Rank Estimation in Reduced-Rank Regression

Efstathia Bura

Department of Statistics, The George Washington University
Washington, DC 20052

E-mail: ebura@gwu.edu

and

R. Dennis Cook

School of Statistics, University of Minnesota, St. Paul, MN 55108

E-mail: dennis@stat.umn.edu

Reduced rank regression assumes that the coefficient matrix in a multivariate regression model is not of full rank. The unknown rank is traditionally estimated under the assumption of normal responses. We derive an asymptotic test for the rank that requires no distributional assumptions on the response vector. The test is extended to the non-constant covariance case. Linear combinations of the components of the predictor vector that are estimated to be significant for modelling the responses are obtained.

Key Words: asymptotic test; chi-squared; weighted chi-squared.

1. INTRODUCTION

Let $\mathbf{Y} = (y_j)_{j=1,\dots,m} \in \mathbb{R}^m$ indicate a response vector and let $\mathbf{X} = (x_j)_{j=1,\dots,p} \in \mathbb{R}^p$ indicate a vector of nonconstant predictors. The classical multivariate regression model is of the form

$$\mathbf{Y}^T|\mathbf{X} = \mathbf{A} + \mathbf{X}^T\mathbf{B} + \mathbf{E}^T \tag{1}$$

where $\mathbf{A}$ is a $1 \times m$ vector of intercepts, $\mathbf{B}$ is a $p \times m$ regression coefficient matrix and $\mathbf{E}$ is the $m \times 1$ random error vector with $\mathrm{E}(\mathbf{E}) = 0$ and positive definite $\mathrm{Cov}(\mathbf{E}) = \mathbf{\Sigma}$ of order $m \times m$. When a random sample of size n is available on $(\mathbf{Y}, \mathbf{X})$ the model can be written in the form

$$\mathbf{Y}_n = \mathbf{1}_n \mathbf{A} + \mathbf{X}_n \mathbf{B} + \mathbf{E}_n \tag{2}$$

where $\mathbf{Y}_n = (y_{i1}, y_{i2}, \ldots, y_{im})_{i=1,\ldots,n}$ is an $n \times m$ matrix of responses, $\mathbf{1}_n$ is an $n \times 1$ vector of ones, $\mathbf{X}_n = (x_{i1} - \bar{x}_1, x_{i2} - \bar{x}_2, \ldots, x_{ip} - \bar{x}_p)_{i=1,\ldots,n}$ is an $n \times p$ full rank matrix of centered predictors, $\mathbf{B}$ is the $p \times m$ coefficient matrix defined in (1), and $\mathbf{A}$ is again a $1 \times m$ intercept vector. The error matrix $\mathbf{E}_n$ is assumed to satisfy

$$\mathrm{E}(\mathbf{E}_n) = 0, \quad \mathrm{Cov}(\mathrm{vec}\, \mathbf{E}_n) = \mathbf{\Sigma} \otimes \mathbf{I}_n \tag{3}$$

where $\mathbf{\Sigma}$ does not depend on $\mathbf{X}$, and $\mathrm{vec}(\mathbf{E}_n)$ is the vector produced by concatenating the columns of the error matrix $\mathbf{E}_n$. The symbol $\otimes$ denotes the Kronecker product. No additional distributional assumptions on the errors are made.

Let $\mathcal{S} = \mathrm{span}(\mathbf{B}^T \mathbf{X}_n^T)$ denote the subspace spanned by the rows of $\mathbf{X}_n \mathbf{B}$. This subspace represents the fewest linear combinations of $\mathbf{X}$ that are needed for the regression. If $\mathbf{b}$ is a known matrix whose columns form a basis for $\mathcal{S}$ then we can replace $\mathbf{X}$ with $\mathbf{b}^T \mathbf{X}$ without loss of information on the regression. The dimension of $\mathcal{S}$, the minimum number of reduced predictors, is equal to the rank of $\mathbf{B}$: Since $\mathrm{rank}(\mathbf{B}^T \mathbf{X}_n^T) = \mathrm{rank}(\mathbf{X}_n \mathbf{B})$,

$$\dim(\mathcal{S}) = \mathrm{rank}(\mathbf{X}_n \mathbf{B}) = \mathrm{rank}(\mathbf{B}^T \mathbf{X}_n^T \mathbf{X}_n \mathbf{B}) = \mathrm{rank}(\mathbf{B}) \tag{4}$$

because $\mathbf{X}_n^T \mathbf{X}_n$ is a positive definite matrix (see *A4.4*, Seber, 1977). In consequence, inference on the dimension of $\mathcal{S}$ can be based solely on $\mathbf{B}$ in the sense that an estimate of the rank of $\mathbf{B}$ constitutes an estimate of the dimension of $\mathcal{S}$.

When the dimension of $\mathcal{S}$ is less than $\min(p, m)$, model (2) corresponds to the basic reduced-rank regression model (see Reinsel and Velu, 1998, ch. 2). Reduced-rank regression models were introduced by Anderson (1951) and are used mostly when there is a need to reduce the number of parameters in (2). They have a wide spectrum of applications in fields such as chemometrics, econometrics, psychometrics and engineering. The typical analysis of a reduced rank regression model is based on the assumption that the coefficient matrix $\mathbf{B}$ is not of full rank. The elements of $\mathbf{B}$ are subsequently estimated for a given value of the rank of $\mathbf{B}$. Asymptotic distributions of estimators of the coefficient matrix and asymptotic rank tests are available when the error terms, or equivalently, the responses are normal (Anderson, 1951; Izenman, 1975; Reinsel and Velu, 1998; Schmidli, 1996). Anderson (1999a) computed the asymptotic distribution of the estimated reduced-rank coefficient matrix without requiring the normality condition on the response vector.

We develop a theory for estimating the rank of a regression model and consequently estimating the subspace of the linear combinations of $\mathbf{X}$ components that span the regression subspace $\mathcal{S}$ without imposing distributional assumptions on either the response or the predictor vector. We also extend our results to accommodate the case of a priori known constraints on $\mathbf{B}$ and for non-constant variances.

In Section 2, an asymptotic chi-squared test for the rank of the coefficient matrix is derived. In Section 3 we consider the special case where additional information on the shape of the response curves is available. In this case, a weighted chi-squared test for the rank of the regression is derived in Section 3.1. The nonconstant covariance structure case is considered in Section 4 where the results of Sections 2 and 3 are extended. A concluding discussion is presented in Section 5. The appendix contains the lengthier proofs and auxiliary results.

2. A CHI-SQUARED ASYMPTOTIC RANK TEST

The key components of model (2) are

$$\mathbf{Y}_n = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}, \quad \mathbf{X}_n = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pm} \end{bmatrix}$$

with the error matrix satisfying (3). To estimate the rank (*dimension*) of the model, or equivalently, the rank of $\mathbf{B}$, we estimate the coefficient matrix by ordinary least squares. Let $\hat{\mathbf{B}}_n = (\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{X}_n^T\mathbf{Y}_n$ be the OLSE for $\mathbf{B}$. Let $\mathbf{H}_n = \mathbf{\Sigma} \otimes (\mathbf{X}_n^T\mathbf{X}_n/n)^{-1}$ denote the covariance matrix of $\sqrt{n}\,\mathrm{vec}(\hat{\mathbf{B}}_n - \mathbf{B})$. If $\mathbf{H}_n$ has a positive definite limit matrix $\mathbf{H}$, then

$$\sqrt{n}\ \mathrm{vec}(\hat{\mathbf{B}}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{H}) \tag{5}$$

provided certain regularity conditions are satisfied (see lemma A.1 in the appendix).

