PARTIALLY REDUCED-RANK MULTIVARIATE REGRESSION MODELS

Gregory C. Reinsel and Raja P. Velu

University of Wisconsin-Madison and Syracuse University

Abstract: A multivariate subset (or 'partially') reduced-rank regression model is considered as an extension of the usual multivariate reduced-rank model. In the model, the reduced-rank coefficient structure is specified to occur for a subset of the response variables only, which allows for more general situations and can lead to more efficient modeling than the usual reduced-rank model. The maximum likelihood estimation of parameters, likelihood ratio testing for rank, and large sample properties of estimators for this partially reduced-rank model are developed. An empirical procedure to aid in identification of the possible subset reduced-rank structure is suggested. Two numerical examples are examined to illustrate the methodology for the proposed model.

Key words and phrases: Canonical correlations, covariance adjustment, likelihood ratio test, maximum likelihood estimator, partitioned coefficient matrix, partially reduced-rank regression.

1. Introduction and Partially Reduced-Rank Regression Model

We consider the multivariate linear regression model that relates a set of m response variables $\mathbf{Y}_k = (y_{1k}, \ldots, y_{mk})'$ to a set of n predictor variables $\mathbf{X}_k = (x_{1k}, \ldots, x_{nk})'$ through the model

$$\mathbf{Y}_k = \mathbf{C}\,\mathbf{X}_k + \boldsymbol{\epsilon}_k, \quad k = 1, \ldots, T, \tag{1.1}$$

where $\mathbf{C}$ is a $m \times n$ regression matrix. The $m \times 1$ vectors of errors $\boldsymbol{\epsilon}_k$ are assumed to be distributed as iid multivariate normal with mean vector $\mathbf{0}$ and $m \times m$ nonsingular covariance matrix $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{\epsilon}_k)$. We focus on situations where the number m of response variables is moderately large and so we assume that $m \geq n$, although this is not necessary for the subsequent developments.

To accommodate problems associated with the large number of parameters in $\mathbf{C}$ and the possibility of similarities in relationships with the predictor variables $\mathbf{X}_k$ among different response variables y_{ik}, the reduced–rank multivariate linear regression model has been proposed and examined extensively by Anderson (1951, 1999), Izenman (1975), Reinsel and Velu (1998, Chap.2), and many other

authors. Related methodology occurs in other dimension reduction contexts, such as work on parametric inverse regression by Bura and Cook (2001) and on model-free tests for reduced dimension (rank) in multivariate regression by Cook and Setodji (2003). The main feature of the reduced-rank linear model is that it imposes a reduced-rank restriction on the coefficient matrix $\mathbf{C}$ in (1.1) of the form $\text{rank}(\mathbf{C}) = r < \min(m, n)$, which also yields the decomposition $\mathbf{C} = \mathbf{AB}$, where $\mathbf{A}$ is $m \times r$ and $\mathbf{B}$ is $r \times n$. One interpretation provided by this decomposition is that the lower-dimensional set of r predictors $\mathbf{X}_k^* = \mathbf{BX}_k$ contains all the relevant information in the original set of predictors $\mathbf{X}_k$ for representing the variations in the response variables $\mathbf{Y}_k$.

Standard reduced-rank regression methods may not be adequate to identify more specialized structure in the coefficient matrix $\mathbf{C}$ however. For example, suppose the regression structure of (1.1) were such that the $(m-1)$-dimensional subset $\mathbf{Y}_{1k} = (y_{1k}, \dots, y_{m-1,k})'$ has reduced-rank structure of rank 1, $E(\mathbf{Y}_{1k}) = \boldsymbol{\alpha}_1 \boldsymbol{\beta}_1' \mathbf{X}_k$, and $E(y_{mk}) = \boldsymbol{\beta}_m' \mathbf{X}_k$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_m$ are linearly independent $n \times 1$ vectors. Then the model for $\mathbf{Y}_k$ has reduced-rank structure of rank 2, of the form

$$E(\mathbf{Y}_k) = \begin{bmatrix} \boldsymbol{\alpha}_1 \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_m' \end{bmatrix} \mathbf{X}_k = \begin{bmatrix} \boldsymbol{\alpha}_1 & \mathbf{0} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_m' \end{bmatrix} \mathbf{X}_k \equiv \mathbf{A}_* \mathbf{B}_* \mathbf{X}_k.$$

Since $\mathbf{A}_* \mathbf{B}_* = (\mathbf{A}_* \mathbf{P}^{-1})(\mathbf{P} \mathbf{B}_*)$ for any nonsingular 2×2 matrix $\mathbf{P}$, standard reduced-rank estimation of a rank 2 model might not reveal the more specialized structure that actually exists. There might be considerable gains in estimation of the model with the more specialized structure specified over estimation of the general form of rank 2 model. Hence, it would be desirable to have a methodology to identify the existence of such specialized structure, and to properly account for it in efficient estimation of regression parameters. Thus, as a variant of the usual reduced-rank model, we consider a somewhat more general situation as above in which the reduced-rank coefficient structure occurs for (or is concentrated on) only a subset of the response variables.

As illustrated, the usual reduced-rank specification may be somewhat restrictive, since it requires that all response variables have regressions that are expressible in terms of a lower-dimensional set of linear combinations of the predictor variables in $\mathbf{X}_k$. In practice, it may be that this feature does not hold for a (small) subset of the response variables, but for the remaining set of response variables it does, possibly with few relevant linear combinations of the predictor variables. In fact, in such cases the usual reduced-rank procedures may not be very effective or efficient, since the lower-rank feature of the submatrix of $\mathbf{C}$ may not be revealed if the requirement is that the reduced-rank specification be determined on the entire set of response variables. For these cases, the reduced-rank

structure should be imposed on a partitioned (row-wise) submatrix of $\mathbf{C}$ for more proper modeling.

Initially, for estimation, it will be assumed that the components of $\mathbf{Y}_k = (\mathbf{Y}'_{1k}, \mathbf{Y}'_{2k})'$ have a known arrangement so that the first m_1 components $\mathbf{Y}_{1k}$ possess a reduced-rank feature separate from the remaining components $\mathbf{Y}_{2k}$. We partition the regression coefficient matrix $\mathbf{C}$ in (1.1) as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}, \tag{1.2}$$

where $\mathbf{C}_1$ is $m_1 \times n$ and $\mathbf{C}_2$ is $m_2 \times n$, with $m_1 + m_2 = m$, and assume the reduced-rank feature that

$$\text{rank}(\mathbf{C}_1) = r_1 < \min(m_1, n) \tag{1.3}$$

and $\mathbf{C}_2$ is of full rank, and the rows of $\mathbf{C}_2$ are not linearly related to the rows of $\mathbf{C}_1$. Thus, for now it is assumed that the subset of response variables that is of reduced rank is known, but the more practical case where the subset is to be determined will be addressed later. Note these assumptions imply $\text{rank}(\mathbf{C}) = r_1 + m_2 (\leq n)$ in (1.1). We can write $\mathbf{C}_1 = \mathbf{AB}$, where $\mathbf{A}$ is a $m_1 \times r_1$ matrix and $\mathbf{B}$ is a $r_1 \times n$ matrix, both of full ranks, and note that $\mathbf{C}$ has the 'overall factorization' $\mathbf{C} = \text{diag}(\mathbf{A}, \mathbf{I}_{m_2})[\mathbf{B}', \mathbf{C}'_2]'$. We consider the estimation of $\mathbf{A}$, $\mathbf{B}$, and hence of $\mathbf{C}_1$ and $\mathbf{C}_2$, and other related inference procedures for this 'partially' reduced-rank model. In Section 4 we discuss the practical issue of procedures to identify the subset of response variables for which a reduced-rank structure as in (1.2)$-$(1.3) may hold.

