Regularized Reduced Rank Growth Curve Models

Yoshio Takane *, Kwanghee Jung, Heungsun Hwang

McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada

Abstract

The growth curve model (GCM), also known as GMANOVA, is a useful technique for investigating patterns of change in repeated measurement data over time and examining the effects of predictor variables on temporal trajectories. The reduced rank feature had been introduced previously to GCM for capturing redundant information in the criterion variables in a parsimonious way. In this paper, a ridge type of regularization was incorporated to obtain better estimates of parameters. Separate ridge parameters were allowed in column and row regressions, and the generalized singular value decomposition (GSVD) was applied for rank reduction. It was shown that the regularized estimates of parameters could be obtained in closed form for fixed values of ridge parameters. Permutation tests were used to identify the best dimensionality in the solution, and the K-fold cross validation method was used to choose optimal values of the ridge parameters. A bootstrap method was used to assess the reliability of parameter estimates. The proposed model was further extended to a mixture of GMANOVA and MANOVA. Illustrative examples were given to demonstrate the usefulness of the proposed method.

Key words: The growth curve model (or GMANOVA), Reduced rank approximation, Ridge-type regularization, A mixture of GMANOVA and MANOVA, Generalized singular value decomposition (GSVD), Permutation tests, K-fold cross validation, The bootstrap method

* Corresponding author (tel: 514-398-6125; fax: 514-398-4896).
Email addresses: takane@psych.mcgill.ca (Yoshio Takane), kwanghee.jung@mail.mcgill.ca (Kwanghee Jung), heungsun.hwang@mcgill.ca (Heungsun Hwang).

1 Introduction

The growth curve model (GCM; Potthof and Roy, 1964), also known as generalized MANOVA (or GMANOVA for short), is a useful technique for investigating patterns of change in repeated measurements of a response variable or variables over time and examining the effects of predictor variables on temporal trajectories. This type of model is often used in the analysis of longitudinal or repeated measurements data, often arising in psycho-physiological, biological and medical research. Recently, the reduced rank feature was introduced to GCM (Reinsel and Velu, 1998, 2003) to capture redundant information in the criterion variables in a parsimonious way. This additional feature allows the extraction of components of predictor variables that are most predictive of criterion variables. A series of components called redundancy components are mutually orthogonal and successively account for the maximum variance in the criterion variables.

In experimental studies conducted in biomedicine and psychology, we frequently encounter data with small sample sizes, which tend to produce estimates of parameters with large standard errors. The small sample size problem casts serious doubts about the adequacy of conventional estimation methods, such as the maximum likelihood estimation method, that largely rely on an asymptotic rationale. To remedy this situation, we incorporate a ridge type of regularization in estimating parameters in the reduced rank GCM. This method shrinks estimates of parameters toward zero, thereby reducing the variance of the estimates a great deal, while introducing a small bias. The net result is that estimates of parameters closer to true population values may be obtained. A ridge type of regularization method is particularly attractive when the sample size is small and/or predictor variables are nearly collinear (Hoerl and Kennard, 1970). This has been demonstrated recently in a variety of contexts in multivariate analysis (Hwang, 2009; Takane and Hwang, 2007; Takane, Hwang, and Abdi, 2008; Takane, and Jung, 2008, 2009). In this paper, we extend the basic methodology of ridge regularization to the reduced rank GCM and illustrate its use. We also consider an analogous extension of a mixture of the GMANOVA and MANOVA models (Chinchilli and Elswick, 1985) with the GMANOVA part subject to similar rank reduction, and the MANOVA part capturing the effects of extraneous variables.

This paper is organized as follows. We first present the model and the parameter estimation procedure for the regularized reduced rank GCM (Section 2.1). We then extend the model and the estimation procedure to a mixture of the GMANOVA and MANOVA models (Section 2.2). This is followed by expositions of permutation tests for selecting the best dimensionality in the solution, the K-fold cross validation method for choosing optimal values of ridge parameters, and the bootstrap method for assessing the reliability of param-

eter estimates (Section 2.3). Illustrative examples are given to demonstrate the usefulness of the proposed method in simulated and real data analysis situations (Section 3). The final section concludes the paper (Section 4).

2 The Methods

2.1 The reduced rank GCM and the regularized parameter estimation

Let $\mathbf{Y}$ denote an n by p matrix of criterion variables. In the GCM setting, this matrix typically consists of multiple measurements of a response variable at p time points from a group of n subjects or cases, although in more general settings, it could be any multivariate data matrix. We assume that there is some additional information about the subjects and/or about the variables in $\mathbf{Y}$ that may be used to predict parts of $\mathbf{Y}$. Let $\mathbf{X}$ denote an n by q ($q \leq n$) matrix of predictor variables for subjects such as their group memberships (e.g., treatment groups) and other demographic information. Let $\mathbf{H}$ denote a p by d ($d \leq p$) matrix of predictor variables for time points (or variables) in $\mathbf{Y}$ that capture the relationships among the columns of $\mathbf{Y}$ such as the coefficients of orthogonal polynomials over time. (The matrix $\mathbf{X}$ is often called a between-subjects design matrix, and $\mathbf{H}$ a within-subjects design matrix.) Then, the GCM may be written as

$$\mathbf{Y} = \mathbf{XBH}' + \mathbf{E}, \tag{1}$$

where $\mathbf{B}$ is a q by d matrix of regression coefficients, and $\mathbf{E}$ is an n by p matrix of disturbance terms. In the reduced rank GCM, we assume that there is some redundancy in $\mathbf{B}$, so that

$$\text{rank}(\mathbf{B}) = r \leq \min(q, d) \tag{2}$$

(Reinsel and Velu, 1998, 2003). A model of the above form has existed outside the realm of GCM, e.g., 2-way CANDELINC (CANonical DEcomposition under LINnear Constraints; Carrol, Pruzansky, and Kruskal, 1980), and as a special case of CPCA (Constrained Principal Component Analysis; Takane and Shibayama, 1991; Takane and Hunter, 2001). Note that, if there is no obvious $\mathbf{H}$ available, we set $\mathbf{H} = \mathbf{I}$, and the model reduces to a simple MANOVA or redundancy analysis model (Van den Wollenberg, 1977; van der Leeden, 1990).

Parameters in the reduced rank GCM are usually estimated by the maximum likelihood (ML) method (Reinsel and Velu, 1998, 2003) or by the least squares

(LS) method (Carroll et al., 1980; Takane et al., 1991, 2001). We use the latter with the provision of its extension to regularized estimation in mind. The structure of derivations for the regularized estimation is remarkably similar to that for the non-regularized case. In the ordinary LS estimation, we minimize

$$\phi(\mathbf{B}) = \mathrm{SS}(\mathbf{Y} - \mathbf{XBH}') \tag{3}$$

with respect to $\mathbf{B}$ subject to the rank restriction (2). To achieve this goal, we first rewrite $\phi(\mathbf{B})$ as (Takane and Shibayama, 1991; ten Berge, 1993):

$$\begin{aligned}\phi(\mathbf{B}) &= \mathrm{SS}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\mathbf{H}') + \mathrm{SS}(\hat{\mathbf{B}} - \mathbf{B})_{X'X,\,H'H} \\ &= \mathrm{SS}(\mathbf{Y}) - \mathrm{SS}(\mathbf{Y})_{P_X,\,P_H} + \mathrm{SS}(\hat{\mathbf{B}} - \mathbf{B})_{X'X,\,H'H},\end{aligned} \tag{4}$$

where $\mathrm{SS}(\mathbf{A})_{M,N} = \mathrm{tr}(\mathbf{A}'\mathbf{MAN})$, $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ and $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^-\mathbf{H}'$ are orthogonal projectors onto the column spaces of $\mathbf{X}$ and $\mathbf{H}$, respectively, and