Assume $\mathbf{G}_n = (\mathbf{X}_n^T\mathbf{X}_n/n)^{-1}$ has a $p \times p$ positive definite limit matrix $\mathbf{G}$. Also, assume that a consistent estimate $\hat{\mathbf{\Sigma}}_n$ is available, as $\mathbf{\Sigma}$ is usually unknown. For example,

$$\hat{\mathbf{\Sigma}}_n = (n-p-1)^{-1}(\mathbf{Y}_n - \mathbf{1}_n\hat{\mathbf{A}}_n - \mathbf{X}_n\hat{\mathbf{B}}_n)^T(\mathbf{Y}_n - \mathbf{1}_n\hat{\mathbf{A}}_n - \mathbf{X}_n\hat{\mathbf{B}}_n) \tag{6}$$

is consistent and unbiased for $\mathbf{\Sigma}$, where $\hat{\mathbf{A}}_n$ is the OLSE of $\mathbf{A}$. Let

$$\hat{\mathbf{H}}_n = \hat{\mathbf{\Sigma}}_n \otimes (\mathbf{X}_n^T\mathbf{X}_n/n)^{-1} \tag{7}$$

Then,

$$\hat{\mathbf{H}}_n \xrightarrow[n\to\infty]{} \mathbf{H} \qquad \text{in probability} \tag{8}$$

The convergence in (8) is a direct application of the triangle inequality and the fact that continuous functions of consistent estimates are themselves consistent. The remarks above in conjunction with direct application of the multivariate version of Slutsky's theorem (see [A 4.19] in Bunke and Bunke (1986)), and (5) give

$$\sqrt{n}\,\hat{\mathbf{H}}_n^{-1/2}\,\mathrm{vec}(\hat{\mathbf{B}}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_{mp}) = N(0, \mathbf{I}_m \otimes \mathbf{I}_p) \tag{9}$$

Let $d = \dim(\mathcal{S})$. We have shown that $d = \mathrm{rank}(\mathbf{B})$ and thus, since $\mathrm{rank}(\mathbf{B}) = \mathrm{rank}(\mathbf{G}^{-1/2}\mathbf{B}\boldsymbol{\Sigma}^{-1/2})$, we use the standardized matrix

$$\hat{\mathbf{B}}_{\mathrm{std}} = \mathbf{G}_n^{-1/2}\hat{\mathbf{B}}_n\hat{\boldsymbol{\Sigma}}_n^{-1/2}$$

to estimate d. A test statistic Λ_d for $d = \mathrm{rank}(\mathbf{B}) = \dim(\mathcal{S})$ is given in Theorem 2.1.

THEOREM 2.1. *Assume that model (2) holds, that $\mathbf{G}_n$ converges pointwise to a positive definite limit, and that $\hat{\boldsymbol{\Sigma}}_n$ is a consistent estimate of $\boldsymbol{\Sigma}$. Let $\phi_1 \geq \phi_2 \geq \ldots \geq \phi_{\min(p,m)}$ be the ordered singular values of $\hat{\mathbf{B}}_{\mathrm{std}}$. Then*

$$\Lambda_d = n \sum_{j=d+1}^{\min(p,m)} \phi_j^2 \tag{10}$$

is asymptotically distributed as a $\chi^2_{(p-d)(m-d)}$ random variable.

Proof. Consider the singular value decomposition of $\mathbf{G}^{-1/2}\mathbf{B}\boldsymbol{\Sigma}^{-1/2}$,

$$\mathbf{G}^{-1/2}\mathbf{B}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Gamma}_1^T \begin{bmatrix} \boldsymbol{\Delta} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\Gamma}_2^T$$

where $\mathbf{\Delta}$ is a $d \times d$ diagonal matrix of positive singular values. Partition $\mathbf{\Gamma}_1^T = (\mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}) : p \times p$, $\mathbf{\Gamma}_{11} : p \times d$, $\mathbf{\Gamma}_{12} : p \times (p-d)$, $\mathbf{\Gamma}_2^T = (\mathbf{\Gamma}_{21}, \mathbf{\Gamma}_{22})^T : m \times m$, where $\mathbf{\Gamma}_{21}^T : d \times m$, $\mathbf{\Gamma}_{22}^T : (m-d) \times m$. By the Eaton-Tyler result (Eaton and Tyler, 1994) on the asymptotic distribution of the singular values of a random matrix, the limiting distribution of the smallest $\min(p-d, m-d)$ singular values of

$$\sqrt{n}\ (\mathbf{G}_n^{-1/2}\hat{\mathbf{B}}_n\hat{\mathbf{\Sigma}}_n^{-1/2})$$

is the same as the limiting distribution of the singular values of the $(p-d) \times (m-d)$ matrix

$$\sqrt{n}\ \mathbf{B}_n = \sqrt{n}\ (\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\hat{\mathbf{B}}_n\hat{\mathbf{\Sigma}}_n^{-1/2}\mathbf{\Gamma}_{22})$$

By (9) we have

$$\sqrt{n}\ \operatorname{vec}(\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\hat{\mathbf{B}}_n\hat{\mathbf{\Sigma}}_n^{-1/2}\mathbf{\Gamma}_{22}) \xrightarrow{\mathcal{D}} N_{(p-d)(m-d)}(0, \mathbf{I}_{m-d} \otimes \mathbf{I}_{p-d}) \quad (11)$$

Consequently, Λ_d has the same asymptotic distribution as the sum of the squares of the singular values of $\sqrt{n}\ (\mathbf{\Gamma}_{12}^T\mathbf{G}_n^{-1/2}\hat{\mathbf{B}}_n\hat{\mathbf{\Sigma}}_n^{-1/2}\mathbf{\Gamma}_{22})$ which is $\chi^2_{(m-d)(p-d)}$ by (11). ∎

We use Λ_k as a test statistic for the rank of $\mathbf{B}$. For example, to test the hypothesis that $d = 1$, compare Λ_1 to the percentage points of a chi-squared distribution with $(p-1)(m-1)$ degrees of freedom. As an aside, it is easy to see that the asymptotic test in Theorem 2.1 coincides with the usual F-test for testing $d = 0$; that is, that all the coefficients are zero, when $m = 1$.

The d left singular vectors $\mathbf{u}_j$ of $\hat{\mathbf{B}}_n$ that correspond to its d largest singular values provide an estimated basis for $\mathcal{S}$. The linear combinations of $\mathbf{X}$ needed for the regression are then constructed as $\mathbf{X}_n\mathbf{u}_j$, $j = 1, \ldots, d$.

3. REDUCED RANK REGRESSION WITH RESTRICTIONS

In the presence of additional information, or guided by scatterplots of components of $\mathbf{Y}$ versus components of $\mathbf{X}$, the reduced-rank regression model can be formulated by allowing linear restrictions on the coefficient matrix $\mathbf{B}$ in model (2), as follows: Let $\boldsymbol{\beta} = \text{vec}(\mathbf{B}) \in \mathbb{R}^{pm}$. Assume that

$$\mathbf{D}\boldsymbol{\beta} = 0 \tag{12}$$

for some matrix $\mathbf{D}$ of zeroes and ones which has order $r \times pm$ and rank r. Each row and each column of $\mathbf{D}$ contains exactly one 1. The rank r of $\mathbf{D}$ equals the number of the elements of $\mathbf{B}$ set equal to zero. The sample regressor matrix $\mathbf{X}_n$ is assumed to be of full rank p. The model is now given by (2) subject to (12). This setup can be used to handle other situations as well. For example, models in which an element of $\boldsymbol{\beta}$ is restricted to be a nonzero constant can be handled by offsetting the response and reducing the restriction to the form in (12).

The constraints imposed on $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ allow us to set any element of $\mathbf{B}$ equal to zero. It should be noted that it makes no sense to set any row of the matrix of parameters $\mathbf{B}$ equal to zero since this is equivalent to eliminating the corresponding element of $\mathbf{X}$. Also, the same applies to setting a $\mathbf{B}$-column equal to zero, since this is equivalent to dropping the corresponding $\mathbf{Y}$-element from the model. Furthermore, the matrix that results from $\mathbf{B}$ after setting r of its entries equal to zero may not be full rank. This can be easy to determine by inspection, in which case we have a *deterministic dimension reduction* of the regression. In what follows, we assume that there is no such deterministic reduction of the rank of the constrained matrix.