2. Maximum Likelihood Estimation of Parameters

We now obtain the maximum likelihood (ML) estimators of parameters for the model (1.1) with the (partially) reduced-rank restriction (1.3). We represent the $m \times T$ data matrix $\mathbf{Y}$ as $\mathbf{Y} = [\mathbf{Y}'_1, \mathbf{Y}'_2]'$, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are $m_1 \times T$ and $m_2 \times T$ matrices of values of the response vectors $\mathbf{Y}_{1k}$ and $\mathbf{Y}_{2k}$, respectively. The full-rank or least squares (LS) estimators of $\mathbf{C}_1$ and $\mathbf{C}_2$ are $\widetilde{\mathbf{C}}_1 = \mathbf{Y}_1 \mathbf{X}'(\mathbf{XX}')^{-1}$ and $\widetilde{\mathbf{C}}_2 = \mathbf{Y}_2 \mathbf{X}'(\mathbf{XX}')^{-1}$, where $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$. We partition the error vectors as $\boldsymbol{\epsilon}_k = (\boldsymbol{\epsilon}'_{1k}, \boldsymbol{\epsilon}'_{2k})'$ and partition the error covariance matrix $\boldsymbol{\Sigma} = \text{cov}(\boldsymbol{\epsilon}_k)$ such that $\boldsymbol{\Sigma}_{11} = \text{cov}(\boldsymbol{\epsilon}_{1k})$, $\boldsymbol{\Sigma}_{12} = \text{cov}(\boldsymbol{\epsilon}_{1k}, \boldsymbol{\epsilon}_{2k})$, and $\boldsymbol{\Sigma}_{22} = \text{cov}(\boldsymbol{\epsilon}_{2k})$.

We first note that the decomposition $\mathbf{C}_1 = \mathbf{AB}$ is not unique. Therefore we need to impose some normalization conditions, chosen as follows:

$$\mathbf{B}\widehat{\boldsymbol{\Sigma}}_{xx}\mathbf{B}' = \boldsymbol{\Lambda}^2 \quad \text{and} \quad \mathbf{A}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{A} = \mathbf{I}_{r_1}, \tag{2.1}$$

where $\widehat{\boldsymbol{\Sigma}}_{xx} = (1/T)\mathbf{X}\mathbf{X}'$, $\boldsymbol{\Lambda}^2 = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{r_1}^2)$, and $\mathbf{I}_{r_1}$ denotes an $r_1 \times r_1$ identity matrix. We select these because in the basic multivariate reduced–rank regression model, the estimator of $\mathbf{A}$ and $\mathbf{B}$ can be related to eigenvectors associated with the 'r' largest canonical correlations (see (Reinsel and Velu (1998, p.28)). Thus the number of free parameters in the regression coefficient matrix structure of the model is $r_1(m_1 + n - r_1) + m_2 n$ compared to mn parameters in the full-rank model. Also note that the number of free parameters in the usual reduced-rank model, which considers only that the rank of the overall matrix $\mathbf{C}$ is $r = r_1 + m_2$, is $r(m + n - r) = r_1(m_1 + n - r_1) + m_2(m_1 + n - r_1)$. Hence if m_1 is relatively large, there can be substantial reductions in the number of model parameters.

Assuming the $\boldsymbol{\epsilon}_k$ are iid, following a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, apart from irrelevant constants the log-likelihood is

$$L(\mathbf{C}_1, \mathbf{C}_2, \boldsymbol{\Sigma}) = \left(\frac{T}{2}\right) \left[\log |\boldsymbol{\Sigma}^{-1}| - \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{W}\right)\right], \qquad (2.2)$$

where $\mathbf{W} = (1/T)(\mathbf{Y} - \mathbf{CX})(\mathbf{Y} - \mathbf{CX})'$ and $|\boldsymbol{\Sigma}^{-1}|$ is the determinant of the matrix $\boldsymbol{\Sigma}^{-1}$. Maximizing (2.2) with respect to $\boldsymbol{\Sigma}$ yields $\widehat{\boldsymbol{\Sigma}} = \mathbf{W}$. Hence, the concentrated log-likelihood is $L(\mathbf{C}_1, \mathbf{C}_2, \widehat{\boldsymbol{\Sigma}}) = -(T/2)(\log |\mathbf{W}| + m)$. We can proceed to directly derive the ML estimates of $\mathbf{C}_1 = \mathbf{AB}$ and $\mathbf{C}_2$ as values that minimize $|\mathbf{W}|$, but as an alternative we consider an argument based on a conditional distribution approach for additional insight.

Consider maximizing the likelihood expressed in terms of the marginal distribution for $\mathbf{Y}_1$ and the conditional distribution for $\mathbf{Y}_2$ given $\mathbf{Y}_1$. The model for $\mathbf{Y}_1$ is of course the reduced rank model $\mathbf{Y}_{1k} = \mathbf{C}_1\mathbf{X}_k + \boldsymbol{\epsilon}_{1k}$, $k = 1, \ldots, T$, with parameters $\mathbf{C}_1 = \mathbf{AB}$ and $\boldsymbol{\Sigma}_{11} = \mathrm{cov}(\boldsymbol{\epsilon}_{1k})$, and the conditional model for $\mathbf{Y}_2$, given $\mathbf{Y}_1$, is representable as

$$\mathbf{Y}_{2k} = \mathbf{C}_2\mathbf{X}_k + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left(\mathbf{Y}_{1k} - \mathbf{C}_1\mathbf{X}_k\right) + \boldsymbol{\epsilon}_{2k}^* \equiv \mathbf{C}_2^*\mathbf{X}_k + \mathbf{D}^*\mathbf{Y}_{1k} + \boldsymbol{\epsilon}_{2k}^* \qquad (2.3)$$

with parameters $\mathbf{C}_2^* = \mathbf{C}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{C}_1$, $\mathbf{D}^* = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$, and $\boldsymbol{\Sigma}_{22}^* = \mathrm{cov}(\boldsymbol{\epsilon}_{2k}^*) = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, and $\mathrm{cov}(\boldsymbol{\epsilon}_{1k}, \boldsymbol{\epsilon}_{2k}^*) = \mathbf{0}$. The two sets of parameters $\{\mathbf{C}_1, \boldsymbol{\Sigma}_{11}\}$ and $\{\mathbf{C}_2^*, \mathbf{D}^*, \boldsymbol{\Sigma}_{22}^*\}$ are functionally independent, so ML estimation can be performed separately on them.

ML estimation for $\mathbf{C}_1$ and $\boldsymbol{\Sigma}_{11}$ is the same as ML estimation of the standard reduced-rank model, and is equivalent to simultaneously minimizing the eigenvalues of $\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}(\widetilde{\mathbf{C}}_1 - \mathbf{AB})\widehat{\boldsymbol{\Sigma}}_{xx}(\widetilde{\mathbf{C}}_1 - \mathbf{AB})'\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}$, where $\widetilde{\mathbf{C}} = [\widetilde{\mathbf{C}}_1', \widetilde{\mathbf{C}}_2']' = \mathbf{YX}'(\mathbf{XX}')^{-1}$ is the (full-rank) LS estimate of $\mathbf{C}$, and $\widetilde{\boldsymbol{\Sigma}}_{11}$ is the upper-left block

of the corresponding estimate $\widetilde{\boldsymbol{\Sigma}} = (1/T)(\mathbf{Y} - \widetilde{\mathbf{C}}\mathbf{X})(\mathbf{Y} - \widetilde{\mathbf{C}}\mathbf{X})'$. From results given by Reinsel and Velu (1998, p.30), this yields the ML estimators of $\mathbf{A}$ and $\mathbf{B}$ as

$$\widehat{\mathbf{A}} = \widetilde{\boldsymbol{\Sigma}}_{11}^{\frac{1}{2}}\widehat{\mathbf{V}}_{(r_1)}, \quad \widehat{\mathbf{B}} = \widehat{\mathbf{V}}'_{(r_1)}\widetilde{\boldsymbol{\Sigma}}_{11}^{-\frac{1}{2}}\widetilde{\mathbf{C}}_1, \tag{2.4}$$

where $\widehat{\mathbf{V}}_{(r_1)} = [\widehat{\mathbf{V}}_1, \ldots, \widehat{\mathbf{V}}_{r_1}]$, and $\widehat{\mathbf{V}}_j$ is the (normalized) eigenvector that corresponds to the jth largest eigenvalue $\hat{\lambda}_j^2$ of the matrix $\widehat{\mathbf{R}}_1 = \widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}\widetilde{\mathbf{C}}_1\widehat{\boldsymbol{\Sigma}}_{xx}\widetilde{\mathbf{C}}_1'\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}$. The ML estimator of $\mathbf{C}_1$ is $\widehat{\mathbf{C}}_1 = \widehat{\mathbf{A}}\widehat{\mathbf{B}}$, and the ML estimator of $\boldsymbol{\Sigma}_{11}$ under the reduced-rank structure is given by $\widehat{\boldsymbol{\Sigma}}_{11} = (1/T)(\mathbf{Y}_1 - \widehat{\mathbf{C}}_1\mathbf{X})(\mathbf{Y}_1 - \widehat{\mathbf{C}}_1\mathbf{X})'$.