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{YH}(\mathbf{H}'\mathbf{H})^-, \tag{5}$$

is a rank free LS estimate of $\mathbf{B}$. Here "$-$" indicates a generalized inverse (g-inverse). Note that while $\hat{\mathbf{B}}$ in (5) is not unique if $\mathbf{X}$ or $\mathbf{H}$ is singular, the decomposition (4) is unique. To obtain a unique estimate of $\hat{\mathbf{B}}$, we can use the Moore-Penrose inverse for $(\mathbf{X}'\mathbf{X})^-$ and $(\mathbf{H}'\mathbf{H})^-$. Since the first and the second terms on the right hand side of (4) are unrelated to $\mathbf{B}$, the reduced rank estimate of $\mathbf{B}$ can be obtained by minimizing the third term. This can be done via the generalized singular value decomposition (GSVD) of $\hat{\mathbf{B}}$ with metric matrices $\mathbf{X}'\mathbf{X}$ and $\mathbf{H}'\mathbf{H}$. This GSVD problem is written as

$$\mathrm{GSVD}(\hat{\mathbf{B}})_{X'X,\,H'H}. \tag{6}$$

Note that here GSVD refers to an SVD under nonidentity metrics (Greenacre, 1984), that is,

$$\hat{\mathbf{B}} = \mathbf{UDV}', \tag{7}$$

where $\mathbf{U}$ is called the matrix of left (generalized) singular vectors such that $\mathbf{U}'\mathbf{X}'\mathbf{XU} = \mathbf{I}$, $\mathbf{V}$ the matrix of right (generalized) singular vectors such that $\mathbf{V}'\mathbf{H}'\mathbf{HV} = \mathbf{I}$, and $\mathbf{D}$ a positive definite (pd) diagonal matrix of (generalized) singular values. The reduced rank estimate of $\mathbf{B}$ is obtained from the above GSVD by retaining only the portions of $\mathbf{U}$, $\mathbf{D}$, and $\mathbf{V}$ pertaining to the r dominant (generalized) singular values (assuming that the rank of $\hat{\mathbf{B}}$ is at

least r), that is,

$$\tilde{\mathbf{B}} = \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r', \tag{8}$$

where $\mathbf{U}_r$, $\mathbf{D}_r$, and $\mathbf{V}_r$ are the portions of $\mathbf{U}$, $\mathbf{D}$, and $\mathbf{V}$ pertaining to the r largest (generalized) singular values of $\hat{\mathbf{B}}$ with metrics $\mathbf{X}'\mathbf{X}$ and $\mathbf{H}'\mathbf{H}$. Further details on GSVD can be found in Takane and Hunter (2001, p. 415) and Takane (2003).

We now extend the above method to the ridge LS (RLS) estimation. As before, the reduced-rank RLS estimate of regression parameters can be obtained by first obtaining a rank free RLS estimate of $\mathbf{B}$, followed by a GSVD. (In essence, equations (3), (4), and (5) in the LS estimation become (9), (16), and (17), respectively, in the RLS estimation.) In the RLS estimation, we minimize

$$\begin{aligned}\phi_{\lambda,\rho}(\mathbf{B}) &= \mathrm{SS}(\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{H}') \\ &\quad + \lambda \mathrm{SS}(\mathbf{B})_{P_{X'}, H'H} + \rho \mathrm{SS}(\mathbf{B})_{X'X, P_{H'}} + \lambda\rho \mathrm{SS}(\mathbf{B})_{P_{X'}, P_{H'}},\end{aligned} \tag{9}$$

where λ and ρ are small positive numbers called ridge parameters, and $\mathbf{P}_{X'} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-}\mathbf{X}$ and $\mathbf{P}_{H'} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-}\mathbf{H}$ are the orthogonal projectors onto the row spaces of $\mathbf{X}$ and $\mathbf{H}$. Note that $\mathbf{P}_{X'}$ and $\mathbf{P}_{H'}$ reduce to $\mathbf{I}_q$ and $\mathbf{I}_d$, respectively, when $\mathbf{X}$ and $\mathbf{H}$ are columnwise nonsingular. The terms added to the LS criterion (3), that is, the second, third, and fourth terms in (9), have the effects of "regularizing" an estimate of $\mathbf{B}$ by shrinking it toward zero.

Let

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X})^{-}\mathbf{X}', \tag{10}$$

and

$$\mathbf{R}_H(\rho) = \mathbf{H}(\mathbf{H}'\mathbf{M}_H(\rho)\mathbf{H})^{-}\mathbf{H}' \tag{11}$$

denote ridge operators (Takane and Yanai, 2008), where

$$\mathbf{M}_X(\lambda) = \mathbf{I}_n + \lambda(\mathbf{X}\mathbf{X}')^{+}, \tag{12}$$

and

$$\mathbf{M}_H(\rho) = \mathbf{I}_p + \rho(\mathbf{H}\mathbf{H}')^{+} \tag{13}$$

are called ridge metric matrices (Takane and Yanai, 2008), and "+" indicates the Moore-Penrose inverse. Note that

$$\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X} = \mathbf{X}'\mathbf{X} + \lambda\mathbf{P}_{X'}, \tag{14}$$

and

$$\mathbf{H}'\mathbf{M}_H(\rho)\mathbf{H} = \mathbf{H}'\mathbf{H} + \rho\mathbf{P}_{H'}. \tag{15}$$

To minimize (9) with respect to $\mathbf{B}$ subject to the rank restriction (2), we first rewrite the criterion as:

$$\begin{aligned}\phi_{\lambda,\rho}(\mathbf{B}) &= \mathrm{SS}(\mathbf{Y}) \\ &\quad - \mathrm{SS}(\mathbf{Y})_{R_X(\lambda),\, R_H(\rho)} + \mathrm{SS}(\hat{\mathbf{B}}(\lambda,\rho) - \mathbf{B})_{X'M_X(\lambda)X,\, H'M_H(\rho)H},\end{aligned} \tag{16}$$

where

$$\hat{\mathbf{B}}(\lambda,\rho) = (\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\mathbf{H}(\mathbf{H}'\mathbf{M}_H(\rho)\mathbf{H})^{-} \tag{17}$$

is a rank free RLS estimate of $\mathbf{B}$. Eq. (16) indicates that the reduced rank estimate of $\mathbf{B}$ can be obtained by minimizing the third term on the right hand side of (16), which is achieved via

$$\mathrm{GSVD}(\hat{\mathbf{B}}(\lambda,\rho))_{X'M_X(\lambda)X,\, H'M_H(\rho)H}. \tag{18}$$

Equivalence between (9) and (16) can be shown as follows. From (9), we have

$$\begin{aligned}\phi_{\lambda,\rho}(\mathbf{B}) &= \mathrm{tr}(\mathbf{Y}'\mathbf{Y} - 2\mathbf{B}'\mathbf{X}'\mathbf{Y}\mathbf{H} + \mathbf{B}'\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X}\mathbf{B}\mathbf{H}'\mathbf{M}_H(\rho)\mathbf{H}) \\ &= \mathrm{tr}(\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{R}_X(\lambda)\mathbf{Y}\mathbf{R}_H(\rho) + \mathbf{Y}'\mathbf{R}_X(\lambda)\mathbf{Y}\mathbf{R}_H(\rho) \\ &\qquad - 2\mathbf{B}'\mathbf{X}'\mathbf{Y}\mathbf{H} + \mathbf{B}'\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X}\mathbf{B}\mathbf{H}'\mathbf{M}_H(\rho)\mathbf{H}) \\ &= \mathrm{SS}(\mathbf{Y}) - \mathrm{SS}(\mathbf{Y})_{R_X(\lambda),\, R_H(\rho)} \\ &\qquad + \mathrm{SS}(\hat{\mathbf{B}}(\lambda,\rho) - \mathbf{B})_{X'M_X(\lambda)X,\, H'M_H(\rho)H},\end{aligned} \tag{19}$$

which is (16).