In this formulation, $\text{span}(\mathbf{B}^T\mathbf{X}_n^T) = \mathcal{S}$, provided $\mathbf{D}\,\text{vec}(\mathbf{B}) = \mathbf{D}\boldsymbol{\beta} = 0$. As in Section 2, inference on the dimension of $\mathcal{S}$ can be based on an estimate of the rank of $\mathbf{B}$ under the condition $\mathbf{D}\boldsymbol{\beta} = 0$. Since the constraint is placed on $\boldsymbol{\beta}$, we consider model (2) in its vector format

$$\operatorname{vec}(\mathbf{Y}_n) = (\mathbf{I}_m \otimes \mathbf{1}_n)\operatorname{vec}(\mathbf{A}) + (\mathbf{I}_m \otimes \mathbf{X}_n)\boldsymbol{\beta} + \operatorname{vec}(\mathbf{E}_n) \tag{13}$$

subject to (12).

Suppose that $\boldsymbol{\beta}_1 \in \mathbb{R}^{mp}$ is such that $\mathbf{D}\boldsymbol{\beta}_1 = 0$. Let $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \ldots, \beta_{1,mp})^T$. Also, let

$$\mathbf{B}_1 = \begin{bmatrix} \beta_{1,1} & \beta_{1,p+1} & \cdots & \beta_{1,(m-1)p+1} \\ \beta_{1,2} & \beta_{1,p+2} & \cdots & \beta_{1,(m-1)p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1,p} & \beta_{1,2p} & \cdots & \beta_{1,pm} \end{bmatrix} \tag{14}$$

Obviously $\operatorname{rank}(\mathbf{B}_1) = d$. Let $\hat{\boldsymbol{\beta}}_1$ be the OLS estimate of $\boldsymbol{\beta}_1$. Also, let $\hat{\mathbf{B}}_1$ be the $p \times m$ matrix produced by the pm-vector $\hat{\boldsymbol{\beta}}_1$ as $\mathbf{B}_1$ was created from $\boldsymbol{\beta}_1$ in (14). The asymptotic distribution of $\hat{\mathbf{B}}_1$ will be used to estimate the rank of $\mathbf{B}_1$. Let

$$\mathbf{H}_n^\star = \operatorname{Cov}(\operatorname{vec}(\hat{\mathbf{B}}_1 - \mathbf{B}_1)) = \operatorname{Cov}(\hat{\boldsymbol{\beta}}_1)$$

Recall that $\boldsymbol{\beta}_1$ has r zero entries, so $\hat{\boldsymbol{\beta}}_1$ has r zero entries as well. Therefore, $\mathbf{H}_n^\star$ is a positive *semi-definite* $pm \times pm$ matrix with r zero columns and rows that correspond to the r zero elements of $\hat{\boldsymbol{\beta}}_1$. Without loss of generality, we can write

$$\mathbf{H}_n^\star = \begin{bmatrix} \mathbf{C}_n & 0 \\ 0 & 0 \end{bmatrix} \tag{15}$$

where $\mathbf{C}_n$ is the $(pm - r) \times (pm - r)$ positive definite covariance matrix of the non-zero $\hat{\boldsymbol{\beta}}_1$ elements. The constrained model (13) can be reduced to an unconstrained one of the form

$$\operatorname{vec}(\mathbf{Y}_n) = (\mathbf{I}_m \otimes \mathbf{1}_n)\operatorname{vec}(\mathbf{A}) + \tilde{\mathbf{X}}_n\boldsymbol{\beta}_1^* + \operatorname{vec}(\mathbf{E}_n) \tag{16}$$

where $\tilde{\mathbf{X}}_n$ is a $mn \times (mp - r)$ full rank matrix and the parameter vector $\boldsymbol{\beta}_1^* \in \mathbb{R}^{mp-r}$ is completely unrestricted.

Accordingly, the OLS estimate of $\boldsymbol{\beta}_1^*$ is given by

$$\hat{\boldsymbol{\beta}}_1^* = (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{X}}_n)^{-1} \tilde{\mathbf{X}}_n^T \operatorname{vec}(\mathbf{Y}_n) \tag{17}$$

Note that the elements of $\hat{\boldsymbol{\beta}}_1^*$ are equal to the corresponding non-zero elements of $\hat{\boldsymbol{\beta}}_1$. But now, the covariance matrix $\mathbf{V}_n$ of $\sqrt{n}\ (\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*)$ is the $(mp - r) \times (mp - r)$ positive definite matrix given by

$$\mathbf{V}_n = n(\tilde{\mathbf{X}}_n^T \tilde{\mathbf{X}}_n)^{-1} \tilde{\mathbf{X}}_n^T (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) \tilde{\mathbf{X}}_n (\tilde{\mathbf{X}}_n^T \tilde{\mathbf{X}}_n)^{-1} \tag{18}$$

A straightforward application of Lemma A.1 in the appendix obtains that

$$\sqrt{n}\ (\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*) \xrightarrow{\mathcal{D}} N(0, \mathbf{V}) \tag{19}$$

provided $\mathbf{V}_n$ has a positive definite limit matrix $\mathbf{V}$ and the conditions of Lemma A.1 are satisfied.

When $\boldsymbol{\Sigma}$ is known, $\mathbf{V}_n$ can be used in place of $\mathbf{V}$ in (19). When $\boldsymbol{\Sigma}$ is unknown, it can be estimated consistently with

$$\hat{\boldsymbol{\Sigma}}_n^\star = \frac{(\mathbf{Y}_n - \mathbf{1}_n \hat{\mathbf{A}} - \mathbf{X}_n \hat{\mathbf{B}}_1)^T (\mathbf{Y}_n - \mathbf{1}_n \hat{\mathbf{A}} - \mathbf{X}_n \hat{\mathbf{B}}_1)}{n - p - 1} \tag{20}$$

Let $\hat{\mathbf{V}}_n$ be given by (18) with $\boldsymbol{\Sigma}$ replaced by $\hat{\boldsymbol{\Sigma}}_n^\star$. $\hat{\mathbf{V}}_n$ is a continuous function of a consistent estimator and therefore it is a consistent estimator of $\mathbf{V}$. Thus, by the multivariate version of Slutsky's theorem (see [A 4.19] in Bunke and Bunke, 1986),

$$\sqrt{n}\ \hat{\mathbf{V}}_n^{-1/2} (\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_{pm-r}) \tag{21}$$

3.1. An Asymptotic Weighted Chi-Squared Rank Test

Recall that $d = \dim(\mathcal{S}) = \mathrm{rank}(\mathbf{B}_1)$ and thus, we can use $\hat{\mathbf{B}}_1$ to estimate the dimension of $\mathcal{S}$. Let $\mathbf{Z}_n$ be the $p \times m$ matrix constructed by arranging in m columns the successive elements of the $(pm-r)$-vector $\sqrt{n}\,\hat{\mathbf{V}}_n^{-1/2}(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*)$ according to the non-zero entries of $\hat{\mathbf{B}}_1$. Also let $\mathbf{Z}$ be the $p \times m$ matrix constructed by arranging in m columns the successive elements of the $(pm-r)$-vector $\mathbf{V}^{-1/2}\boldsymbol{\beta}_1^*$ according to the non-zero entries of $\mathbf{B}_1$, where $\mathbf{V}$ is the limit matrix of $\mathbf{V}_n$. The matrix $\mathbf{Z}$ is fixed given the predictors $\mathbf{X}_n$.