For the conditional model (2.3), notice that the parameters $\mathbf{C}_2^*$ and $\mathbf{D}^*$ are full rank by assumption, so ML estimation in the conditional model simply yields the usual full- rank LS estimates. These can be expressed in a convenient form (e.g., see Reinsel and Velu (1998, p.7)) as

$$\widehat{\mathbf{C}}_2^* = \widetilde{\mathbf{C}}_2 - \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widetilde{\mathbf{C}}_1, \quad \widehat{\mathbf{D}}^* = \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}, \tag{2.5}$$

with $\widehat{\boldsymbol{\Sigma}}_{22}^* = \widetilde{\boldsymbol{\Sigma}}_{22} - \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widetilde{\boldsymbol{\Sigma}}_{12}$. The MLE of $\mathbf{C}_2 = \mathbf{C}_2^* + \mathbf{D}^*\mathbf{C}_1$ can then be obtained as

$$\widehat{\mathbf{C}}_2 = \widehat{\mathbf{C}}_2^* + \widehat{\mathbf{D}}^*\widehat{\mathbf{C}}_1 = \widetilde{\mathbf{C}}_2 - \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}(\widetilde{\mathbf{C}}_1 - \widehat{\mathbf{C}}_1). \tag{2.6}$$

The MLEs of $\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$ can be obtained accordingly, from $\widehat{\boldsymbol{\Sigma}}_{21} = \widehat{\mathbf{D}}^*\widehat{\boldsymbol{\Sigma}}_{11} = \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}_{11}$ and

$$\widehat{\boldsymbol{\Sigma}}_{22} = \widehat{\boldsymbol{\Sigma}}_{22}^* + \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}_{12} = \widetilde{\boldsymbol{\Sigma}}_{22} - \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widetilde{\boldsymbol{\Sigma}}_{12} + \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widehat{\boldsymbol{\Sigma}}_{11}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widetilde{\boldsymbol{\Sigma}}_{12}.$$

Thus the MLEs $\widehat{\mathbf{C}}_1$ and $\widehat{\mathbf{C}}_2$ can be obtained using matrix routines for eigenvector (or singular value) decompositions in a non-iterative fashion. The ML estimation strategy essentially corresponds to the usual reduced-rank estimation to obtain $\widehat{\mathbf{C}}_1$ followed by 'covariance-adjustment' estimation (2.6) to obtain $\widehat{\mathbf{C}}_2$, where the LS estimator $\widetilde{\mathbf{C}}_2$ is adjusted by the 'covariates' $\mathbf{Y}_1 - \widehat{\mathbf{C}}_1\mathbf{X}$.

Note that $\widetilde{\mathbf{C}}_1 - \widehat{\mathbf{C}}_1 = (\mathbf{I} - \widehat{\mathbf{P}}_1)\widetilde{\mathbf{C}}_1$ and $\widetilde{\mathbf{C}}_2 - \widehat{\mathbf{C}}_2 = \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}(\widetilde{\mathbf{C}}_1 - \widehat{\mathbf{C}}_1) = \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}(\mathbf{I} - \widehat{\mathbf{P}}_1)\widetilde{\mathbf{C}}_1$, where $\widehat{\mathbf{P}}_1 = \widetilde{\boldsymbol{\Sigma}}_{11}^{1/2}\widehat{\mathbf{V}}_{(r_1)}\widehat{\mathbf{V}}'_{(r_1)}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}$ is an idempotent matrix of rank r_1. Hence, $\widetilde{\mathbf{C}} - \widehat{\mathbf{C}} = \widetilde{\mathbf{Q}}(\mathbf{I} - \widehat{\mathbf{P}}_1)\widetilde{\mathbf{C}}_1$, where $\widetilde{\mathbf{Q}}' = [\mathbf{I}_{m_1}, \widetilde{\boldsymbol{\Sigma}}_{11}^{-1}\widetilde{\boldsymbol{\Sigma}}_{12}]$, and then from basic results we get

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}} &= \widetilde{\boldsymbol{\Sigma}} + (\widetilde{\mathbf{C}} - \widehat{\mathbf{C}})\widehat{\boldsymbol{\Sigma}}_{xx}(\widetilde{\mathbf{C}} - \widehat{\mathbf{C}})' \\
&= \widetilde{\boldsymbol{\Sigma}} + \widetilde{\mathbf{Q}}(\mathbf{I} - \widehat{\mathbf{P}}_1)\widetilde{\mathbf{C}}_1\widehat{\boldsymbol{\Sigma}}_{xx}\widetilde{\mathbf{C}}_1'(\mathbf{I} - \widehat{\mathbf{P}}_1)'\widetilde{\mathbf{Q}}' \\
&= \widetilde{\boldsymbol{\Sigma}} + \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{\Sigma}}_{11}^{\frac{1}{2}}\left[\mathbf{I} - \widehat{\mathbf{V}}_{(r_1)}\widehat{\mathbf{V}}'_{(r_1)}\right]\widehat{\mathbf{R}}_1\left[\mathbf{I} - \widehat{\mathbf{V}}_{(r_1)}\widehat{\mathbf{V}}'_{(r_1)}\right]\widetilde{\boldsymbol{\Sigma}}_1^{\frac{1}{2}}1\widetilde{\mathbf{Q}}'. \tag{2.7}
\end{aligned}$$

3. Likelihood Ratio Test for Rank and Inference Results

We consider the likelihood ratio (LR) test of the hypothesis $H_0 : \text{rank}(\mathbf{C}_1) \leq r_1$. The LR test statistic for testing $\text{rank}(\mathbf{C}_1) = r_1$ is $\lambda = U^{T/2}$, where $U = |\mathbf{S}|/|\mathbf{S}_0| \equiv |\widetilde{\mathbf{\Sigma}}|/|\widehat{\mathbf{\Sigma}}|$, $\mathbf{S} = (\mathbf{Y} - \widetilde{\mathbf{C}}\mathbf{X})(\mathbf{Y} - \widetilde{\mathbf{C}}\mathbf{X})'$ is the residual sum of squares matrix from fitting the full-rank model, while $\mathbf{S}_0 = (\mathbf{Y} - \widehat{\mathbf{C}}^{(r)}\mathbf{X})(\mathbf{Y} - \widehat{\mathbf{C}}^{(r)}\mathbf{X})'$ is the residual sum of squares matrix from fitting the model under the rank condition on $\mathbf{C}_1$. Here $\widehat{\mathbf{C}}^{(r)}$ denotes the estimate of $\mathbf{C}$ under rank condition (1.3). It is known that $|\widehat{\mathbf{\Sigma}}| = |\widehat{\mathbf{\Sigma}}_{11}||\widehat{\mathbf{\Sigma}}_{22}^*| = |\widehat{\mathbf{\Sigma}}_{11}||\widetilde{\mathbf{\Sigma}}_{22}^*|$ and $|\widetilde{\mathbf{\Sigma}}| = |\widetilde{\mathbf{\Sigma}}_{11}||\widetilde{\mathbf{\Sigma}}_{22}^*|$, where $\widehat{\mathbf{\Sigma}}_{22}^* \equiv \widetilde{\mathbf{\Sigma}}_{22}^*$ since reduced-rank estimation in the model for $\mathbf{Y}_1$ does not affect the LS estimation for the conditional model (2.3). Therefore, we have $U = |\widetilde{\mathbf{\Sigma}}_{11}|/|\widehat{\mathbf{\Sigma}}_{11}|$. It follows that the LR testing procedure is the same as in the usual reduced-rank regression model for $\mathbf{Y}_1$, and does not involve the response variables $\mathbf{Y}_2$. Therefore, the criterion $\lambda = U^{T/2}$ is such that (e.g., Reinsel and Velu (1998, Sec. 2.6))

$$-2\log(\lambda) = T \sum_{j=r_1+1}^{m_1} \log\left(1 + \hat{\lambda}_j^2\right) = -T \sum_{j=r_1+1}^{m_1} \log\left(1 - \hat{\rho}_j^2\right), \qquad (3.1)$$

where $\hat{\lambda}_j^2$, $j = r_1 + 1, \ldots, m_1$ are the $(m_1 - r_1)$ smallest eigenvalues of $\widehat{\mathbf{R}}_1$, and $1 + \hat{\lambda}_j^2 = 1/(1 - \hat{\rho}_j^2)$, where the $\hat{\rho}_j^2$ are the squared sample canonical correlations between $\mathbf{Y}_{1k}$ and $\mathbf{X}_k$ (adjusting for sample means if constant terms are allowed for). Then (3.1) follows asymptotically the $\chi^2_{(m_1-r_1)(n-r_1)}$ distribution under the null hypothesis (see Anderson (1951, Them. 3). A simple correction factor for the LR statistic in (3.1), to improve the approximation to the $\chi^2_{(m_1-r_1)(n-r_1)}$ distribution, is given by $\mathcal{M} = -2\{[T - n + (n - m_1 - 1)/2]/T\}\log(\lambda) = -[T - n + (n - m_1 - 1)/2]\sum_{j=r_1+1}^{m_1}\log(1 - \hat{\rho}_j^2)$. This approximation is known to work well when T is large (see Anderson (1984, Chap.8)).