2.2 A mixture of the GMANOVA-MANOVA models

Now, we consider an extension of the GCM to a mixture of the GMANOVA and MANOVA models (Chinchilli and Elswick, 1985). Suppose that $\mathbf{X}$ is split

into two subsets, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ where $\mathbf{X}_1$ is an n by q_1 matrix of predictor variables (analogous to the entire $\mathbf{X}$ in the previous section), and $\mathbf{X}_2$ is an n by q_2 ($q_1 + q_2 = q$) matrix of extraneous variables whose effects are to be partialled out in a manner analogous to the analysis of covariance. We consider the following model:

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' + \mathbf{X}_2\mathbf{B}_2 + \mathbf{E}, \tag{20}$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are q_1 by d and q_2 by p matrices, respectively, of regression coefficients. As before, $\mathbf{B}_1$ is subject to the rank restriction

$$\text{rank}(\mathbf{B}_1) \leq \min(q_1, d), \tag{21}$$

while no such restriction is imposed on $\mathbf{B}_2$. The term related to $\mathbf{X}_2$ is included in the model for eliminating its effects in fitting the GCM. (Note that $\mathbf{H}$ in the previous section is now written as $\mathbf{H}_1$.)

As before, we first discuss the LS estimation, and then an extension to the RLS estimation. In the LS estimation, we minimize

$$\phi(\mathbf{B}_1, \mathbf{B}_2) = \text{SS}(\mathbf{Y} - \mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' - \mathbf{X}_2\mathbf{B}_2) \tag{22}$$

with respect to $\mathbf{B}_1$ and $\mathbf{B}_2$ subject to $\text{rank}(\mathbf{B}_1) \leq r$. We first rewrite the model by orthogonalizing the two terms in the model:

$$\mathbf{Y} = \mathbf{Q}_{X_2}\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' + \mathbf{X}_2\mathbf{B}_2^* + \mathbf{E}, \tag{23}$$

where

$$\mathbf{B}_2^* = \mathbf{B}_2 + (\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1', \tag{24}$$

and $\mathbf{Q}_{X_2} = \mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'$ is the orthogonal projector on the null space of $\mathbf{X}_2'$. The criterion (22) can also be rewritten as

$$\phi(\mathbf{B}_1, \mathbf{B}_2^*) = \text{SS}(\mathbf{Y} - \mathbf{Q}_{X_2}\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' - \mathbf{X}_2\mathbf{B}_2^*), \tag{25}$$

which can be further rewritten as

$$\begin{aligned}\phi(\mathbf{B}_1, \mathbf{B}_2^*) = {} & \text{SS}(\mathbf{Y}) - \text{SS}(\mathbf{Y})_{P_{Q_{X_2}X_1}, P_{H_1}} - \text{SS}(\mathbf{Y})_{P_{X_2}, I} \\ & + \text{SS}(\hat{\mathbf{B}}_1 - \mathbf{B}_1)_{X_1'Q_{X_2}X_1, H_1'H_1} + \text{SS}(\hat{\mathbf{B}}_2^* - \mathbf{B}_2^*)_{X_2'X_2, I},\end{aligned} \tag{26}$$

where

$$\hat{\mathbf{B}}_1 = (\mathbf{X}_1'\mathbf{Q}_{X_2}\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{Q}_{X_2}\mathbf{Y}\mathbf{H}_1(\mathbf{H}_1'\mathbf{H}_1)^-, \tag{27}$$

and

$$\hat{\mathbf{B}}_2^* = (\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{Y}. \tag{28}$$

Equivalence between (25) and (26) is shown in Appendix (A). Since the first and second terms on the right side of (26) are unrelated to $\mathbf{B}_1$, and the fourth term can always be made equal to zero by taking $\mathbf{B}_2^* = \hat{\mathbf{B}}_2^*$, (26) can be minimized by minimizing the third term with respect to $\mathbf{B}_1$ subject to the rank restriction (21). This is again obtained via

$$\mathrm{GSVD}(\hat{\mathbf{B}}_1)_{X_1'Q_{X_2}X_1,\, H_1'H_1}. \tag{29}$$

An estimate of $\mathbf{B}_2$ is obtained by

$$\tilde{\mathbf{B}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'(\mathbf{Y} - \mathbf{X}_1\tilde{\mathbf{B}}_1\mathbf{H}_1') = \hat{\mathbf{B}}_2^* - (\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{X}_1\tilde{\mathbf{B}}_1\mathbf{H}_1', \tag{30}$$

where $\tilde{\mathbf{B}}_1$ is obtained from (29).

We now extend the LS estimation to the RLS estimation. Again, a reduced rank RLS estimate of $\mathbf{B}_1$ is obtained by first obtaining a rank free RLS estimate of $\mathbf{B}_1$ followed by a GSVD, which is similar to the simple GMANOVA case. Equations (22) through (30) in the LS estimation turn into equations (31), (35), (37), (38), (44), (41), (42), (45), and (46), respectively, in the RLS estimation. In the RLS estimation, we minimize

$$\begin{aligned}\phi_{\lambda,\rho}(\mathbf{B}_1, \mathbf{B}_2) &= \mathrm{SS}(\mathbf{Y} - \mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' - \mathbf{X}_2\mathbf{B}_2) \\ &\quad + \lambda\mathrm{SS}(\mathbf{C}_1)_{P_{X'},\, I_p} + \rho\mathrm{SS}(\mathbf{C}_2)_{R'X'XR,\, I_p} + \lambda\rho\mathrm{SS}(\mathbf{C}_2)_{R'P_{X'}R,\, I_p},\end{aligned} \tag{31}$$

with respect to $\mathbf{B}_1$ and $\mathbf{B}_2$ under the rank restriction (21), where

$$\mathbf{C}_1 = \begin{bmatrix} \mathbf{B}_1\mathbf{H}_1' \\ \mathbf{B}_2 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} \mathbf{B}_1\mathbf{H}_1'(\mathbf{H}_1\mathbf{H}_1')^{-1/2} \\ \mathbf{B}_2 \end{bmatrix}, \tag{32}$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{P}_{X_1'} & \mathbf{0} \\ -\mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1 & \mathbf{0} \end{bmatrix}, \tag{33}$$

and

$$\mathbf{G} = \mathbf{X}_2'\mathbf{M}_X(\lambda)\mathbf{X}_2. \tag{34}$$

We then rewrite the model (20) as

$$\mathbf{Y} = \mathbf{S}_{X_2}(\lambda)\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' + \mathbf{X}_2\mathbf{B}_2^* + \mathbf{E}, \tag{35}$$

where

$$\mathbf{S}_{X_2}(\lambda) = \mathbf{I}_n - \mathbf{X}_2\mathbf{G}^-\mathbf{X}_2', \tag{36}$$

and

$$\mathbf{B}_2^* = \mathbf{B}_2 + \mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1'. \tag{37}$$

Note that the first two terms on the right hand side of (35) are orthogonal with respect to $\mathbf{M}_X(\lambda)$ defined in (12). We also rewrite the criterion (31) according to this rewritten model:

$$\begin{aligned}\phi_{\lambda,\rho}(\mathbf{B}_1, \mathbf{B}_2^*) = {} & \mathrm{SS}(\mathbf{Y} - \mathbf{S}_{X_2}(\lambda)\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' - \mathbf{X}_2\mathbf{B}_2^*) \\ & + \lambda\mathrm{SS}(\mathbf{C}_1^*)_{N'P_{X'}N,\, I_p} + \rho\mathrm{SS}(\mathbf{B}_1)_{X_1'S_{X_2}(\lambda)^2X_1,\, P_{H_1'}} \\ & + \lambda\rho\mathrm{SS}(\mathbf{B}_1)_{P_{X_1'} + X_1'X_2G^-P_{X_2'}G^-X_2'X_1,\, P_{H_1'}},\end{aligned} \tag{38}$$

where

$$\mathbf{C}_1^* = \begin{bmatrix} \mathbf{B}_1\mathbf{H}_1' \\ \mathbf{B}_2^* \end{bmatrix}, \tag{39}$$

and

$$\mathbf{N} = \begin{bmatrix} \mathbf{P}_{X_1'} & \mathbf{0} \\ -\mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1 & \mathbf{P}_{X_2'} \end{bmatrix}. \tag{40}$$

Note that $\mathbf{N}$ is such that $\mathbf{C}_1 = \mathbf{N}\mathbf{C}_1^*$.