The matrix $\mathbf{Z}_n$ is the "standardized" version of $\hat{\mathbf{B}}_1$ in the following sense: Let $\mathbf{I}^-_{pm-r}$ denote the $pm \times pm$ diagonal matrix with zeroes off the main diagonal and zeroes and ones on the main diagonal; zeroes if the corresponding $\hat{\mathbf{B}}_1$ entry is zero. Then, by (21) we have

$$\mathrm{vec}(\mathbf{Z}_n) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}^-_{pm-r}) \tag{22}$$

Observe that the rank of $\mathbf{Z}_n$ is equal to the rank of $\hat{\mathbf{B}}_1$. Therefore, inference on d can be based on the following test statistic

$$\Lambda_k^\star = n \sum_{j=k+1}^{\min(p,m)} (\phi_j^\star)^2 \tag{23}$$

where $\phi_1^\star \geq \phi_2^\star \geq \ldots \geq \phi^\star_{\min(p,m)}$ are the ordered singular values of the $p \times m$ matrix constructed in the manner of $\mathbf{Z}_n$ but from the vector $\sqrt{n}\,\hat{\mathbf{V}}_n^{-1/2}\hat{\boldsymbol{\beta}}_1^*$.

Consider the singular value decomposition of $\mathbf{Z}$,

$$\mathbf{Z} = \boldsymbol{\Gamma}_1^{\star T} \begin{bmatrix} \boldsymbol{\Delta}^\star & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\Gamma}_2^{\star T}$$

$\mathbf{\Delta}^\star$ is a $d \times d$ diagonal matrix with the positive singular values of $\mathbf{Z}$ along its diagonal. Partition $\mathbf{\Gamma}_1^{\star T} = (\mathbf{\Gamma}_{11}^\star, \mathbf{\Gamma}_{12}^\star) : p \times p$, $\mathbf{\Gamma}_{11}^\star : p \times d$, $\mathbf{\Gamma}_{12}^\star : p \times (p-d)$,

$$\mathbf{\Gamma}_2^{\star T} = \begin{bmatrix} \mathbf{\Gamma}_{21}^{\star T} \\ \mathbf{\Gamma}_{22}^{\star T} \end{bmatrix} \quad : m \times m$$

where $\mathbf{\Gamma}_{21}^{\star T} :\ d \times m$, $\mathbf{\Gamma}_{22}^{\star T} : (m-d) \times m$.

When (22) holds, by the Eaton-Tyler (1994) result, the limiting distribution of the smallest $p-d$ singular values of $\mathbf{Z}_n$ is the same as the limiting distribution of the singular values of the $(p-d) \times (m-d)$ matrix

$$\hat{\mathbf{B}}_{\text{std}}^\star = \sqrt{n}\ (\mathbf{\Gamma}_{12}^{\star T} \mathbf{Z}_n \mathbf{\Gamma}_{22}^\star)$$

From (21), the asymptotic distribution of $\text{vec}(\hat{\mathbf{B}}_{\text{std}}^\star)$ is $N_{(p-d)(m-d)}(0, (\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})\mathbf{I}_{mp-r}^{-}(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star))$. Unfortunately the asymptotic covariance matrix

$$(\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})\mathbf{I}_{pm-r}^{-}(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star)$$

is not the identity and we cannot conclude that the sum of the squares of the singular values of $\text{vec}(\hat{\mathbf{B}}_{\text{std}}^\star)$ has a chi-squared asymptotic distribution.

The asymptotic covariance matrix $(\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})\mathbf{I}_{pm-r}^{-}(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star)$ is a $(m-d)(p-d) \times (m-d)(p-d)$ matrix that can be computed as follows: Let $\mathbf{\Gamma}_{22}^\star = (\gamma_{ts}^\star)$, $t = 1, \ldots, m$, $s = 1, \ldots, m-d$. The (s,k)th block of $(\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})\mathbf{I}_{pm-r}^{-}(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star)$ is computed as the product of the sth row block of $(\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})$ with $\mathbf{I}_{pm-r}^{-}$ and the kth block of $(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star)$,

$$[(\mathbf{\Gamma}_{22}^{\star T} \otimes \mathbf{\Gamma}_{12}^{\star T})\mathbf{I}_{pm-r}^{-}(\mathbf{\Gamma}_{22}^\star \otimes \mathbf{\Gamma}_{12}^\star)]_{sk} = \sum_{l}^{m} \sum_{t}^{m} \gamma_{ts}^\star \gamma_{lk}^\star \mathbf{\Gamma}_{12}^{\star T} \mathbf{I}_{tl}^{-} \mathbf{\Gamma}_{12}^\star \tag{24}$$

where $\mathbf{I}_{tl}^{-}$ is the tlth $p \times p$ block matrix of $\mathbf{I}_{pm-r}^{-}$, $t, l = 1, \ldots, m$. Obviously, $\mathbf{I}_{tl}^{-} = 0$ when $t \neq l$, and $\mathbf{I}_{tt}^{-}$ is a diagonal matrix with ones, or ones and zeroes, on the diagonal. If, for some t, $\mathbf{I}_{tt}^{-} = \mathbf{I}_m$, then $\mathbf{\Gamma}_{12}^{\star T} \mathbf{I}_{tt}^{-} \mathbf{\Gamma}_{12}^\star = \mathbf{I}_{p-d}$

and (24) equals

$$\sum_{l}^{m}\sum_{t}^{m}\gamma_{ts}^{\star}\gamma_{lk}^{\star}\mathbf{I}_{p-d}$$

which, in turn, equals either $\mathbf{I}_{p-d}$, when $s = k$, or 0 when $s \neq k$, since $\mathbf{\Gamma}_{22}^{\star T}\mathbf{\Gamma}_{22}^{\star} = \mathbf{I}_{m-d}$. Otherwise, $\mathbf{\Gamma}_{12}^{\star T}\mathbf{I}_{tt}^{-}\mathbf{\Gamma}_{12}^{\star}$ does not equal the identity matrix, and (24) equals a nontrivial diagonal matrix when $s = k$.

The discussion above is summarized in Lemma 3.1 next.

LEMMA 3.1. *Assume that the restricted formulation of model (2) holds, that $\mathbf{V}_n$ converges pointwise to a positive definite limit, and that $\hat{\mathbf{\Sigma}}_n^{\star}$ is a consistent estimate of $\mathbf{\Sigma}$. Let $\phi_1^{\star} \geq \phi_2^{\star} \geq \ldots \geq \phi_{\min(p,m)}^{\star}$, be the ordered singular values of $\hat{\mathbf{B}}_{\text{std}}^{\star}$. Then $\Lambda_d^{\star}$ is asymptotically distributed as a linear combination of $(m-d)(p-d)$ independent chi-squared random variables with one degree of freedom.*

The coefficients of the linear combination of the asymptotic distribution of $\Lambda_d^{\star}$ are unknown, as they are based on the singular value decomposition of the unknown matrix $\mathbf{Z}$. Nevertheless, they can be consistently estimated by replacing the matrices $\mathbf{\Gamma}_{12}^{\star}$ and $\mathbf{\Gamma}_{22}^{\star}$ with the corresponding singular value decomposition matrices of the consistent estimate of $\mathbf{Z}$ resulting from the vector $\hat{\mathbf{V}}_n^{-1/2}\hat{\boldsymbol{\beta}}_1^{*}$.

Alternatively, instead of calculating percentage points for the distribution of a combination of chi-squares, we could employ a Satterthwaite (1941) approximation in which the mean and variance of Λ_d^{*} are used to construct an adjusted version that can be compared to the percentage points of a chi-squared distribution. Fouladi (1997) found in a simulation study that this adjusted statistic performs better than competitors.

4. THE NON-CONSTANT COVARIANCE CASE

In Sections 2 and 3 the parametric models that were used assumed that the covariance structure of the error matrix given $\mathbf{X}$ was constant; that is, independent of $\mathbf{X}$. Occasionally, this assumption may be seriously violated resulting in rank estimation errors. In this section the non-constant error covariance structure case is addressed.