The alternative hypothesis in the above testing procedure is that the matrix $\mathbf{C}$ is of full rank. There may be situations where $r_1 + m_2 = r < \min(m, n)$ so that the matrix $\mathbf{C}$ would still have reduced rank r to begin with. We might want to test the subset reduced-rank model assumptions of (1.2)−(1.3) against the alternative of the usual reduced-rank model, $\text{rank}(\mathbf{C}) = r = r_1 + m_2$. The form of the LR statistic for this test can readily be developed, and

$$-2\log(\lambda) = -T\left[\sum_{j=r_1+1}^{m_1} \log\left(1 - \hat{\rho}_j^2\right) - \sum_{j=r+1}^{m} \log\left(1 - \tilde{\rho}_j^2\right)\right], \qquad (3.2)$$

where $\tilde{\rho}_j^2$, $j = r + 1, \ldots, m$, are the $(m - r) \equiv (m_1 - r_1)$ smallest squared sample canonical correlations between $\mathbf{Y}_k$ and $\mathbf{X}_k$, with $-2\log(\lambda)$ distributed as $\chi^2_{m_2(m_1-r_1)}$ asymptotically.

Concerning distributional properties of the ML estimators $\widehat{\mathbf{C}}_1$ and $\widehat{\mathbf{C}}_2$, the approximate normal distribution and approximate covariance matrix of the reduced-rank estimator $\widehat{\mathbf{C}}_1$ follow directly from results by Anderson (1999) and Reinsel and Velu (1998, Secs. 2.5 and 3.5). In particular, the results from Anderson (1999, pp.1147-1148) imply that the large sample (as $T \to \infty$) approximate covariance matrix of $\widehat{\mathbf{C}}_1$ is given by

$$\text{cov}[\text{vec}(\widehat{\mathbf{C}}_1)] = \boldsymbol{\Sigma}_{11} \otimes \left(\mathbf{X}\mathbf{X}'\right)^{-1} - \left[\boldsymbol{\Sigma}_{11} - \mathbf{A}\left(\mathbf{A}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}'\right]$$
$$\otimes \left[\left(\mathbf{X}\mathbf{X}'\right)^{-1} - \mathbf{B}'\left(\mathbf{B}\left(\mathbf{X}\mathbf{X}'\right)\mathbf{B}'\right)^{-1}\mathbf{B}\right]. \tag{3.3}$$

For the general case, arguments similar to those by Ahn and Reinsel (1988) and (Reinsel and Velu (1998, Sec. 3.5)) establish that the asymptotic covariance matrix of the joint ML estimator $\widehat{\gamma} = [\text{vec}(\widehat{\mathbf{C}}_1'), \text{vec}(\widehat{\mathbf{C}}_2')]'$ is given by

$$\text{cov}\left(\widehat{\gamma}\right) = \mathbf{M}\left[\mathbf{M}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}\mathbf{X}'\right)\mathbf{M}\right]^{-1}\mathbf{M}', \tag{3.4}$$

where $\mathbf{M} = \text{diag}\left(\mathbf{M}_1, \mathbf{I}_{m_2 n}\right)$ and $\mathbf{M}_1 = \partial\gamma_1/\partial\boldsymbol{\theta}'$, with $\gamma_1 = \text{vec}(\mathbf{C}_1') \equiv \text{vec}(\mathbf{B}'\mathbf{A}')$ and $\boldsymbol{\theta} = [\text{vec}(\mathbf{A}')', \text{vec}(\mathbf{B}')']'$. It can be verified that this approach yields the same result for $\text{cov}[\text{vec}(\widehat{\mathbf{C}}_1')]$ as the expression given in (3.3) and, furthermore, that we obtain the asymptotic expression

$$\text{cov}[\text{vec}(\widehat{\mathbf{C}}_2')] = \boldsymbol{\Sigma}_{22}^* \otimes \left(\mathbf{X}\mathbf{X}'\right)^{-1} + (\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} \otimes \mathbf{I}_n)\text{cov}[\text{vec}(\widehat{\mathbf{C}}_1')](\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \otimes \mathbf{I}_n),$$
$$\tag{3.5}$$

where $\text{cov}[\text{vec}(\mathbf{C}_1')]$ is given by the expression in (3.3), and $\boldsymbol{\Sigma}_{22}^* = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Consider, for instance, the extreme case in which $r_1 = \text{rank}(\mathbf{C}_1)$ is taken as 0, so that $\mathbf{C}_1 = 0$ and estimation of $\mathbf{C}_1$ is not involved. Then, in (2.6), $\widehat{\mathbf{C}}_2 = \widehat{\mathbf{C}}_2^*$ and (3.5) collapses to $\text{cov}[\text{vec}(\widehat{\mathbf{C}}_2')] = \boldsymbol{\Sigma}_{22}^* \otimes (\mathbf{X}\mathbf{X}')^{-1}$. This can be recognized as a familiar result in the context of 'covariance-adjustment', where in this case the entire set of response variables in $\mathbf{Y}_{1k}$ would be used as covariates for $\mathbf{Y}_{2k}$ since $\mathbf{C}_1 = 0$ implies that the complete vector of responses $\mathbf{Y}_{1k}$ is entirely unrelated to $\mathbf{X}_k$.

For the covariance matrix of $\widehat{\mathbf{C}}_2$, it may also be instructive to mention a more direct argument. Since $\widetilde{\boldsymbol{\Sigma}}$ is consistent for $\boldsymbol{\Sigma}$, we can write the ML estimator $\widehat{\mathbf{C}}_2$ as

$$\widehat{\mathbf{C}}_2 = \widetilde{\mathbf{C}}_2 - \widetilde{\boldsymbol{\Sigma}}_{21}\widetilde{\boldsymbol{\Sigma}}_{11}^{-1}(\widetilde{\mathbf{C}}_1 - \widehat{\mathbf{C}}_1) = \widetilde{\mathbf{C}}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\widetilde{\mathbf{C}}_1 - \widehat{\mathbf{C}}_1) + o_p\left(T^{-\frac{1}{2}}\right)$$
$$= (\widetilde{\mathbf{C}}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\widetilde{\mathbf{C}}_1) + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\widehat{\mathbf{C}}_1 + o_p\left(T^{-\frac{1}{2}}\right)$$
$$= \mathbf{H}\widetilde{\mathbf{C}} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\widehat{\mathbf{C}}_1 + o_p\left(T^{-\frac{1}{2}}\right), \tag{3.6}$$

where $\mathbf{H} = \left[-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}, \mathbf{I}\right]$ is such that $\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}' = \boldsymbol{\Sigma}_{22}^*$. Because $(\widetilde{\mathbf{C}}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\widetilde{\mathbf{C}}_1)$ and $\widetilde{\mathbf{C}}_1$ have zero covariance, it follows that the two terms in (3.6) that comprise $\widehat{\mathbf{C}}_2$ are uncorrelated (independent) asymptotically. So we obtain the asymptotic covariance matrix of $\widehat{\mathbf{C}}_2$ using (3.6) from $\mathrm{cov}[\mathrm{vec}(\widehat{\mathbf{C}}_2')] = \mathrm{cov}[\mathrm{vec}(\widetilde{\mathbf{C}}'\mathbf{H}')] + \mathrm{cov}[\mathrm{vec}(\widehat{\mathbf{C}}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})]$, the result in (3.5).