Let

$$\hat{\mathbf{B}}_1(\lambda, \rho) = \mathbf{A}^-\mathbf{X}_1'\mathbf{S}_{X_2}(\lambda)\mathbf{Y}\mathbf{H}_1(\mathbf{H}_1'\mathbf{M}_{H_1}(\rho)\mathbf{H}_1)^- \tag{41}$$

and

$$\hat{\mathbf{B}}_2^*(\lambda) = \mathbf{G}^- \mathbf{X}_2' \mathbf{Y} \tag{42}$$

be rank free estimates of $\mathbf{B}_1$ and $\mathbf{B}_2^*$ that minimize (38), where

$$\mathbf{A} = \mathbf{X}_1' \mathbf{S}_{X_2}(\lambda) \mathbf{M}_X(\lambda) \mathbf{S}_{X_2}(\lambda) \mathbf{X}_1. \tag{43}$$

Then, (38) can further be rewritten as

$$\begin{aligned} \phi_{\lambda,\rho}(\mathbf{B}_1, \mathbf{B}_2^*) = \mathrm{SS}(\mathbf{Y}) - \mathrm{SS}(\mathbf{Y})_{R_{S_{X_2}(\lambda)X_1}(\lambda),\, R_{H_1}(\rho)} - \mathrm{SS}(\mathbf{Y})_{R_{X_2}(\lambda),\, I} \\ + \mathrm{SS}(\hat{\mathbf{B}}_1(\lambda,\rho) - \mathbf{B}_1)_{A,\, H_1' M_{H_1}(\rho) H_1} + \mathrm{SS}(\hat{\mathbf{B}}_2^*(\lambda) - \mathbf{B}_2^*)_{G,\, I}, \end{aligned} \tag{44}$$

where $\mathbf{R}_{S_{X_2}(\lambda)X_1}(\lambda)$, and $\mathbf{R}_{X_2}(\lambda)$ are defined analogously to (10), $\mathbf{R}_{H_1}(\rho)$ to (11), and $\mathbf{M}_{H_1}(\rho)$ to (13). Note that $\mathbf{A} = \mathbf{X}_1' \mathbf{S}_{X_2}(\lambda) \mathbf{M}_X(\lambda) \mathbf{S}_{X_2}(\lambda) \mathbf{X}_1 = \mathbf{X}_1' \mathbf{S}_{X_2}(\lambda) \mathbf{X}_1 + \lambda \mathbf{P}_{X_1'}$. Equivalence between (38) and (44) is shown in Appendix (B). The first three terms on the right hand side of (44) are unrelated to $\mathbf{B}_1$, and the fifth term can always be made equal to zero by taking $\mathbf{B}_2^* = \hat{\mathbf{B}}_2^*(\lambda)$. Thus (44) can be minimized by minimizing the fourth term with respect to $\mathbf{B}_1$ subject to $\mathrm{rank}(\mathbf{B}_1) \leq \min(q_1, d)$. A reduced rank ridge estimate of $\mathbf{B}_1$ is obtained via

$$\mathrm{GSVD}(\hat{\mathbf{B}}_1(\lambda,\rho))_{A,\, H_1' M_{H_1}(\rho) H_1}. \tag{45}$$

The estimate of $\mathbf{B}_2$ is obtained by

$$\tilde{\mathbf{B}}_2(\lambda) = \hat{\mathbf{B}}_2^*(\lambda) - \mathbf{G}^- \mathbf{X}_2' \mathbf{X}_1 \tilde{\mathbf{B}}_1(\lambda,\rho) \mathbf{H}_1', \tag{46}$$

where $\tilde{\mathbf{B}}_1(\lambda,\rho)$ is obtained from (45).

2.3 Permutation tests, cross validation and the bootstrap method

Two important decisions have to be made in the application of regularized reduced rank GCM: The choice of dimensionality and the choice of optimal values of the ridge (regularization) parameters λ and ρ. We discuss these topics in turn and then, introduce the bootstrap method to assess the reliability of parameter estimates.

We use permutation tests to choose the best dimensionality (the number of components) r in the solution. In the permutation tests, rows of $\mathbf{X}$ are randomly permuted many times. The permutation operation effectively eliminates

any (systematic) associations between $\mathbf{X}$ and $\mathbf{Y}$, thereby allowing us to obtain the null distribution of singular values. Our proposed model is applied to each permuted $\mathbf{X}$ and the original $\mathbf{Y}$ repeatedly to obtain the null distribution of the largest singular value. If these singular values are larger than the largest singular value obtained from the original $\mathbf{X}$ and $\mathbf{Y}$ less than $100\alpha\%$ times, the first component is considered statistically significant at the α level. If the first component is significant, we eliminate its effect from $\mathbf{X}$ (deflating $\mathbf{X}$) and apply the same procedure as above to test the significance of the second component, and so on. We continue this procedure until we find a nonsignificant component or reach the maximum possible number of components. See Takane and Hwang (2002) for more applications of permutation tests in a similar situation, and Legendre and Legendre (1998) and ter Braak and Šmilauer (1998) for more general discussions on the permutation tests in similar contexts. We may apply the above procedure with different values of λ and ρ in cases where the best dimensionality may depend on the values of λ and ρ.

A bit of a cautionary remark is in order. The permutation tests are based on the assumption of exchangeability of observations (rows of $\mathbf{X}$ or of $\mathbf{Y}$) under the null hypothesis that $\mathbf{X}$ has no predictability on $\mathbf{Y}$. The most representative cases of exchangeability include those in which observations are independently and identically distributed, and in which they are normally distributed with homogeneous variances and covariances across observations (Good, 2005, p. 24). There is some evidence, however, showing that the permutation tests are not so robust when the exchangeability assumption is violated (e.g., Boik, 1987; Romano, 1990).

We use K-fold cross validation (Hastie, Tibshirani, and Friedman, 2001) to choose an optimal value of λ and ρ. In this method, the data are randomly divided into K subsets. One of the K sets is set aside as the test sample, and model parameters are estimated from the remaining $K-1$ subsets (called the calibration or training sample). These estimates are then applied to the test sample to estimate the amount of prediction error. This is repeated K times with the test sample changed systematically, and the prediction errors accumulated over the K test samples. We then compare the prediction error for different pairs of λ and ρ (e.g., $\lambda, \rho = 0, .5, 1, 3, 5$), and choose the λ and ρ pair associated with the smallest value of prediction error.

A bootstrap method (Efron and Tibshirani, 1993) is used to assess the reliability of parameter estimates. In this method, random samples of size n (equal to the size of the original data set) are repeatedly sampled from the original data set with replacement. Estimates of parameters are obtained for each bootstrap sample. We then calculate the means and the variances of the estimates across the bootstrap samples to estimate biases and standard errors of the original estimates. Significance tests of estimated coefficients may also be performed as a by-product of the bootstrap procedure. We simply count the number of

times bootstrap estimates "cross" over zero (if the original estimate is positive, we count the number of times the bootstrap estimates turn out to be negative, and vice versa). If the relative frequency (p-value) of crossovers is less than a prescribed value of α, we conclude that the coefficient is significantly positive (or negative).

3 Empirical Demonstrations

In this section we provide empirical demonstrations of the usefulness of the proposed method. The first study investigates the effect of regularization using a Monte Carlo technique. The second and third studies pertain to the analysis of real data sets. The second study employs the reduced rank GCM (GMANOVA), while the third the mixture of the reduced rank GCM and MANOVA.