We assume that the regression model (2) holds but now $\mathrm{Cov}(\mathbf{Y}|\mathbf{X})$ is a function of $\mathbf{X}$:

$$\mathrm{Cov}(\mathbf{Y}|\mathbf{X}) = \mathbf{\Sigma}(\mathbf{X}) = (\sigma_{ij}(\mathbf{X}))_{i,j=1}^{m}$$

In this case, the covariance structure of the error matrix $\mathbf{E}_n$ can no longer be represented by the Kronecker product of $\mathbf{\Sigma}$ and the identity $\mathbf{I}_n$, for

$$\mathrm{Cov}(Y_{ki}, Y_{kj}|\mathbf{X} = \mathbf{X}_k) = \sigma_{ij}(\mathbf{X}_k)$$

for $k = 1, \ldots, n$, $i, j = 1, \ldots, m$. Hence, the covariance matrix of $\mathrm{vec}(\mathbf{Y}_n)$, and consequently of $\mathrm{vec}(\mathbf{E}_n)$, is a $nm \times nm$ symmetric matrix consisting of m^2 blocks of order $n \times n$, where the ijth block is the diagonal $n \times n$ matrix with $\sigma_{ij}(\mathbf{X}_1), \ldots, \sigma_{ij}(\mathbf{X}_n)$ along its main diagonal for $i, j = 1, \ldots, m$.

In vector form, $\hat{\mathbf{B}}_n$ can be written as $\mathrm{vec}(\hat{\mathbf{B}}_n) = \mathrm{vec}(\mathbf{W}_n\mathbf{Y}_n) = (\mathbf{I}_m \otimes \mathbf{W}_n)\,\mathrm{vec}(\mathbf{Y}_n)$, where $\mathbf{W}_n = (\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{X}_n^T$ is a $p \times n$ known matrix of weights. The covariance matrix of $\mathrm{vec}\,\hat{\mathbf{B}}_n$ equals $(\mathbf{I}_m \otimes \mathbf{W}_n)\,\mathrm{Cov}(\mathrm{vec}\,\mathbf{Y}_n)(\mathbf{I}_m \otimes \mathbf{W}_n^T)$. Thus, $\mathrm{Cov}(\mathrm{vec}\,\hat{\mathbf{B}}_n)$ is a $pm \times pm$ block matrix, whose ijth block is given by

$$\mathbf{W}_n\ \mathrm{diag}(\sigma_{ij}(\mathbf{X}_1), \ldots, \sigma_{ij}(\mathbf{X}_n))\ \mathbf{W}_n^T \tag{25}$$

for $i, j = 1, \ldots, m$, and $\mathbf{H}_n = \mathrm{Cov}[n^{1/2}\,\mathrm{vec}\,(\mathbf{W}_n\mathbf{Y}_n - \mathbf{B})]$ is a $pm \times pm$ block matrix, whose ijth block is given by (25) multiplied by n.

Assuming that the conditional covariance of Y_i and Y_j given $\mathbf{X}$ is bounded and that $\sigma_{ii}(\mathbf{X}) > 0$, for all $i, j = 1, \dots, m$, we obtain

$$\sqrt{n}\ \mathrm{vec}(\mathbf{W}_n \mathbf{Y}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{H})$$

provided $n\mathbf{W}_n\mathbf{W}_n^T = (\mathbf{X}_n^T\mathbf{X}_n/n)^{-1}$ has a positive definite limit matrix $\mathbf{G}$, where $\mathbf{H}$ is the positive definite limit matrix of $\mathbf{H}_n$ (see lemma A.2 in the appendix).

Let $\hat{\mathbf{\Sigma}}_n(\mathbf{X}) = (\hat{\sigma}_{ij}(\mathbf{X}))$ be a consistent estimate of $\mathbf{\Sigma}(\mathbf{X}) = (\sigma_{ij}(\mathbf{X}))$, for $i, j = 1, \dots, m$. Let $\hat{\mathbf{H}}_n$ be the $pm \times pm$ matrix whose ijth block is given by (25) multiplied by n, with $\hat{\sigma}_{ij}(\mathbf{X}_k)$ in place of $\sigma_{ij}(\mathbf{X}_k)$.

Since $\hat{\mathbf{H}}_n$ is nonsingular, if it were also consistent for $\mathbf{H}$, then by the multivariate version of Slutsky's theorem we would obtain

$$\sqrt{n}\hat{\mathbf{H}}_n^{-1/2}\, \mathrm{vec}\,(\mathbf{W}_n\mathbf{Y}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{I}_m \otimes \mathbf{I}_p)$$

The dimension d is not affected by this nonsingular transformation. It is straightforward to establish that $\hat{\mathbf{H}}_n \to \mathbf{H}$ in probability given a weakly consistent estimator of $\mathbf{\Sigma}(\mathbf{X})$. Moreover, by placing further conditions on the entries of $\hat{\mathbf{\Sigma}}_n(\mathbf{X})$ we obtain that $\hat{\mathbf{H}}_n$ is L^2-consistent for $\mathbf{H}$ (see lemma A.3 in the appendix).

Let $\mathrm{vec}(\hat{\mathbf{B}})_{\mathrm{std}} = \hat{\mathbf{H}}_n^{-1/2}\, \mathrm{vec}(\mathbf{W}_n\mathbf{Y}_n)$. The $p \times m$ matrix $\hat{\mathbf{B}}_{\mathrm{std}}$ that results from the arrangement of $\mathrm{vec}(\hat{\mathbf{B}}_{\mathrm{std}})$ into m columns, satisfies $\mathrm{rank}(\hat{\mathbf{B}}_{\mathrm{std}}) = \mathrm{rank}(\mathbf{W}_n\mathbf{Y}_n)$. Also, let

$$\Lambda_d = n \sum_{j=d+1}^{\min(p,m)} \phi_j^2 \tag{26}$$

where ϕ_j, $j = 1, \dots, \min(p, m)$, denote the ordered singular values of $\hat{\mathbf{B}}_{\mathrm{std}}$. The following theorem states the conditions under which the asymptotic distribution of Λ_d is chi-squared.

THEOREM 4.1. *Assume that all conditions of lemma A.2 are satisfied. If* $d = \operatorname{rank}(\mathbf{B}) = \dim(\mathcal{S})$, *then* Λ_d *as defined in (26) is asymptotically distributed as a* $\chi^2_{(m-d)(p-d)}$ *random variable.*

Proof. The proof is analogous to the proof of Theorem 2.1. ∎

The inferential procedure on d is the same as in the constant covariance case, provided $\mathbf{\Sigma}(\mathbf{X})$ can be estimated consistently. The computation of a consistent estimate of $\mathbf{\Sigma}$ is presented next.

Assume that $\mathrm{E}(Y_j^2) < \infty$, for all $j = 1, \dots, m$. Let

$$\begin{aligned} \hat{\sigma}_{ij}(\mathbf{X}) &= \widehat{\mathrm{Cov}}_n(Y_i, Y_j|\mathbf{X}) \\ &= \hat{\mathrm{E}}_n(Y_iY_j|\mathbf{X}) - \hat{\mathrm{E}}_n(Y_i|\mathbf{X})\hat{\mathrm{E}}_n(Y_j|\mathbf{X}) \end{aligned} \tag{27}$$

for $i,\ j \in \{1, 2, \dots, m\}$, where $\hat{\mathrm{E}}_n(\cdot|\mathbf{X})$ denotes the least squares estimate of $\mathrm{E}(\cdot|\mathbf{X})$ from regressing the argument in $(\cdot)$ on $\mathbf{X}_n$. The choice of the regression model to be fitted on the argument in $(\cdot)$ is guided by the data. Under regularity conditions, least squares estimates are consistent and thus,

$$\hat{\sigma}_{ij}(\mathbf{X}) \to \sigma_{ij}(\mathbf{X}) \tag{28}$$

in probability, for all $i,\ j \in \{1, 2, \dots, m\}$.

Let $\hat{\mathbf{\Sigma}}_n(\mathbf{X}_k)$ be the $m \times m$ matrix with entries given by (27), computed at $\mathbf{X}_k$ for $k = 1, \dots, n$. Then, (28) implies that $\hat{\mathbf{\Sigma}}_n(\mathbf{X}_k)$ is a consistent estimate of $\mathbf{\Sigma}(\mathbf{X}_k)$, for all $k = 1, \dots, n$.

