{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}}\mathbf{J}_{\boldsymbol{\Omega}}^{-1}\mathbf{J}_{\boldsymbol{\Omega}\boldsymbol{\Gamma}}\right)^{-1}.
$$

Follow Cook et al. (2010), we have

$$
\begin{aligned}
\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\eta},\mathbf{B}})] &= [\boldsymbol{\eta}\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T\boldsymbol{\eta}^T \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T - 2\mathbf{I}_u \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^T \\
&\quad + \boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T]^{\dagger},
\end{aligned}
$$

and by replacing $\boldsymbol{\eta}\mathbf{B} \to \boldsymbol{\eta}$ in Cook et al. (2010), it is easy to obtain the following results

$$
\mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})] = \left[\widetilde{\mathbf{R}}_1(\widetilde{\mathbf{R}}_1^T\mathbf{J}_{\mathbf{h}}\widetilde{\mathbf{R}}_1)^{\dagger}\widetilde{\mathbf{R}}_1^T\right]_{11},
$$

where $[]_{11}$ means the upper left block of a block-wise matrix. The above equality explains the contribution from the first column block of $\widetilde{\mathbf{R}}$, which is the first term in (D.7).

Therefore, (D.7) can be written as

$$\begin{aligned}\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})] &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})] \\ &\quad +\mathbf{M}_B \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + (\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} - \mathbf{M}_B) \otimes \boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T \qquad \text{(D.9)} \\ &= \mathrm{avar}[\sqrt{n}\mathrm{vec}(\mathbf{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\eta},\mathbf{B}})] + \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}})], \qquad \text{(D.10)}\end{aligned}$$

where the last equality follows from the asymptotic covariance of $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\mathrm{RR}})$ in (C.8) and from Lemma 1 that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}}$ is $\boldsymbol{\Gamma}$ times the reduced-rank regression estimator of regression $\boldsymbol{\Gamma}^T\mathbf{Y}$ on $\mathbf{X}$.

D.2.2 Asymptotic covariance when $\boldsymbol{\Gamma}$ and B are known

The asymptotic covariance for $\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma},\mathbf{B}})$ is

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma},\mathbf{B}})] = \mathbf{J}_{\boldsymbol{\eta}}^{-1} = (\mathbf{B}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{B}^T)^{-1} \otimes \boldsymbol{\Omega}. \qquad \text{(D.11)}$$

Notice that $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\mathbf{B}}) = \mathrm{vec}(\boldsymbol{\Gamma}\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma},\mathbf{B}}\mathbf{B}) = (\mathbf{B}^T \otimes \boldsymbol{\Gamma})\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\Gamma},\mathbf{B}})$, we have

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\mathbf{B}})] = \mathbf{M}_B \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T. \qquad \text{(D.12)}$$

D.2.3 Asymptotic covariance when $\boldsymbol{\Gamma}$ and $\boldsymbol{\eta}$ are known

The asymptotic covariance for $\mathrm{vec}(\widehat{\mathbf{B}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}})$ is

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\mathbf{B}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}})] = \mathbf{J}_{\mathbf{B}}^{-1} = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes (\boldsymbol{\eta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta})^{-1}. \qquad \text{(D.13)}$$

Notice that $\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}}) = \mathrm{vec}(\boldsymbol{\Gamma}\boldsymbol{\eta}\widehat{\mathbf{B}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}}) = (\mathbf{I}_p \otimes \boldsymbol{\Gamma}\boldsymbol{\eta})\mathrm{vec}(\widehat{\mathbf{B}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}})$, we have

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}})] = \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T. \qquad \text{(D.14)}$$

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma},\boldsymbol{\eta}}\mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}^T)] = (\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} - \mathbf{M}_B) \otimes \boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T. \qquad \text{(D.15)}$$

D.2.4 Decomposition

Finally, plugging (D.12) and (D.15) into (D.9), we have proven this Proposition.