3.1 A Monte Carlo study

We first examine the quality of the RLS estimator compared to the OLS (ordinary least squares) estimator using a Monte-Carlo technique. The quality of an estimator can be measured by how close it is on average to its population counterpart. For this purpose we use the mean square error (MSE). MSE is the expected value of $\mathrm{SS}(\hat{\theta}-\theta)$, where θ is the vector of true parameters and $\hat{\theta}$ is the vector of their estimators. MSE can be decomposed into two parts: $\mathrm{MSE} = \mathrm{E}[\mathrm{SS}(\theta - \mathrm{E}(\hat{\theta}))] + \mathrm{E}[\mathrm{SS}(\hat{\theta} - \mathrm{E}(\hat{\theta}))]$. The first term on the right hand side is the squared bias, and the second term is the variance of the estimator $\hat{\theta}$. The estimator with a smaller value of MSE is considered a better estimator. The ridge LS (RLS) estimator is shown to provide smaller MSE's than OLS (Takane and Hwang, 2007; Takane and Jung, 2008). This is due to the fact that the OLS estimator tends to have a large variance, while it is unbiased. The RLS estimator, on the other hand, is slightly biased, but it has much smaller variance, resulting in a smaller MSE.

A small Monte Carlo study was conducted to verify the above expectation. First, a population reduced rank GCM was postulated, from which 100 replicated data sets of varying sample sizes ($N = 20, 50, 80, 100$) were generated. The reduced rank GCM was then fitted using the RLS estimation method to derive the estimates of regression coefficients with the values of λ and ρ systematically varied ($\lambda, \rho = 0, .5, 1, 3, 5$). Average MSE, squared bias, and variance were calculated using the assumed population values of regression coefficients. In the assumed population model, the number of criterion variables (p) was set to 15, and that of predictor variables (q) to 5. (We also tried $p = 5$, but the results were similar to those for $p = 15$.) Each row of $\mathbf{Y}$ was generated

according to $\mathbf{y}'_j = \mathbf{x}'_j\mathbf{B}\mathbf{H}' + \mathbf{e}'_j$, where $\mathbf{x}'_j \sim N(\mathbf{0}, \mathbf{\Sigma})$, and $\mathbf{e}'_j \sim \mathrm{N}(\mathbf{0}, \gamma^2\mathbf{I}_p)$ for $j = 1, \ldots, N$. The diagonal elements of $\mathbf{\Sigma}$ were set to unity, and off-diagonal elements varied at three levels (0, .5, and .9). The value of γ^2 was also varied at three levels (.5, 1, and 2). The design matrix for criterion variables $\mathbf{H}$ was set to the second order orthogonal polynomials with a constant term, that is,

$$\mathbf{H}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 91 & 52 & 19 & -8 & -29 & -44 & -53 & -56 & -53 & -44 & -29 & -8 & 19 & 52 & 91 \end{pmatrix}.$$

Matrix $\mathbf{Y}$ was columnwise centered, and $\mathbf{X}$ and $\mathbf{H}$ were columnwise standardized before the analysis was conducted. For each data set, elements of $\mathbf{B}$ were initially generated by uniform random numbers. Matrix $\mathbf{B}$ was then subjected to GSVD to reduce its rank to 1.

Figure 1 presents the main results of the study for a particular combination of conditions, namely the medium level of correlations among the predictor variables (off-diagonal elements of $\mathbf{\Sigma} = .5$), a high level of error variance ($\gamma^2 = 2$), and rank($\mathbf{B}$) = 1. The figure shows the MSE of regression coefficients in the reduced rank GCM as a function of the sample size and the ridge parameters (λ and ρ). The top left panel displays MSE as a function of sample size and λ (0, .5, 1, 3, and 5) for $\rho = 0$, the top right panel the same for $\rho = .5$. In both cases, MSE is larger at $\lambda = 0$ (this case corresponds with the non-regularized case), decreases as the value of λ departs from 0, but then rises again. This tendency is clearer for small sample sizes, although it can still be observed for larger sample sizes. The bottom panels show MSE's as a function of sample size and ρ (0, .5, 1, 3, and 5), but for fixed values of λ ($\lambda = 0$ for the left panel, and $\lambda = 3$ for the right panel). A similar tendency (that MSE decreases initially, but eventually picks up again as ρ increases) can be observed in the left panel, where $\lambda = 0$. However, this tendency is no longer observed (MSE monotonically increases) at $\lambda = 3$. MSE takes minimum values for $\lambda = 3$ and $\rho = 0$ in all cases presented in Figure 1. Overall, the effect of ρ on MSE is much weaker than that of λ. This may be because in the present case the columns of $\mathbf{H}$ are always orthogonal as they are the coefficients of orthogonal polynomials, and there is no instance of multicollinearity.

Table 1 compares MSE, squared bias, and variance between non-regularized and optimally regularized ($\lambda = 3$, and $\rho = 0$) cases as a function of sample size. It can be observed that squared bias increases and the variance decreases as a result of regularization, but that their sum, MSE, decreases across all sample sizes. This is because the amount of decrease in variance is much larger than the amount of increase in squared bias.

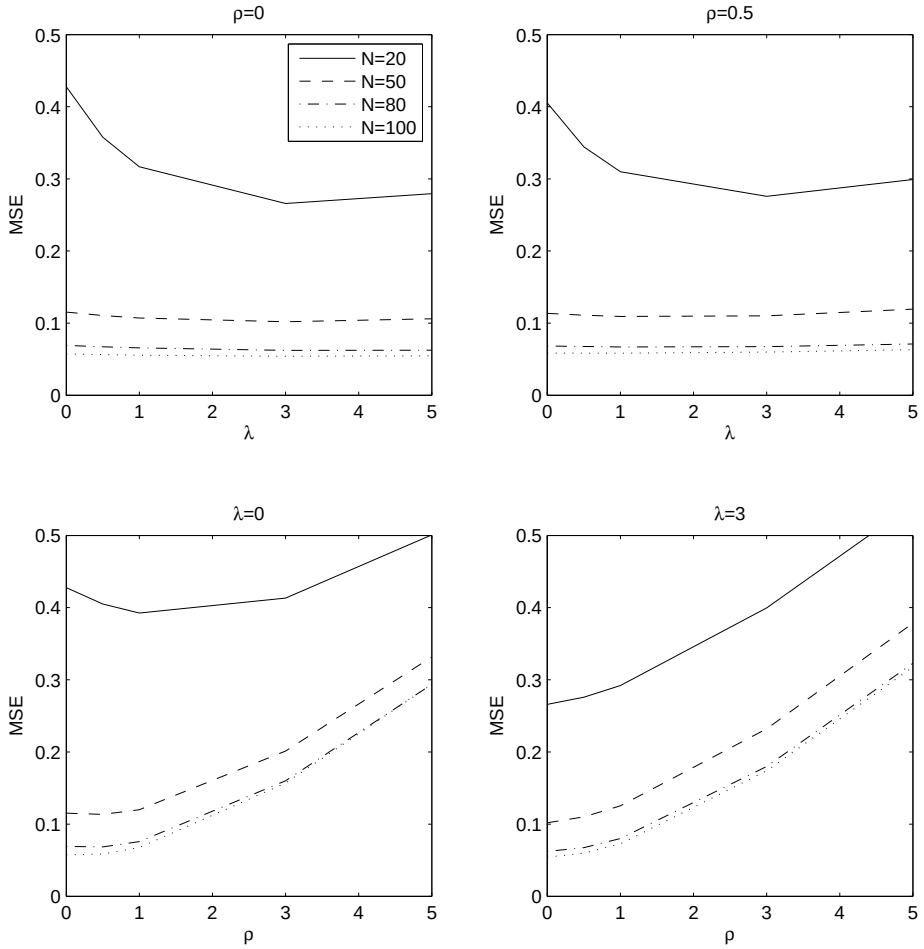

Fig. 1. Plot of MSE as a function of sample size N and the ridge parameters λ and ρ. Top left: $\rho = 0$, and $\lambda = 0, .5, 1, 3, 5$; Top right: $\rho = .5$ and $\lambda = 0, .5, 1, 3, 5$; Bottom left: $\lambda = 0$, and $\rho = 0, .5, 1, 3, 5$; Bottom right: $\lambda = 3$, and $\rho = 0, .5, 1, 3, 5$.