To obtain L^2 consistency for the covariance matrix estimate $\hat{\mathbf{H}}_n$, we should also require that $\hat{\sigma}_{ij}(\mathbf{X}_k)$ be L^2 consistent for $\sigma_{ij}(\mathbf{X}_k)$, for all $i, j = 1, \dots, m,\ k = 1, \dots, n$, and that $\mathrm{Cov}(\hat{\sigma}_{ij}(\mathbf{X}_k), \hat{\sigma}_{ij}(\mathbf{X}_l)) \xrightarrow[n\to\infty]{} 0$, for all $k, l = 1, \dots, n,\ k \neq l$ (see lemma A.3).

4.1. The Restricted Model Case

The covariance matrix of $\hat{\boldsymbol{\beta}}_1^*$ is not as computationally convenient as in the unrestricted model. Let $\tilde{\mathbf{W}}_n = (\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\tilde{\mathbf{X}}_n^T$ be a $(pm-r) \times mn$ matrix of weights that can be written as

$$\tilde{\mathbf{W}}_n = \begin{bmatrix} \tilde{\mathbf{W}}_{11} & \tilde{\mathbf{W}}_{12} & \cdots & \tilde{\mathbf{W}}_{1m} \\ \tilde{\mathbf{W}}_{21} & \tilde{\mathbf{W}}_{22} & \cdots & \tilde{\mathbf{W}}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{W}}_{pm-r,1} & \tilde{\mathbf{W}}_{pm-r,2} & \cdots & \tilde{\mathbf{W}}_{pm-r,m} \end{bmatrix}$$

where $\tilde{\mathbf{W}}_{ij}$ is $1 \times n$ ijth block row vector of $\tilde{\mathbf{W}}_n$, for $i = 1, \ldots, pm-r$, $j = 1, \ldots, m$.

As in the unconstrained model, assuming that the conditional covariance of Y_i and Y_j given $\mathbf{X}$ is bounded and that $\sigma_{ii}(\mathbf{X}) > 0$, for all $i, j = 1, \ldots, m$, we obtain

$$\sqrt{n}\ (\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*) \xrightarrow{\mathcal{D}} N_{pm-r}(0, \mathbf{V}) \tag{29}$$

provided $n\tilde{\mathbf{W}}_n\tilde{\mathbf{W}}_n^T = (\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n/n)^{-1}$ has a positive definite limit matrix, where $\mathbf{V}$ is the positive definite limit matrix of $\mathbf{V}_n$ (see lemma A.2 in the appendix).

The asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}_1^*)$ is also given by (29) when a weakly consistent estimate $\hat{\boldsymbol{\Sigma}}_n(\mathbf{X})$ of $\boldsymbol{\Sigma}(\mathbf{X})$ is used; that is, when $\hat{\mathbf{V}}_n$ is the matrix resulting from $\mathbf{V}_n$ with $\sigma_{ij}(\mathbf{X})$ replaced by $\hat{\sigma}_{ij}(\mathbf{X})$. The argument is similar to the one used in the unconstrained model case.

Inference on $d = \dim(\mathcal{S})$ is based on the test statistic $\Lambda_d^\star$ defined analogously to (23). The computation of the asymptotic distribution of $\Lambda_d^\star$ is similar to that of the constant covariance structure case obtaining that the asymptotic distribution is a weighted sum of independent chi-squared with one degree of freedom random variables.

5. MINNEAPOLIS ELEMENTARY SCHOOL DATA

To illustrate selected aspects of the development in Section 2, we use data on the performance of students in $n = 63$ Minneapolis Schools obtained from Cook (1998). The $m = 4$ dimensional response vector $\mathbf{P}$ consists of the percentages $P_{(\cdot)}$ of students in a school scoring above (A) and below (B) average on standardized fourth and sixth grade reading comprehension tests, $\mathbf{P} = (P_{A4}, P_{B4}, P_{A4}, P_{B6})^T$. Subtracting either pair of grade specific percentages from 100 gives the percentage of students scoring about average on the test. Five predictors ($p = 5$) were used to characterize a school: (1) the percentage of children receiving Aid to Families with Dependent Children (AFDC), (2) the percentage of children not living with biological parents, (3) the percentage of persons in the area below the federal poverty level, (4) percent of adults in the school area who completed high school, and (5) the pupil-teacher ratio.

We began the analysis by inspecting the four univariate regressions. Our results indicated that using the square-root of all percentages resulted in reasonable linear models that passed the standard diagnostic tests. Thus, we next turn to the multivariate analysis using the response $\mathbf{Y} = (P_{(\cdot)}^{1/2})$, and the square-roots of the four percentage predictors and the pupil-teacher ratio. The transformed predictor vector is denoted by $\mathbf{X}$.

Before analyzing the data, we conducted a small simulation study to check on the accuracy of the asymptotic results. In all cases we used the 63 observations on $\mathbf{X}$ in combination with constructed responses $\mathbf{Y}^\star$. In the first simulation, $\mathbf{Y}^\star$ was a 4×1 vector of standard normal variates. According to our asymptotic results, $d = 0$ and Λ_0 defined in (10) should be distributed approximately as a chi-squared variate with $mp = 20$ degrees of freedom. Our results supported this conclusion, although the test may have resulted in a few too many rejections. For instance, in one run of

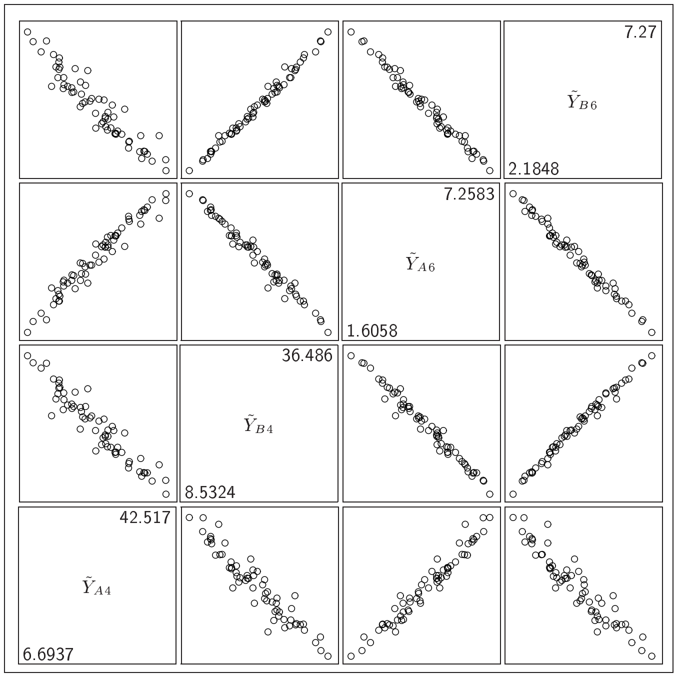

FIG. 1. Scatterplot of $\tilde{\mathbf{Y}}_{(\cdot)}$

1000 simulations we observed about 8.5 percent rejections for a nominal 5 percent test.

In a second set of simulations we used $\mathbf{Y}^{\star} = \mathbf{b}\mathrm{AFDC}^{1/2} + .2N(0, I)$ for various choices of the 4×1 vector $\mathbf{b}$. In this setting, $d = 1$ and Λ_1 should be distributed approximately as a chi-squared variate with 12 degrees of freedom. Again the results indicated that the asymptotic approximation is reasonable, perhaps with a few too many rejections of the order indicated for the case $d = 0$ described previously.