Finally, we can compare the distributional properties of the ML estimators $\widehat{\mathbf{C}}_1$ and $\widehat{\mathbf{C}}_2$ under the subset reduced-rank model assumptions of $(1.2)-(1.3)$ with those for ML estimators obtained under the usual or standard reduced-rank model assumptions. As noted in Section 1, $(1.2)-(1.3)$ imply that $\mathrm{rank}(\mathbf{C}) = r_1 + m_2 = r$ in (1.1) with an overall factorization as $\mathbf{C} = \mathrm{diag}(\mathbf{A}, \mathbf{I}_{m_2})[\mathbf{B}', \mathbf{C}_2']' \equiv \mathbf{A}_*\mathbf{B}_*$. When $r = \min(m, n)$ there is no reduced rank overall and one would consider the usual LS estimator $\widetilde{\mathbf{C}}$ with $\mathrm{cov}[\mathrm{vec}(\widetilde{\mathbf{C}}')] = \boldsymbol{\Sigma} \otimes (\mathbf{X}\mathbf{X}')^{-1}$. When $r < \min(m, n)$, however, one may still obtain the ML estimator of $\mathbf{C}$ for the usual reduced-rank model, denoted as $\overline{\mathbf{C}} = \overline{\mathbf{A}}_*\overline{\mathbf{B}}_*$ with components $\overline{\mathbf{C}}_1$ and $\overline{\mathbf{C}}_2$. Using similar methods as previously, it can be shown that the asymptotic covariance matrices of $\mathrm{vec}(\overline{\mathbf{C}}_1')$ and $\mathrm{vec}(\overline{\mathbf{C}}_2')$ are of the same form as the ML estimators in (3.3) and (3.5) under the subset reduced-rank model, but with $(\mathbf{X}\mathbf{X}')^{-1} - \mathbf{B}_*'(\mathbf{B}_*(\mathbf{X}\mathbf{X}')\mathbf{B}_*')^{-1}\mathbf{B}_*$ in place of $(\mathbf{X}\mathbf{X}')^{-1} - \mathbf{B}'(\mathbf{B}(\mathbf{X}\mathbf{X}')\mathbf{B}')^{-1}\mathbf{B}$. These results can thus be directly compared with the results in (3.3) and (3.5) to indicate the increase in the covariance matrix relative to the subset reduced-rank model.

4. Procedures for Identification of Subset Reduced-Rank Structure

In practice, we typically would not know or be able to specify a priori the subset of response variables for which $(1.2)-(1.3)$ holds. In (1.1) with $m > n$, however, there will exist $m - n$ linear combinations $\ell_j'\mathbf{Y}_k$ such that $\ell_j'\mathbf{C} = 0$. When $(1.2)-(1.3)$ holds, $m_1 - r_1$ of the vectors ℓ_j can take the form $\ell_j' = [\ell_{1j}', \mathbf{0}']$, with ℓ_{1j} such that $\ell_{1j}'\mathbf{C}_1 = 0$. As described by Reinsel and Velu (1998, Chap.2), the ℓ_j can be estimated by $\hat{\ell}_j = \widetilde{\boldsymbol{\Sigma}}^{-1/2}\widehat{V}_j$, $j = n + 1, \ldots, m$, where the $\widehat{V}_j$ are normalized eigenvectors of the matrix $\widehat{\mathbf{R}} = \widetilde{\boldsymbol{\Sigma}}^{-1/2}\widetilde{\mathbf{C}}\widehat{\boldsymbol{\Sigma}}_{xx}\widetilde{\mathbf{C}}'\widetilde{\boldsymbol{\Sigma}}^{-1/2}$ associated with its $m - n$ smallest (zero) eigenvalues. If a certain number $(m_1 - r_1 > m_1 - n)$ of the estimates $\hat{\ell}_j$ satisfy $\hat{\ell}_j' \approx [\hat{\ell}_{1j}', \mathbf{0}']$, for a particular partition of the response variables, then this feature can serve as a preliminary tool to identify the existence and nature of the partially reduced-rank structure for $\mathbf{C}$.

If desired, one can also systematically examine all possible subsets of the response variables for each given dimension $m_1 < m$, and compute the LR statistics for tests of rank and other summary statistics for each subset case. An information criterion such as AIC could then be used to select the 'best' subset of response variables that exhibits the most desirable (partially) reduced-rank

regression features. The procedure of examining all possible subsets of the response variables for tests of rank and other features may not be practical or desirable when the dimension m is relatively large. Thus we suggest and describe briefly a stepwise "backward-elimination" procedure. In this, we start with all $m(\geq n)$ response variables included in $\mathbf{Y}_{1k} \equiv \mathbf{Y}_k$ and carry out the LR test procedure for reduced rank. Assume that r is identified as the rank of the overall coefficient matrix $\mathbf{C}$ by the LR test at this stage, that is, r is the smallest value such that the hypothesis $H_0 : \text{rank}(\mathbf{C}) \leq r$ is not rejected by the LR test procedure. Then we consider each of the m distinct subsets of $\mathbf{Y}_k$, of the form $(y_{1k}, \ldots, y_{i-1,k}, y_{i+1,k}, \ldots, y_{mk})'$, obtained by excluding one response variable (the ith variable) at a time. For each $(m-1)$-dimensional subset we calculate the LR statistic for testing the corresponding hypothesis that $\text{rank}(\mathbf{C}_1) = r - 1$. If $y_{i'k}$ is the response variable that yields the *smallest* value of the LR statistic for testing $\text{rank}(\mathbf{C}_1) = r - 1$ and if this value leads to not rejecting this null hypothesis, then the subset $\mathbf{Y}_{1k}$ with variable $y_{i'k}$ excluded is chosen. After excluding $y_{i'k}$, we test if it is reasonable to reduce the rank of $\mathbf{C}_1$ without discarding any more y-variables. At the next step we consider each of the $m - 1$ distinct subsets of this $\mathbf{Y}_{1k}$ obtained by excluding a remaining response variable one at a time. For each $(m - 2)$-dimensional subset we calculate the LR statistic for testing the corresponding hypothesis $\text{rank}(\mathbf{C}_1) \leq r - 2$. If $y_{i''k}$ is the response variable that yields the *smallest* value of the LR test statistic and if this leads to not rejecting this null hypothesis, then the new $(m - 2)$-dimensional subset $\mathbf{Y}_{1k}$ with variable $y_{i''k}$ excluded (in addition to the previously excluded variable $y_{i'k}$) is chosen. The stepwise procedure continues until no further response variables can be 'eliminated'.

5. Illustrative Examples

In this section we present two numerical examples to illustrate our (partially) reduced-rank methods. The first example involves chemometrics data for which ignoring the subset structure leads to no rank reduction. The second example involves macroeconomic time series data for which consideration of subset structure leads to further rank reduction.

Chemometrics Data. We first consider a multivariate chemometrics data set obtained from simulation of a low-density polyethylene tubular reactor. The data are from Skager, MacGregor and Kiparissides (1992), who used partial least squares (PLS) multivariate regression modeling applied to these data both for predicting properties of the produced polymer and for multivariate process control. The data were also considered by Breiman and Friedman (1997) and Reinsel (1999) to illustrate the relative performance of different multivariate

prediction methods, and by Reinsel and Velu (1998, Sec. 3.3) to illustrate the use of partial canonical correlation methods in reduced-rank modeling.

The data set consists of $T = 56$ multivariate observations, with $m = 6$ response variables and 22 'original' predictor (or process) variables. The response variables are the following output properties of the polymer produced: y_1, number-average molecular weight; y_2, weight-average molecular weight; y_3, frequency of long chain branching; y_4, frequency of short chain branching; y_5, content of vinyl groups in the polymer chain; y_6, content of vinylidene groups in the polymer chain. The process variable measurements employed consist of the wall temperature of the reactor (x_1) and the feed rate of the solvent (x_2) that also acts as a chain transfer agent, complemented with 20 different temperatures measured at equal distances along the reactor. For interpretational convenience, the response variables y_3, y_5, and y_6 were rescaled by the multiplicative factors 10^2, 10^3, and 10^2, respectively, for the analysis presented here, so that all six response variables would have variability of the same order of magnitude. The predictor variable x_1 was also rescaled by the factor 10^{-1}. The temperature measurements in the temperature profile along the reactor are expected to be highly correlated and therefore methods to reduce the dimensionality and complexity of the input or predictor set of data may be especially useful. Reinsel and Velu (1998, Sec. 3.3) used partial canonical correlation analysis between the response variables and the 20 temperature measurements, given x_1 and x_2, to exhibit that only the first two (partial) canonical variates of the temperature measurements were necessary for adequate representation of the six response variables y_1 through y_6. We denote these two (partial) canonical variates of the temperature measurements as x_3 and x_4, and suppose that these two (partial canonical) variables together with the two original variables x_1 and x_2 represent the set of predictor variables available for modeling of the six response variables, for purposes of illustration of the partially reduced-rank modeling methodology.