Table 1
MSE, squared bias and variance at optimal values of λ and ρ as a function of sample size (N).

	Non-Regularized			Regularized				
N	MSE	BIAS	VAR	MSE	BIAS	VAR	λ_{opt}	ρ_{opt}
20	.428	.004	.424	.266	.063	.203	3	0
50	.115	.001	.114	.102	.014	.088	3	0
80	.069	.001	.068	.062	.004	.058	3	0
100	.057	.001	.056	.054	.004	.050	3	0

3.2 Coronary sinus potassium level over time

The first real data set we analyze comes from a biomedical study (Grizzle and Allen, 1969). The study involved 36 dogs randomly assigned to one of the following four treatment groups: Group 1 - Control group (9 dogs), Group 2 - Extrinsic cardiac denervation three weeks prior to coronary occlusion (10

dogs), Group 3 - Extrinsic cardiac denervation immediately prior to coronary occlusion (8 dogs), Group 4 - Bilateral thoracic sympathectomy and stellectomy three weeks prior to coronary occlusion (9 dogs). This data set clearly has a small sample size (36 cases). The level of coronary sinus potassium was measured for each dog 1, 3, 5, 7, 9, 11 and 13 minutes after coronary occlusion (7 time points). The design matrix for time **H** consists of coefficients of orthogonal polynomials of up to the third order with a constant term. (This is based on the finding by Grizzle and Allen.) The criterion variables (**Y**) were columnwise centered, and the predictor variables (**X**) indicating group membership and the design matrix **H** were normalized before the analysis.

The reduced rank GCM was fitted to the data. Permutation tests were first applied with varying values of λ and ρ (0, .5, 1, and 5), which consistently indicated one and only one significant dimension. The 36-fold cross validation method was then applied with the dimensionality equal to one. As shown in Table 2, the prediction error takes the minimum value of .787 at $\lambda = 1$ and $\rho = 1$. This compares favorably with the value of .803 for the non-regularized case.

Table 2
K-fold Cross Validation Results for Grizzle and Allen's (1969) data

λ	ρ	Prediction Error
0	0	.803
.5	0	.795
1	0	.791
5	0	.795
0	.5	.795
.5	.5	.790
1	.5	.788
5	.5	.801
0	1	.790
.5	1	.788
1	1	.787*
5	1	.807
0	5	.803
.5	5	.808
1	5	.813
5	5	.849

*Optimal combination

A bootstrap method was then used to assess the reliability of the parameter estimates. This was done for both $\lambda = 0$ and $\rho = 0$ (the non-regularized case) and the optimal values of λ and ρ (the regularized case) for comparison. One thousand bootstrap samples were generated, parameter estimates were obtained for each sample, and the standard errors of the estimated parameters were calculated. Table 3 compares the non-regularized (ordinary) LS estimates and the best regularized LS estimates of regression coefficients. Each cell in the table has an estimate and its standard error in parentheses. The p-values were also calculated, as described in Section 2.3, and the significance of esti-

mated coefficients is indicated by one (5%) or two asterisks (1%) according to the p-values. The overall pattern of significance of estimated coefficients remains the same for both non-regularized and regularized cases. As expected, the estimates of coefficients tend to be smaller in the regularized estimation, showing the shrinkage effect of regularization. The standard errors are also smaller in the regularized estimation. This, however, does not necessarily imply that the regularized estimates are more reliable. Note, however, that the ratio of estimated coefficients to their standard errors is consistently larger in the regularized case, indicating that the smaller standard errors are not merely due to the shrinkage effect of regularization, but reflect the fact that the regularized estimates are indeed more reliable.

Table 3
Comparison of estimated regression coefficients **B** from Grizzle and Allen's (1969) data.

	Non-Regularized ($\lambda = 0$ and $\rho = 0$)				Regularized ($\lambda = 1$ and $\rho = 1$)			
	Const.	Linear	Quad.	Cub.	Const.	Linear	Quad.	Cub.
Group 1	**.566	**.195	-.057	**-.095	**.446	**.154	-.045	**-.075
	(.152)	(.061)	(.052)	(.037)	(.118)	(.047)	(.040)	(.029)
Group 2	**-.498	**.172	.050	**.083	-.396	-.137	.040	**.066
	(.156)	(.046)	(.044)	(.029)	(.121)	(.036)	(.034)	(.022)
Group 3	.116	.040	-.012	-.020	.090	.031	-.009	-.015
	(.204)	(.069)	(.031)	(.030)	(.152)	(.051)	(.024)	(.023)
Group 4	-.116	-.040	.012	.020	-.092	-.032	.009	.015
	(.162)	(.056)	(.024)	(.028)	(.125)	(.043)	(.018)	(.022)

3.3 The rat growth data

The second data set, which was taken from Box (1950), concerns the prediction of growth in rats. In predicting the weight change in rats, it is essential to take into account differences in their initial weight. We thus include the initial weight of rats as a covariate in the analysis, and eliminate its effect in assessing weekly gains in weight under different drug treatments. The study involved 27 rats randomly assigned to one of the following three treatment groups: Group 1 - Control group (10 rats), Group 2 - Thyroxin treatment (7 rats), Group 3 - Thiouracil treatment (10 rats). Each rat was weighed after one, two, three and four weeks (4 time points). The design matrix **H** for time consists of up to the second order orthogonal polynomials with a constant term. The criterion variables (**Y**) were columnwise centered, and the predictor variables (**X**) indicating drug treatments and the design matrix **H** were normalized before the analysis.

A mixture of the GMANOVA and MANOVA models was applied to the data set. The results of permutation tests with varying λ and ρ (0, .5, 1, and 5) showed that the dimensionality was consistently one. Results of 24-fold cross

validation are given in Table 4. This table indicates that the prediction error takes the minimum value of .648 at $\lambda = 1$ and $\rho = 0$, which is slightly better than the value of .655 for the non-regularized case. (The effect of regularization seems much smaller in this example than in the previous one.) Table 5 compares regularized and non-regularized estimates in a manner similar to Table 3. Results similar to Table 3 are also observed in Table 5, although the effect of regularization is not as remarkable as in the former. We also tried to assess the effect of the covariate in prediction. Specifically, we compared the mixture model of GMANOVA and MANOVA and the reduced rank GCM in their predictability. The result showed that the cross validated normalized prediction error of .515 for the mixture model was smaller than that of .539 for the reduced rank GCM, indicating that more stable predictions of rat growth could be made with the former.

Table 4
K-fold Cross Validation Results for Box's (1950) data.

λ	ρ	Prediction Error
0	0	.655
.5	0	.650
1	0	.648*
5	0	.681
0	.5	.652
.5	.5	.652
1	.5	.655
5	.5	.702

*Optimal combination

Table 5
Comparison of estimated regression coefficients **B** from Box's (1950) data.

	Non-Regularized ($\lambda = 0$ and $\rho = 0$)			Regularized ($\lambda = 1$ and $\rho = 0$)		
	Const.	Linear	Quad.	Const.	Linear	Quad.
Group 1	**2.898	**1.548	-0.651	**2.646	**1.409	-.584
	(1.165)	(.585)	(.417)	(1.034)	(.516)	(.367)
Group 2	*4.228	*2.259	-.950	*3.695	*1.967	-.815
	(2.144)	(1.207)	(.811)	(1.812)	(1.026)	(.696)
Group 3	**-5.857	**-3.130	*1.316	**-5.334	**-2.839	1.176
	(.941)	(.799)	(.776)	(.858)	(.727)	(.693)

4 Concluding Remarks

In this paper, we proposed a ridge type of regularized estimation method for the reduced rank GCM to deal with the problem of small sample size in biomedical studies. We have shown that the reduced rank ridge LS (RLS) estimates of regression coefficients can be obtained in closed form, given fixed values of the regularization parameters λ and ρ. The best dimensionality of the

solution and the optimal values of ridge parameters were determined, respectively, by permutation tests and the cross validation method. Furthermore, we extended the RLS estimation to a mixture of the GMANOVA and MANOVA models.