Back to the Minneapolis data, we constructed the test statistics Λ_k, $k = 1, \ldots, 4$ obtaining 299.98, 9.1619, 1.5405, and 0.46212 on 20, 12, 6 and 2 degrees of freedom. The corresponding p-values are about 0, .69, 0.96 and 0.79. Accordingly we inferred that $d = 1$ and thus that a single

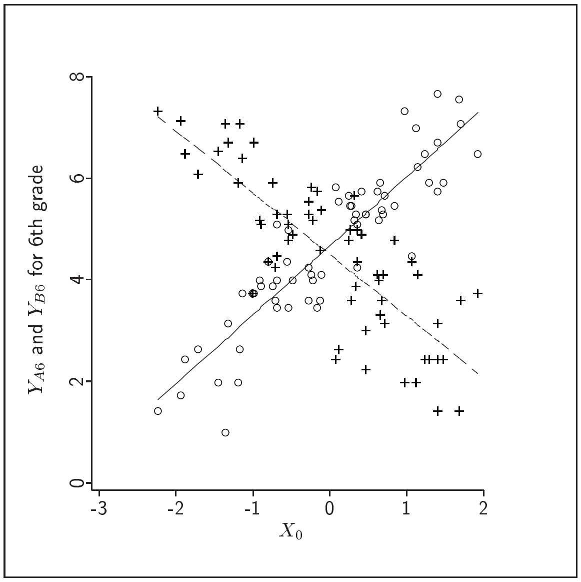

FIG. 2. Y_{A6} and Y_{B6} versus X_0

linear combination X_0 of the predictors carries the information that $\mathbf{X}$ has to furnish about $\mathbf{Y}$. If this conclusion is reasonable then we might expect the fitted values $\tilde{\mathbf{Y}}_{(\cdot)}$ from the four individual regressions to be highly correlated. A scatterplot of these fitted values is shown in Figure 1. A single plot showing both Y_{A6} and Y_{B6} versus X_0 is shown in Figure 2. The plot for the fourth grade percentages is similar.

Having concluded that $d = 1$ is a good choice for this regression, we could next use results from Anderson (1999a) to infer about the rank-reduced coefficient matrix without requiring normal responses.

6. CONCLUDING REMARKS

The relationship between canonical correlation analysis and reduced rank regression has been used to estimate the rank in a reduced rank regression model using Bartlett's (1947) test for the significance of the last $m - d$

canonical correlations. The number of nonzero canonical correlations is the rank of the matrix $\mathrm{Cov}(\mathbf{Y}, \mathbf{X})$ and hence of the coefficient matrix $\mathbf{B}$. Reinsel and Velu (1998) showed that the test is a likelihood ratio test resting on the assumption of normal responses. In particular, Anderson (1999b, sec. 7) computed a likelihood ratio test for the number of zero canonical correlations. He showed that the asymptotic distribution of the log-likelihood ratio to be chi-squared with $(m-d)^2$ degrees of freedom.

Schmidli (1996) used the *information criterion* and the mean squared error for prediction for selecting the model, and hence the coefficient matrix rank, maximizing predictive power. He studied the reliability of the likelihood based estimates and cross validation estimates using the two above measures of the predictive power of models of different ranks.

We have approached the problem of rank estimation from a completely different angle. Starting from the classic multivariate linear regression model, we estimate the rank using the least squares coefficient matrix estimator and its asymptotic normality without imposing distributional assumptions on $\mathbf{Y}$. Thus, we bypass the model selection step that is essential in the existing theory of rank estimation. As an important by-product of our rank estimation procedure, we obtain the linear combinations of the $\mathbf{X}$-components that are found to be significant for modelling $\mathbf{Y}$ as a function of $\mathbf{X}$, and as such can be used in place of the regressor vector resulting in a *dimension reduction* of the regression model.

APPENDIX

Lemma A.1. *Let $\mathcal{M}_{pq}^{>}$ be the space of all $pq \times pq$ positive definite matrices and let $\mathcal{F}$ be the space of distributions of the errors $\mathbf{E}_n$. If*

$$\mathbf{H}_n \xrightarrow[n\to\infty]{} \mathbf{H} \in \mathcal{M}_{pq}^{>} \tag{A.1}$$

then

$$\sqrt{n}\ \mathrm{vec}(\hat{\mathbf{B}}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pq}(0, \mathbf{H})$$

provided the following three conditions are satisfied

$$\|(\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T\|_{\max} = o(n^{-1/2}) \qquad (I)$$

$$\sup_{F \in \mathcal{F}} \int_{\|x\|>c} \|x\|^2 dF(x) \longrightarrow 0 \quad as \ \ c \to \infty \qquad (II)$$

$$\inf_{\mathbf{\Sigma} \in \mathcal{M}(\mathcal{F})} \lambda_{\min}(\mathbf{\Sigma}) \geq r > 0 \qquad (III)$$

where $\mathcal{M}(\mathcal{F}) = \{\int_{R^p} xx^T dF(x) : F \in \mathcal{F}\} \subset \mathcal{M}_p^{>}$.

The notation $\|\cdot\|_{\max}$ identifies the norm on the vector space of matrices defined by $\|(a_{ij})\|_{\max} = \max_{i,j} |a_{ij}|$, for a matrix $\mathbf{A} = (a_{ij})$. $\lambda_{\min}(\mathbf{\Sigma})$ is the smallest eigenvalue of $\mathbf{\Sigma}$ and r is some positive real. The error distributions that are usually considered satisfy Conditions (II) and (III).

Proof. The lemma follows readily from Theorem 2.4.3, Bunke and Bunke (1986), and the multivariate version of Slutsky's theorem (see [A 4.19], Bunke and Bunke (1986)). ▮

Lemma A.2. *Suppose that there exists a positive real number c such that* $0 \leq |\sigma_{ij}(\mathbf{X}) = \mathrm{Cov}(Y_i, Y_j|\mathbf{X})| \leq c$ *and* $\sigma_{ii}(\mathbf{X}) > 0$ *for all* $i, j = 1, \ldots, p$, *and all* $\mathbf{X}$ *in the* $\mathbf{X}$*-sample space. If* $n\mathbf{W}_n\mathbf{W}_n^T = (\mathbf{X}_n^T\mathbf{X}_n/n)^{-1}$ *has a positive definite limit matrix* $\mathbf{G}$, *then* $\mathbf{H}_n$ *has a positive definite limit matrix* $\mathbf{H}$, *and*

$$\sqrt{n}\ \mathrm{vec}\,(\mathbf{W}_n\mathbf{Y}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{H})$$

provided

$$\|\mathbf{W}_n\|_{\max} = \|(\mathbf{X}_n^T\mathbf{X}_n)^{-1}\mathbf{X}_n^T\|_{\max} = o(n^{-1/2}) \qquad \text{(A.2)}$$

and Conditions (II) and (III) of Lemma A.1 hold.

Proof. Consider the ijth block of $\mathbf{H}_n$,

$$n\mathbf{W}_n \mathrm{diag}(\sigma_{ij}(\mathbf{X}_1), \ldots, \sigma_{ij}(\mathbf{X}_n))\mathbf{W}_n^T \tag{A.3}$$

which is obviously a linear transformation of $n\mathbf{W}_n\mathbf{W}_n^T$. Therefore, since $n\mathbf{W}_n\mathbf{W}_n^T \longrightarrow \mathbf{G}$, and $|\sigma_{ij}(\mathbf{X})| \leq c$, (A.3) has a $p \times p$ limit matrix, for all $i, j = 1, \ldots, m$, and hence all m^2 $p \times p$ block matrices of $\mathbf{H}_n$ have limit matrices. This in turn implies that there exists a $pm \times pm$ matrix $\mathbf{H}$, so that $\mathbf{H}_n \longrightarrow \mathbf{H}$. In addition, since $n\mathbf{W}_n\mathbf{W}_n^T$ is positive definite and $\sigma_{ii}(\mathbf{X}) > 0$ for all $i = 1, \ldots, m$, $\mathbf{H}_n$ is also positive definite (Harville, 1997, thm. 12.2.9). The latter obtains that the limit $\mathbf{H}$ of $\mathbf{H}_n$ is also positive definite. By a direct application of Slutsky's theorem ([A 4.19], Bunke and Bunke 1986), we obtain that

$$\mathbf{H}_n^{-1/2}\ \mathrm{vec}\,(\mathbf{W}_n\mathbf{Y}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{I}_{mp} = \mathbf{I}_m \otimes \mathbf{I}_p)$$

if and only if

$$n^{1/2}\ \mathrm{vec}\,(\mathbf{W}_n\mathbf{Y}_n - \mathbf{B}) \xrightarrow{\mathcal{D}} N_{pm}(0, \mathbf{H}) \tag{A.4}$$