We consider a multivariate linear regression model for the kth vector of responses of the form

$$\mathbf{Y}_k = \mathbf{D} + \mathbf{D}\mathbf{X}_k + \boldsymbol{\epsilon}_k, \quad k = 1, \ldots, T, \tag{5.1}$$

with $T = 56$, where $\mathbf{Y}_k = (y_{1k}, \ldots, y_{6k})'$ is a 6×1 vector, $\mathbf{X}_k = (x_{1k}, \ldots, x_{4k})'$ is a 4×1 vector (hence $n = 4$), and $\mathbf{D}$ allows for the 6×1 vector of constant terms in the regression model. For convenience of notation, let $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$ denote the 6×56 and 4×56 data matrices of response and predictor variables, respectively, after adjustment by subtraction of overall sample means. Then the least squares

estimate of the 6×4 regression coefficient matrix $\mathbf{C}$ is

$$\widetilde{\mathbf{C}} = \widetilde{\mathbf{Y}}\widetilde{\mathbf{X}}'(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')^{-1} = \begin{bmatrix} -0.36855 & 5.47679 & 0.03961 & 0.06727 \\ (0.04151) & (0.07792) & (0.00300) & (0.00894) \\ -0.21289 & 27.65198 & 0.10804 & 0.03777 \\ (0.13960) & (0.26210) & (0.01008) & (0.03007) \\ -1.20633 & -0.80070 & -0.04714 & 0.45245 \\ (0.14690) & (0.27570) & (0.01060) & (0.03164) \\ -1.02840 & -1.40210 & 0.19635 & 0.22933 \\ (0.09776) & (0.18350) & (0.00706) & (0.02106) \\ -0.37562 & -0.49059 & 0.07631 & 0.08210 \\ (0.05658) & (0.10620) & (0.00409) & (0.01219) \\ -0.39426 & -0.54438 & 0.07415 & 0.09133 \\ (0.04133) & (0.07759) & (0.00298) & (0.00890) \end{bmatrix},$$

with estimated standard errors of the individual elements of $\widetilde{\mathbf{C}}$ displayed in parentheses below the estimates, and $\widetilde{\mathbf{D}} = [24.525, 18.930, 109.556, 97.631, 41.281, 40.722]'$. The ML estimate $\widetilde{\boldsymbol{\Sigma}} = (1/56)(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{C}}\widetilde{\mathbf{X}})(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{C}}\widetilde{\mathbf{X}})'$ of the 6×6 covariance matrix $\boldsymbol{\Sigma}$ of the errors $\boldsymbol{\epsilon}_k$ has diagonal elements $\tilde{\sigma}_{jj}$, $j = 1, \ldots, 6$, given by 0.026194, 0.296424, 0.328035, 0.145301, 0.048671, 0.025973, with $\log(|\widetilde{\boldsymbol{\Sigma}}|) = -16.6529$. Moderate sample correlations, of the order of 0.5, are found between most of the pairs of residual variables $\hat{\epsilon}_{jk}$, $j = 1, \ldots, 6$, except for residuals $\hat{\epsilon}_{2k}$ for the second response variable y_{2k} which exhibit little correlation with residuals from response variables y_{3k} through y_{6k}. By comparison with these (partial) correlations among the y_{jk} after adjustment for $\mathbf{X}_k$, the original response variables y_{1k} and y_{2k} show a correlation of about 0.985, variables y_{4k}, y_{5k} and y_{6k} form another strongly correlated group with correlations of about 0.975, while y_{3k} has moderate negative correlations with y_{4k}, y_{5k} and y_{6k} of the order of -0.5.

From the values and significance of the elements in $\widetilde{\mathbf{C}}$, the regression coefficients corresponding to variables y_4 through y_6 share strong similarities, indicating that rank one may be possible for this (sub)set of response variables. But with the inclusion of response variables y_1 through y_3, the rank of $\mathbf{C}$ might be four, that is, full rank. In fact, a LR test of $\text{rank}(\mathbf{C}) \leq 3$ versus $\text{rank}(\mathbf{C}) = 4$ gives a chi-squared test statistic value of $\mathcal{M} = 10.172$ with 3 degrees of freedom, so a hypothesis of reduced rank of three or less for $\mathbf{C}$ is rejected. Moreover, the estimated vectors $\hat{\ell}'_j = \widehat{\mathbf{V}}'_j \widetilde{\boldsymbol{\Sigma}}^{-1/2}$, $j = 5, 6$, associated with the zero eigenvalues of $\widehat{\mathbf{R}} = \widetilde{\boldsymbol{\Sigma}}^{-1/2}\widetilde{\mathbf{C}}\widehat{\boldsymbol{\Sigma}}_{xx}\widetilde{\mathbf{C}}'\widetilde{\boldsymbol{\Sigma}}^{-1/2}$ are found to be $\hat{\ell}'_5 = [0.23983, -0.05415, 0.01428, -2.20349,$

3.73557, 1.95009] and $\hat{\ell}_6' = [-0.24202,\ 0.04324,\ 0.12969,\ 0.67320,\ 4.13553,$ $-5.88991]$. These vectors have the feature that coefficients corresponding to the first three response variables y_1, y_2, and y_3 are close to zero, with relatively large nonzero coefficients corresponding to variables y_4, y_5, and y_6. Based on the discussion at the beginning of Section 4, this feature is highly suggestive of a reduced-rank structure for the coefficient matrix associated with variables y_4 through y_6. The stepwise backward-elimination procedure discussed in Section 4 also performs quite well in this example. In particular, at the first stage the LR statistics $\mathcal{M}$ for testing $\mathrm{rank}(\mathbf{C}_1) \leq 3$ give extremely small and nonsignificant values of 0.025, 0.024, and 0.243, each with 2 degrees of freedom, when the single response variable y_{1k}, y_{2k}, or y_{3k}, respectively, is excluded from $\mathbf{Y}_k$. As the procedure continues, it leads to clearly retaining only the variables y_{4k}, y_{5k}, and y_{6k} in $\mathbf{Y}_{1k}$ and identification of coefficient structure of rank one for this subset.

We now rearrange the response variables with $\mathbf{Y}_{1k} = (y_{4k}, y_{5k}, y_{6k})'$ and $\mathbf{Y}_{2k} = (y_{1k}, y_{2k}, y_{3k})'$, and let $\mathbf{C}_1$ denote the upper 3×4 submatrix of $\mathbf{C}$ corresponding to the (rearranged) response variables in $\mathbf{Y}_{1k}$. The LR test of H_0 : $\mathrm{rank}(\mathbf{C}_1) \leq 2$ gives the test statistic value of $\mathcal{M} = 0.032$ with 2 degrees of freedom, and the LR test of H_0 : $\mathrm{rank}(\mathbf{C}_1) \leq 1$ gives the test statistic value of $\mathcal{M} = 2.285$ with 6 degrees of freedom, so the hypothesis of reduced rank of one for $\mathbf{C}_1$ is quite acceptable. The hypothesis that $\mathbf{C}_1 = 0$ has a LR statistic value of $\mathcal{M} = 201.310$ with 12 degrees of freedom, so we should clearly retain the rank one hypothesis. We mention that the squared sample canonical correlations between $\mathbf{Y}_{1k}$ and $\mathbf{X}_k$ (adjusting for sample means) are $\hat{\rho}_1^2 = 0.97981$, $\hat{\rho}_2^2 = 0.04322$, and $\hat{\rho}_3^2 = 0.00063$.