We have demonstrated the usefulness of the RLS estimation through examples. In the Monte Carlo study, we reported that the RLS estimation could provide better estimates than the OLS counterparts in terms of achieving smaller MSE's (mean square errors), particularly in small samples. In two empirical examples, we have shown that the RLS estimates are more stable than the non-regularized estimates in the reduced rank GCM as well as in the mixture of reduced rank GMANOVA (GCM) and MANOVA models. In sum, the proposed method can serve as an efficient tool for handling potential problems with data with small sample sizes.

For further extensions of the proposed method, we may consider analogous extensions of the maximum likelihood (ML) estimator in GCM. The rank free ML estimate of the regression coefficients in model (1) is given by (e.g., Grizzle and Allen, 1969)

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{H}(\mathbf{H}'\mathbf{S}^{-1}\mathbf{H})^{-1}, \tag{47}$$

where $\mathbf{S} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$, under the distributional assumption that $\mathbf{E} \sim \mathrm{MN}(\mathbf{0}, \Sigma \otimes \mathbf{I})$. There are at least a couple of ways to incorporate a ridge type of regularization in this estimate. First of all, $\mathbf{S}$ can be regularized by $\mathbf{S}+\rho_1\mathbf{I}$. Secondly, $\mathbf{H}'\mathbf{S}^{-1}\mathbf{H}$ can be regularized by $\mathbf{H}'\mathbf{S}^{-1}\mathbf{H}+\rho_2\mathbf{I}$. Thirdly, these two procedures can be combined into one. In somewhat different contexts, Jung and Takane (2009) have successfully demonstrated the usefulness of this type of regularized estimation.

We might also consider a similar regularization method for extended growth curve models (EGCM; Vervyla and Venables, 1988; Tian and Takane, 2009). A basic form of EGCM posits that

$$\mathbf{Y} = \sum_{j=1}^{J} \mathbf{X}_i\mathbf{B}_i\mathbf{H}_i' + \mathbf{E} \tag{48}$$

where J is generally greater than 1. A closed form solution exists for both ML and LS estimation, if no rank restrictions are imposed on $\mathbf{B}_j$, although some form of iterative algorithm, like the one used in Takane, Kiers, and de Leeuw (1995), is necessary when the reduced rank restriction is imposed. In any case, it is relatively straightforward to incorporate the regularized estimation into these procedures.

Another possible area of extension involves hierarchical linear models (HLM)

(e.g., Croon and van Veldhoven 2007). In all of these, it may also be worthwhile to consider a more generalized form of ridge regression that incorporates $\lambda\mathbf{L}$ (instead of $\lambda\mathbf{P}_X^{'}$) as the regularization term, where $\mathbf{L}$ is a q by q nonnegative-definite (nnd) matrix such that $\mathrm{Sp}(\mathbf{L}) = \mathrm{Sp}(\mathbf{X}')$. This generalized form of ridge regression is useful for incorporating more complicated forms of regularization such as smoothness (Ramsay and Silverman, 2005).

5 Appendix

5.1 (A): Proof of the equivalence between (25) and (26)

$\phi(\mathbf{B}_1, \mathbf{B}_2^*)$ in (25) can be expanded as

$$\begin{aligned}\phi(\mathbf{B}_1, \mathbf{B}_2^*) = \mathrm{SS}((&\mathbf{Y} - \mathbf{Q}_{X_2}\mathbf{X}_1\hat{\mathbf{B}}_1\mathbf{H}_1' - \mathbf{X}_2\hat{\mathbf{B}}_2^*) \\ &+\mathbf{Q}_{X_2}\mathbf{X}_1(\hat{\mathbf{B}}_1 - \mathbf{B}_1)\mathbf{H}_1' + \mathbf{X}_2(\hat{\mathbf{B}}_2^* - \mathbf{B}_2^*)).\end{aligned} \tag{49}$$

It can be easily verified that the three terms in SS are mutually orthogonal, so that $\phi(\mathbf{B}_1, \mathbf{B}_2^*)$ can be further rewritten as

$$\begin{aligned}\phi(\mathbf{B}_1, \mathbf{B}_2^*) = &\mathrm{SS}(\mathbf{Y} - \mathbf{Q}_{X_2}\mathbf{X}_1\hat{\mathbf{B}}_1\mathbf{H}_1' - \mathbf{X}_2\hat{\mathbf{B}}_2^*) \\ &+ \mathrm{SS}(\hat{\mathbf{B}}_1 - \mathbf{B}_1)_{X_1'Q_{X_2}X_1,\, H_1'H_1} + \mathrm{SS}(\hat{\mathbf{B}}_2^* - \mathbf{B}_2^*)_{X_2'X_2,\, I},\end{aligned} \tag{50}$$

where the first SS reduces to

$$\begin{aligned}&\mathrm{SS}(\mathbf{Y} - \mathbf{Q}_{X_2}\mathbf{X}_1\hat{\mathbf{B}}_1\mathbf{H}_1' - \mathbf{X}_2\hat{\mathbf{B}}_2^*) = \\ &\qquad \mathrm{SS}(\mathbf{Y}) - \mathrm{SS}(\mathbf{Y})_{P_{Q_{X_2}X_1},\, P_{H_1}} - \mathrm{SS}(\mathbf{Y})_{P_{X_2},\, I}.\end{aligned} \tag{51}$$

Here, $\hat{\mathbf{B}}_1 = (\mathbf{X}_1'\mathbf{Q}_{X_2}\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{Q}_{X_2}\mathbf{Y}\mathbf{H}_1(\mathbf{H}_1'\mathbf{H}_1)^-$, and $\hat{\mathbf{B}}_2^* = (\mathbf{X}_2'\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{Y}$, as defined in (27) and (29), respectively.

5.2 (B): Proof of the equivalence between (38) and (44)

Criterion (38) can be expanded as

$$\begin{aligned}&\phi_{\lambda,\rho}(\mathbf{B}_1, \mathbf{B}_2^*) = \\ &\mathrm{tr}[(\mathbf{Y}'\mathbf{Y} + \mathbf{H}_1\mathbf{B}_1'\mathbf{X}_1'\mathbf{S}_{X_2}(\lambda)^2\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' + \mathbf{B}_2^{*\prime}\mathbf{X}_2'\mathbf{X}_2\mathbf{B}_2^* \\ &- 2\mathbf{H}_1\mathbf{B}_1'\mathbf{X}_1'\mathbf{S}_{X_2}(\lambda)\mathbf{Y} - 2\mathbf{B}_2^{*\prime}\mathbf{X}_2'\mathbf{Y} + 2\mathbf{B}_2^{*\prime}\mathbf{X}_2'\mathbf{S}_{X_2}(\lambda)\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1'\end{aligned}$$

$$
\begin{aligned}
&+ \lambda(\mathbf{H}_1\mathbf{B}_1'\mathbf{B}_1\mathbf{H}_1' + \mathbf{H}_1\mathbf{B}_1'\mathbf{X}_1'\mathbf{X}_2\mathbf{G}^-\mathbf{P}_{X_2'}\mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' \\
&- \mathbf{H}_1\mathbf{B}_1'\mathbf{X}_1'\mathbf{G}^-\mathbf{B}_2^* - \mathbf{B}_2^{*\prime}\mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1\mathbf{B}_1\mathbf{H}_1' \\
&+ \mathbf{B}_2^{*\prime}\mathbf{B}_2^*) + \rho(\mathbf{B}_1'\mathbf{X}_1'\mathbf{S}_{X_2}(\lambda)^2\mathbf{X}_1\mathbf{B}_1\mathbf{P}_{H_1'}) \\
&+ \lambda\rho(\mathbf{B}_1'\mathbf{B}_1 + \mathbf{B}_1'\mathbf{X}_1'\mathbf{X}_2\mathbf{G}^-\mathbf{P}_{X_2'}\mathbf{G}^-\mathbf{X}_2'\mathbf{X}_1\mathbf{B}_1\mathbf{P}_{H_1'})]. \qquad (52)
\end{aligned}
$$