E Proof for Corollary 1

By noticing $\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, we can write

$$\begin{aligned}
\mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})} &= \boldsymbol{\Gamma}\boldsymbol{\eta}(\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Gamma}\mathbf{P}_{\boldsymbol{\eta}(\boldsymbol{\Omega}^{-1})}\boldsymbol{\Gamma}^T. \\
\boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T &= \boldsymbol{\Gamma}\mathbf{P}_{\boldsymbol{\eta}(\boldsymbol{\Omega}^{-1})}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T = \boldsymbol{\Gamma}\mathbf{P}_{\boldsymbol{\eta}(\boldsymbol{\Omega}^{-1})}\boldsymbol{\Gamma}^T \cdot \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma} = \mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}.
\end{aligned}$$

Then, from (D.9), we have

$$\begin{aligned}
\operatorname{avar}[\sqrt{n}\operatorname{vec}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Gamma}})] &= \mathbf{M}_B \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + (\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} - \mathbf{M}_B) \otimes \boldsymbol{\Gamma}\mathbf{M}_\eta\boldsymbol{\Gamma}^T \\
&= \mathbf{P}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma} + \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma} \\
&= \left\{\mathbf{P}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{I}_r + \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\right\}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma} \\
&= \left\{\mathbf{I}_{pr} - \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\right\}\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma} \\
&= \left\{\mathbf{I}_{pr} - \mathbf{Q}_{\mathbf{B}^T(\boldsymbol{\Sigma}_{\mathbf{X}})} \otimes \mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma}^{-1})}\right\}(\mathbf{I}_p \otimes \mathbf{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}.
\end{aligned}$$

F Proof for Proposition 8

From Proposition 2 and Proposition 4, we see that the minimizer $\widehat{\mathbf{h}}_{\mathrm{RE}} = \mathbf{h}(\widehat{\boldsymbol{\phi}})$ of $\mathcal{F}(\mathbf{h}(\boldsymbol{\phi}), \widehat{\mathbf{h}}_{\mathrm{OLS}})$ is Fisher consistent. The rest of the proof relies on Shapiro's (1986) results on the asymptotics of overparameterized structural models. In order to apply Shapiro's (1986) theory in our context, we first can check that $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\mathrm{OLS}})$ satisfies: (1) $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\mathrm{OLS}}) \geq 0$ for all $\widehat{\mathbf{h}}_{\mathrm{OLS}}$ and $\mathbf{h}$; (2) $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\mathrm{OLS}}) = 0$ if and only if $\widehat{\mathbf{h}}_{\mathrm{OLS}} = \mathbf{h}$; and (3) $\mathcal{F}(\mathbf{h}, \widehat{\mathbf{h}}_{\mathrm{OLS}})$ is twice continuously differentiable in $\mathbf{h}$ and $\widehat{\mathbf{h}}_{\mathrm{OLS}}$. Recall from Section 4.2 that we use the subscript 0 to emphasize the true parameter: $\mathbf{h}_0$ and $\boldsymbol{\phi}_0$ correspond to the true distribution of $\boldsymbol{\epsilon}$. Then $\widehat{\boldsymbol{h}}_{\mathrm{OLS}}$ is $\sqrt{n}$-consistent for $\mathbf{h}_0$. Notice that $\widehat{\mathbf{h}}_{\mathrm{OLS}}$ is a smooth function of the sample covariance matrices which converges in distribution to the population covariance matrices, then by the delta method we know $\sqrt{n}(\widehat{\mathbf{h}}_{\mathrm{OLS}} - \mathbf{h}_0) \to N(0, \mathbf{K})$, for some positive definite covariance $\mathbf{K}$. Using Shapiro's (1986) Proposition 3.1 and Proposition 4.1, we will have $\sqrt{n}$-consistency results for $\widehat{\mathbf{h}}_{\mathrm{RE}} = \mathbf{h}(\widehat{\boldsymbol{\phi}})$ as shown in Proposition 8.