To obtain the ML estimate $\widehat{\mathbf{C}}_1 = \widehat{\mathbf{A}}\widehat{\mathbf{B}}$ under the rank one hypothesis, we find the normalized eigenvector of the matrix $\widehat{\mathbf{R}}_1 = \widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}\widetilde{\mathbf{C}}_1\widehat{\boldsymbol{\Sigma}}_{xx}\widetilde{\mathbf{C}}_1'\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}$ associated with the largest eigenvalue $\hat{\lambda}_1^2 \equiv \hat{\rho}_1^2/(1 - \hat{\rho}_1^2) = 48.52315$. This normalized eigenvector is $\widehat{\mathbf{V}}_1 = [0.77466,\ 0.23619,\ 0.58661]'$. So as in (2.4) we compute $\widehat{\mathbf{A}} = \widetilde{\boldsymbol{\Sigma}}_{11}^{1/2}\widehat{\mathbf{V}}_1$ and $\widehat{\mathbf{B}} = \widehat{\mathbf{V}}_1'\widetilde{\boldsymbol{\Sigma}}_{11}^{-1/2}\widetilde{\mathbf{C}}_1$ to obtain $\widehat{\mathbf{A}} = [0.321726,\ 0.126565,\ 0.121561]$, $\widehat{\mathbf{B}} = [-3.22504,\ -4.42765,\ 0.610427,\ 0.731985]$, so that the reduced-rank ML estimate of $\mathbf{C}_1$ is

$$\widehat{\mathbf{C}}_1 = \widehat{\mathbf{A}}\widehat{\mathbf{B}} = \begin{bmatrix} -1.03758 & -1.42449 & 0.19639 & 0.23550 \\ -0.40818 & -0.56039 & 0.07726 & 0.09264 \\ -0.39204 & -0.53823 & 0.07420 & 0.08898 \end{bmatrix}.$$

This rank-one estimate quite accurately recovers the corresponding LS estimates, displayed previously as the last three rows of $\widetilde{\mathbf{C}}$ in the original ordering of response variables. The associated ML estimate of $\mathbf{C}_2$ (last three rows of $\mathbf{C}$ under

rearrangement) is

$$\widehat{\mathbf{C}}_2 = \begin{bmatrix} -0.37690 & 5.45910 & 0.03987 & 0.06973 \\ -0.21307 & 27.65157 & 0.10805 & 0.03785 \\ -1.26164 & -0.91897 & -0.04551 & 0.47003 \end{bmatrix}.$$

The corresponding ML estimate $\widehat{\boldsymbol{\Sigma}}$ of the error covariance matrix $\boldsymbol{\Sigma}$ has diagonal elements, corresponding to the original ordering of the response variables, of 0.026266, 0.296424, 0.331456, 0.145592, 0.049883, 0.026013, with $\log(|\widehat{\boldsymbol{\Sigma}}|) = -16.6081$, which are very close to values from $\widetilde{\boldsymbol{\Sigma}}$ under (full-rank) LS estimation. For graphical illustration in support of the reduced-rank feature among the subset of response variables y_4, y_5 and y_6, Figure 1 displays scatter plots of each of the response variables against the single predictive index variable $x_{1k}^* = \widehat{\mathbf{B}}\mathbf{X}_k$ determined from the rank one modeling of the coefficient matrix $\mathbf{C}_1$ for the variables y_4, y_5 and y_6. (The linear fits of each of y_4, y_5 and y_6 with x_1^* obtained from the reduced-rank model are also indicated in the graphs.) This illustration confirms that each of y_4, y_5 and y_6 has a quite strong linear relationship with the single index x_1^*, whereas response variables y_1, y_2 and y_3 do not show any particular relationship with x_1^*, consistent with the modeling results.

UK Macroeconomic Data. We now consider UK macroeconomic data originally presented and analyzed by Klein, Ball, Hazlewood and Vandome (1961). The data were also previously considered by Gudmundsson (1977) and Reinsel and Velu (1998, Chap.4). Klein, Ball, Hazlewood and Vandome presented a detailed description of an econometric model based on an extensive data base (more than twenty macroeconomic time series) including the data used in our analysis. The endogenous (response) variables considered are $y_1 =$ index of industrial production, $y_2 =$ consumption of food, drinks, and tobacco at constant prices, $y_3 =$ total unemployment, $y_4 =$ index of volume of total imports, $y_5 =$ index of volume of total exports. The exogenous variables are $x_1 =$ total civilian labor force, $x_2 =$ index of weekly wage rates, $x_3 =$ price index of total imports, $x_4 =$ price index of total exports, $x_5 =$ price index of total consumption. The relationships between these two sets of variables can be taken to reflect the demand side of the macrosystem of the UK economy. The observations are quarterly for 1948-1956. The first three response variable series are seasonally adjusted, and time series plots of the resulting endogenous and exogenous variables are given by Reinsel and Velu (1998, p.96).

From preliminary LS regression of each response variable y_{ik} on the predictor variables x_{ik}, some moderate degree of autocorrelation is noted in the residuals, particularly for the second and third response variables. Therefore, based on

LS estimation of a model which included a lagged response variable in these two equations, in the analysis described below we use the 'adjusted' response variables $y_{2k}^* = y_{2k} + 0.6y_{2,k-1}$ and $y_{3k}^* = y_{3k} - 0.635y_{3,k-1}$ to account for autocorrelation.
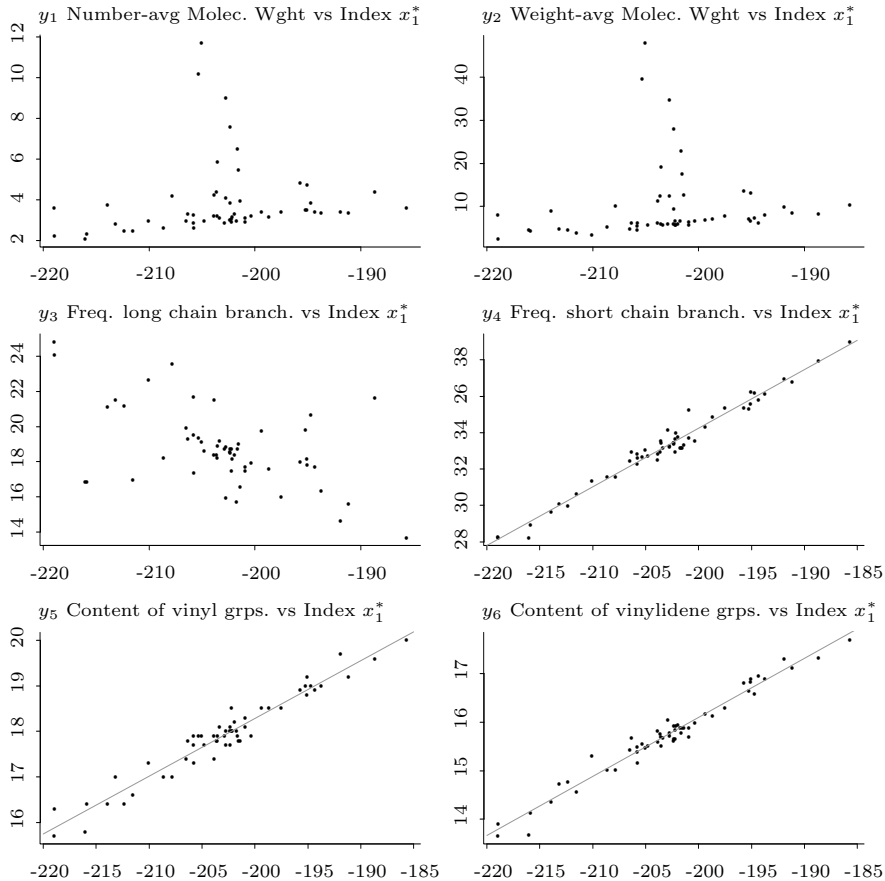
PSfrag replacements

Figure 1. Response variables $y_1, \ldots, y_6$ for the chemometrics data versus the single index predictor variable $x_1^* = \widehat{\mathbf{B}}\mathbf{X}$ determined from the partially reduced-rank regression model.

Then we consider a multivariate linear regression model of the form $\mathbf{Y}_k = \mathbf{D} + \mathbf{C}\mathbf{X}_k + \boldsymbol{\epsilon}_k$, $k = 1, \ldots, T$, with $T = 36$, where $\mathbf{Y}_k = (y_{1k}, y_{2k}^*, y_{3k}^*, y_{4k}, y_{5k})'$, $\mathbf{X}_k = (x_{1k}, \ldots, x_{5k})'$, and $\mathbf{D}$ allows for the 5×1 vector of constant terms in the regression model. For convenience of notation, let $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{X}}$ denote the 5×36 data matrices of response and predictor variables, respectively, after adjustment by subtraction of overall sample means. The least squares estimate of the 5×5

regression coefficient matrix $\mathbf{C}$ is

$$\widetilde{\mathbf{C}} = \tilde{\mathbf{Y}}\tilde{\mathbf{X}}'(\tilde{\mathbf{X}}\tilde{\mathbf{X}}')^{-1} = \begin{bmatrix} 5.5010* & -0.8392* & 0.3044 & -0.8910* & 2.0452* \\ 1.9134* & -0.1506 & 0.2745* & -0.5852* & 0.9334* \\ -2.7096 & 0.6824 & -1.2594* & 2.4762* & -2.5368* \\ 4.4956* & -0.5684 & 0.4521 & -0.8408* & 1.5797* \\ 8.9506* & -0.1536 & 0.4797 & -0.4363 & -0.1680 \end{bmatrix},$$

where the entries with asterisk indicate that the usual t-ratios of these estimates are greater than 1.65 in absolute value. The ML estimate $\widetilde{\mathbf{\Sigma}} = (1/36)(\tilde{\mathbf{Y}} - \widetilde{\mathbf{C}}\tilde{\mathbf{X}})(\tilde{\mathbf{Y}} - \widetilde{\mathbf{C}}\tilde{\mathbf{X}})'$ of the 5×5 covariance matrix $\mathbf{\Sigma}$ of the errors ϵ_k has diagonal elements $\tilde{\sigma}_{jj}$, $j = 1, \ldots, 5$, given by 10.0280, 2.7598, 69.5145, 25.1619, 34.1522, with $\log(|\widetilde{\mathbf{\Sigma}}|) = 13.5484$.