Note that the sixth, ninth, and tenth terms in the trace (tr) on the right hand side of the above equation add up to 0, the third and eleventh terms add up to

$$
\mathrm{tr}(\mathbf{B}_2^{*\prime}\mathbf{G}\mathbf{B}_2^*), \qquad (53)
$$

and the second, seventh, eighth, twelfth, thirteenth, and fourteenth terms add up to

$$
\mathrm{tr}(\mathbf{B}_1'\mathbf{A}\mathbf{B}_1\mathbf{H}_1'\mathbf{M}_{H_1}(\rho)\mathbf{H}_1). \qquad (54)
$$

This may be seen as follows: First of all, the second, seventh, and eighth terms add up to

$$
\mathrm{tr}(\mathbf{B}_1'\mathbf{A}\mathbf{B}_1\mathbf{H}_1'\mathbf{H}_1), \qquad (55)
$$

and the twelfth and fourteenth terms add up to

$$
\mathrm{tr}(\mathbf{B}_1'\mathbf{A}\mathbf{B}_1\mathbf{P}_{H_1'}), \qquad (56)
$$

so that (55), (56), and the thirteenth term in (52) add up to (54).

On the other hand, (44) can be expanded as

$$
\begin{aligned}
&\phi_{\lambda,\rho}(\mathbf{B}_1, \mathbf{B}_2^*) = \\
&\mathrm{tr}[\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{R}_{S_{X_2}(\lambda)X_1}(\lambda)\mathbf{Y}\mathbf{R}_{H_1} - \mathbf{Y}'\mathbf{P}_{X_2}(\lambda)\mathbf{Y} \\
&+ \mathbf{H}_1'\hat{\mathbf{B}}_1(\lambda,\rho)'\mathbf{A}\hat{\mathbf{B}}_1(\lambda,\rho) + \mathbf{B}_1'\mathbf{A}\mathbf{B}_1\mathbf{H}_1'\mathbf{M}_{H_1}(\rho)\mathbf{H}_1 \\
&- 2\mathbf{H}_1'\mathbf{B}_1'\mathbf{A}\hat{\mathbf{B}}_1(\lambda,\rho) + \hat{\mathbf{B}}_2^*(\lambda)'\mathbf{G}\hat{\mathbf{B}}_2^*(\lambda) \\
&+ \mathbf{B}_2^{*\prime}\mathbf{G}\mathbf{B}_2^* - 2\mathbf{B}_2^{*\prime}\mathbf{G}\hat{\mathbf{B}}_2^*(\lambda)]. \qquad (57)
\end{aligned}
$$

As noted earlier, $\mathbf{A} = \mathbf{X}_1'\mathbf{S}_{X_2}(\lambda)\mathbf{X}_1 + \lambda\mathbf{P}_{X_1'}$. Note that the second and fourth terms in the trace on the right hand side of the above equation cancel out, and so do the third and seventh terms. Observe that the first term on the right hand side of (57) is equal to the first term on the right hand side of (52), the sixth term in (57) to the fourth term in (52), the ninth term in (57) to the fifth term in (52), the eighth term in (57) to the sum of the third and

eleventh terms in (52), and the fifth term in (57) is equal to (54), establishing the equivalence between (38) and (44).

6 Computer Software

Matlab programs that carried out the computations reported in the paper are available upon request from the second author.

7 Acknowledgement

The work reported in this paper is supported by Grants 10603 and 290439 from the Natural Sciences and Engineering Research Council of Canada to the first and the third authors, respectively. Correspondence regarding this article should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC H3A 1B1 Canada.

References

Boik, R.J. 1987. The Fisher-Pitman permutation test: A nonrobust alternative to the normal theory F test when variances are heterogeneous. Brit J Math Stat Psy, 40, 26-42.

Box, G.E.P. 1950. Problems in the analysis of growth and wear curves. Biometrics, 6, 362-389.

Carroll, J.D., Pruzansky, S., Kruskal, J.B., 1980. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. Psychometrika, 45, 3-24.

Chinchilli, V.M., Elswick, R.K., 1985. A mixture of the MANOVA and GMANOVA models. Commun Stat-Theor M, 14, 3075-3089.

Croon, M.A., van Veldhoven, M.J.P.M., 2007. Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. Psychol Methods, 12, 45-57.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap, CRC Press, Boca Raton, Florida.

Good, P. I. 2005. Permutation, Parametric, and Bootstrap Tests of Hypothses (Third Edition), Springer, New York.

Greenacre, M.J. 1984. Theory and Applications of Correspondence Analysis, Academic Press, London.

Grizzle, L.E., Allen, D.M., 1969. Analysis of growth and dose response curves. Biometrics, 25, 357-381.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Leaning: Data Mining, Inference, and Prediction, Springer, New York.
Hoerl, K.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.
Hwang, H., 2009. Regularized generalized structured component analysis. Psychometrika, 74, 517-530.
Jung, S., Takane, Y., 2009. Regularized structural equation models with latent variables. Submitted for publication.
Legendre, P., Legendre, L., 1998. Numerical Ecology, (Second English edition), Elsevier, Oxford.
Potthof, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis (Second Edition), Springer, New York.
Reinsel, G.C., Velu, R.P., 1998. Multivariate Reduced-rank Regression: Theory and Applications, Springer, New York.
Reinsel, G.C., Velu, R.P., 2003. Reduced-rank growth curve models. J Stat Plan Infer, 114, 107-129.
Romano, J.P. 1990. On the behavior of randomization tests without a group invariance assumption. J Am Stat Assoc, 85, 686-692.
Takane, Y., 2003. Relationships among various kinds of eigenvalue and singular value decompositions. In H. Yanai, A, Okada, K. Shigemasu, Y. Kano, and J. Meulman, (Eds.), New Developments in Psychometrics (pp. 45-56). Springer, Tokyo.
Takane, Y., Hunter, M.A., 2001. Constrained principal component analysis: A comprehensive theory. Appl Algebr Eng Comm, 12, 391-419.
Takane, Y., Hwang, H., 2002. Generalized constrained canonical correlation analysis. Multivar Behav Res, 37, 163-195.
Takane, Y., Hwang, H., 2007. Regularized linear and kernel redundancy analysis. Comput Stat Data An, 52, 394-405.
Takane, Y., Hwang, H., Abdi, H., 2008. Regularized multiple-set canonical correlation analysis. Psychometrika, 73, 753-775.
Takane, Y., Jung, S., 2008. Regularized partial and/or constrained redundancy analysis. Psychometika, 73, 671-690.
Takane, Y., Jung, S., 2009. Regularized nonsymmetric correspondence analysis. Comput Stat Data An, 53, 3159-3170.
Takane, Y., Kiers, H.A.L., de Leeuw, J., 1995. Component analysis with different constraints on different dimensions. Psychometrika, 60, 259-280.
Takane, Y., Shibayama, T., 1991. Principal component analysis with external information on both subjects and variables. Psychometrika, 56, 97-120.
Takane, Y., Yanai, H., 2008. On ridge operators. Linear Algebra Appl, 428, 1778-1790.
ten Berge, J.M.F. 1993. Least Squares Optimization. DSWO Press, Leiden,

The Netherlands.
ter Braak, C.J.I., Šmilauer, P., 1998. CANOCO Reference Manual and User's Guide to Canoco for Windows. Microcomputer Power, Ithaka, N.Y.
Tian, Y., Takane, Y., 2009. On consistency, natural restrictions and estimability under classical and extended growth curve models. J Stat Plan Infer, 139, 2445-2458.
Van den Wollenberg, A.L., 1977. Redundancy analysis: An alternative for canonical analysis. Psychometrika, 42, 207-219.
van der Leeden, R., 1990. Reduced Rank Regression with Structured Residuals. DSWO Press, Leiden, The Netherlands.
Verbyla, A.P., Venables, W.N., 1988. An extension of the growth curve model. J Multivariate Anal, 31, 187-200.