Now, a sufficient condition for (A.4) is (A.2) (see Bura 1996, Chapter 5). ∎

Lemma A.3. *Suppose that*

$$n\,\mathbf{W}_n\mathbf{W}_n^T \xrightarrow[n\to\infty]{} \mathbf{G} \in \mathcal{M}_q^{>} \tag{A.5}$$

Suppose that $\hat{\sigma}_{ij}(\mathbf{X})$ converges to $\sigma_{ij}(\mathbf{X})$ in quadratic mean, for all $i, j = 1, \ldots, m$, and all $\mathbf{X}$ in the relevant sample space. Also, suppose that $\mathrm{Cov}(\hat{\sigma}_{ij}(\mathbf{X}_k), \hat{\sigma}_{ij}(\mathbf{X}_l)) \longrightarrow 0$ as $n \to \infty$, for $k, l = 1, \ldots, n$, $k \neq l$. Then, $\hat{\mathbf{H}}_n$ is a L^2-consistent estimate of $\mathbf{H}$.

Proof. Since $\mathbf{H}$ is the limit matrix of $\mathbf{H}_n$, it suffices to show that

$$\hat{\mathbf{H}}_n - \mathbf{H}_n \xrightarrow[n\to\infty]{} 0 \quad \text{in } L^2 \tag{A.6}$$

Then, from the triangle inequality, it follows that $\hat{\mathbf{H}}_n$ is a L^2-consistent estimate of $\mathbf{H}$. Consider the ijth blocks of $\hat{\mathbf{H}}_n$ and $\mathbf{H}_n$. The lqth entry of the ijth block of $\mathbf{H}_n$ is

$$\sum_k^n \sigma_{ij}(\mathbf{X}_k)(nW_{lk}W_{qk}) \tag{A.7}$$

and of $\hat{\mathbf{H}}_n$ is given by (A.7) with $\hat{\sigma}_{ij}$ in place of σ_{ij}, for all $l, q = 1, \dots, p$, and all $i, j = 1, \dots, m$. Then, (A.6) is true if and only if

$$\sum_k^n \hat{\sigma}_{ij}(\mathbf{X}_k)(nW_{lk}W_{qk}) - \sum_k^n \sigma_{ij}(\mathbf{X}_k)(nW_{lk}W_{qk}) \xrightarrow[n\to\infty]{} 0$$

in L^2, for all $l, q = 1, \dots, p$, and $i, j = 1, \dots, m$. Now,

$$\begin{aligned}
&\mathrm{E}[\sum_{k=1}^n \hat{\sigma}_{ij}(\mathbf{X}_k)(nW_{lk}W_{qk}) - \sum_{k=1}^n \sigma_{ij}(\mathbf{X}_k)(nW_{lk}W_{qk})]^2 \\
&\quad = \mathrm{E}[\sum_{k=1}^n (\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k))^2 (nW_{lk}W_{qk})^2] \\
&\quad\quad + \mathrm{E}[\sum_{k=1}^n \sum (\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k))(\hat{\sigma}_{ij}(\mathbf{X}_r) - \sigma_{ij}(\mathbf{X}_r)) \\
&\quad\quad\quad \times (nW_{lk}W_{qk})(nW_{lr}W_{qr})] \\
&\quad = \sum_{k=1}^n \mathrm{E}[(\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k))^2](nW_{lk}W_{qk})^2] \qquad \text{(A.8)} \\
&\quad\quad + \sum_{k=1}^n \sum \mathrm{E}[(\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k))(\hat{\sigma}_{ij}(\mathbf{X}_r) - \sigma_{ij}(\mathbf{X}_r)) \\
&\quad\quad\quad \times (nW_{lk}W_{qk})(nW_{lr}W_{qr})] \qquad \text{(A.9)}
\end{aligned}$$

The integration can be brought inside the sum by the bounded convergence theorem (see Billingsley 1986, p. 214), since (A.5) holds by assumption and $\hat{\sigma}_{ij}(X_k)$ is consistent in quadratic mean for $\sigma_{ij}(\mathbf{X}_k)$, therefore $\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k)$ is L^2 bounded, for all $k = 1, \ldots, n$. But then, we also have that (A.8) vanishes by the L^2 consistency of $\hat{\sigma}_{ij}(\mathbf{X}_k)$. Furthermore, (A.9) goes to zero by assumption. Hence, $\mathrm{E}[\sum_{k=1}^{n}(\hat{\sigma}_{ij}(\mathbf{X}_k) - \sigma_{ij}(\mathbf{X}_k))(nW_{lk}W_{qk})]^2 \xrightarrow[n\to\infty]{} 0$ for all $l, q = 1, \ldots, p$, $i, j = 1, \ldots, m$. ∎

ACKNOWLEDGMENTS

We would like to thank T. W. Anderson for valuable comments and for familiarizing us with relevant work. His help improved this article significantly. The first author would also like to thank the National Science Foundation and Ingram Olkin (DMS-9631278) for giving her the opportunity to visit the Department of Statistics at Stanford during the summer of 2001 where most of this work was completed. The second author's work was supported in part by National Science Foundation grants DMS-9703777 and DMS-0103983.

REFERENCES

Anderson, T. W., Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions, *Ann. Math. Statist.* **22** (1951), 327–351.

Anderson, T. W., Asymptotic Distribution of the Reduced Rank Regression Estimator Under General Conditions, *Ann. Statist.* **22** (1999a), 327–351.

Anderson, T. W., Asymptotic Theory for Canonical Correlation Analysis, *J. Mult. Anal.* **70** (1999b), 1–29.

Bartlett, M. S., Multivariate Analysis, *J. Royal Statist. Soc., Ser. B* **9** (1947), 176–197.

Billingsley, P., "Probability and Measure," New York: Wiley, 1986.

Bunke, H. and Bunke, O. (editors), "Statistical Inference in Linear Models. Statistical Methods of Model Building, Vol I.," Wiley, Chichester, 1986.

Bura, E., "Dimension Reduction via Inverse Regression," Ph.D. Dissertation, School of Statistics, University of Minnesota, 1996.

Cook, R. D., "Regression Graphics: Ideas for Studying Regressions through Graphics," Wiley: New York, 1998.

Eaton, M. L. and Tyler, D. E., The Asymptotic Distribution of Singular Values with Applications to Canonical Correlations and Correspondence Analysis, *J. Mult. Anal.* **34** (1994), 439–446.

Fouladi, R. T., Type I error control of some covariance structure analysis technique under conditions of multivariate non-normality, *Comput. Sci. Statistics* **29** (1997), 526–532.

Harville, D. A., "Matrix Algebra From a Statistician's Perspective," Springer-Verlag: New York, 1997.

Izenman, A. J., Reduced-rank Regression for the Multivariate Linear Model, *J. Mult. Anal.* **5** (1975), 248–264.

Reinsel, G. C. and Velu, R. P., "Multivariate Reduced-Rank Regression: Theory and Applications, Lecture Notes in Statistics," New York: Springer-Verlag, 1998.

Satterthwaite, F. E., Synthesis of variance, *Psychometrika.*, **6** (1941), 309–316.

Seber, G. A. F., "Linear Regression Analysis," New York: Wiley, 1977.

Schmidli, H., "Reduced-rank Regression," Berlin: Physica, 1996.

Tierney, L., "LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics," New York: Wiley, 1990.