From the values and significance of the elements in $\widetilde{\mathbf{C}}$, the regression coefficients corresponding to certain of the response variables y_{ik} share strong similarities. For example, possible similarities among y_{1k}, y_{2k}^*, and y_{4k}, indicate that reduced rank may be possible for the set of response variables and/or for some subset of these variables. We perform LR tests of rank$(\mathbf{C}) \leq r$ for $r = 4, 3, 2, 1$. For reference, the squared sample canonical correlations between $\mathbf{Y}_k$ and $\mathbf{X}_k$ (adjusting for sample means) are $\hat{\rho}_1^2 = 0.96921$, $\hat{\rho}_2^2 = 0.40003$, $\hat{\rho}_3^2 = 0.31975$, $\hat{\rho}_4^2 = 0.11723$, and $\hat{\rho}_5^2 = 0.01675$. For $r = 3$ the LR test gives a chi-squared test statistic value of $\mathcal{M} = 4.176$ with 4 degrees of freedom, so a hypothesis of reduced rank of three or less for $\mathbf{C}$ is clearly not rejected. For $r = 2$, the value $\mathcal{M} = 15.543$ with 9 degrees of freedom is obtained, which hints that a reduced rank of two could even be possible. For the present, however, we adopt the more conservative conclusion that rank$(\mathbf{C}) \leq 3$ with $r = 3$.

ML estimates of the matrix factors $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ in $\widehat{\mathbf{C}} = \widehat{\mathbf{A}}\widehat{\mathbf{B}}$ under the rank 3 model are obtained in the standard way, similar to (2.4). These estimates are

$$\widehat{\mathbf{A}}' = \begin{bmatrix} 2.144* & 1.292* & -0.574* & 1.966* & 2.230* \\ 0.907 & -0.001 & -3.829* & 1.044 & 4.964* \\ 0.631 & 0.785* & -7.272* & -0.073 & -2.049 \end{bmatrix},$$

$$\widehat{\mathbf{B}} = \begin{bmatrix} 1.845 & -0.187 & 0.125 & -0.303 & 0.669 \\ 0.875 & 0.015 & 0.090 & -0.069 & -0.171 \\ -0.274 & -0.060 & 0.117 & -0.273 & 0.344 \end{bmatrix},$$

and the corresponding ML estimate $\widehat{\mathbf{\Sigma}}$ of the error covariance matrix $\mathbf{\Sigma}$ has diagonal elements $\hat{\sigma}_{jj}$, $j = 1, \ldots, 5$, given by 10.5321, 2.8166, 69.6403, 25.6354, 34.1685, quite similar to the full-rank LS results, with $\log(|\widehat{\mathbf{\Sigma}}|) = 13.6900$. For

reference, ML estimates under the rank 2 model ($r = 2$) would merely consist of the first two columns and first two rows of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$, respectively. Notice that, in particular, from the estimate $\widehat{\mathbf{A}}$ there may still be some strong similarities among its rows of coefficients, except that the third row of estimates, corresponding to response variable y_3^*, does seem to show substantial differences from the remaining rows. We examine the estimated vectors $\hat{\ell}_j' = \widehat{\mathbf{V}}_j' \widetilde{\mathbf{\Sigma}}^{-1/2}$, $j = 4, 5$, associated with the 'near' zero eigenvalues $\hat{\lambda}_j^2 = \hat{\rho}_j^2/(1 - \hat{\rho}_j^2)$ of $\widehat{\mathbf{R}} = \widetilde{\mathbf{\Sigma}}^{-1/2} \widetilde{\mathbf{C}} \widehat{\mathbf{\Sigma}}_{xx} \widetilde{\mathbf{C}}' \widetilde{\mathbf{\Sigma}}^{-1/2}$ (such that $\hat{\ell}_j' \widetilde{\mathbf{C}} \approx \mathbf{0}$), which are $\hat{\ell}_4' = [-0.3362, \ 0.4722, \ 0.0029, \ -0.0199, \ 0.0678]$ and $\hat{\ell}_5' = [0.1334, \ 0.0432, \ 0.0109, \ -0.2004, \ 0.0262]$. These vectors have the feature that coefficients corresponding to the response variable y_3^* are close to zero, there are only moderate values corresponding to variable y_5, and some relatively large nonzero coefficients corresponding to variables y_1, y_2^*, and y_4. Based on the discussion at the beginning of Section 4, the above features are highly suggestive of a further reduced-rank structure for the coefficient matrix associated with variables y_1, y_2^*, y_4, and y_5, with variable y_3^* excluded. The stepwise backward-elimination procedure discussed in Section 4 also performs quite well in revealing this feature in this example. In particular, at the first stage the LR statistic $\mathcal{M}$ for testing $\text{rank}(\mathbf{C}_1) \leq 2$ gives a relatively small and nonsignificant value of 4.363 with 6 degrees of freedom, when the single response variable y_{3k}^* is excluded from $\mathbf{Y}_k$, and this LR test statistic value is smaller than when any of the other single response variables y_{ik}, $i \neq 3$ is excluded.

We rearrange the response variables with $\mathbf{Y}_{1k} = (y_{1k}, y_{2k}^*, y_{4k}, y_{5k})'$ and $\mathbf{Y}_{2k} = (y_{3k}^*)$, and let $\mathbf{C}_1$ denote the upper 4×5 submatrix of $\mathbf{C}$ corresponding to the (rearranged) response variables in $\mathbf{Y}_{1k}$. As indicated, the LR test of $H_0 : \text{rank}(\mathbf{C}_1) \leq 2$ gives the test statistic value of $\mathcal{M} = 4.363$ with 6 degrees of freedom, whereas the LR test of $H_0 : \text{rank}(\mathbf{C}_1) \leq 1$ gives $\mathcal{M} = 18.542$ with 12 degrees of freedom. The hypothesis of reduced rank of two for $\mathbf{C}_1$ is clearly acceptable, while reduced rank of one might also be plausible but again the more conservative value $r_1 = 2$ is taken. ML estimates of the factors $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ in $\widehat{\mathbf{C}}_1 = \widehat{\mathbf{A}} \widehat{\mathbf{B}}$ for this subset reduced-rank model under $r_1 = 2$ are obtained from (2.4) and, in particular, the estimate $\widehat{\mathbf{A}}$ is

$$\widehat{\mathbf{A}}' = \begin{bmatrix} 2.372* & 1.429* & 2.172* & 2.475* \\ 0.391 & -0.439 & 0.837 & 5.260* \end{bmatrix}.$$

From these results we see that the coefficient estimates for the second predictive factor (second column of $\widehat{\mathbf{A}}$) are generally small (and nonsignificant) for response variables y_{1k}, y_{2k}^*, and y_{4k}, but much more substantial for the last response variable y_{5k}. This suggests an even further reduced-rank feature for the coefficient

matrix associated with y_1, y_2^*, and y_4, with variable y_5 excluded (in addition to the previously exclude y_3^*). Thus we entertain continuation of the backward elimination procedure for reduced rank, leading to eliminating y_{5k} and retaining only y_{1k}, y_{2k}^*, and y_{4k} in $\mathbf{Y}_{1k}$, and identification of coefficient structure of rank one for this subset.

Finally, we rearrange the response variables with $\mathbf{Y}_{1k} = (y_{1k}, y_{2k}^*, y_{4k})'$, $\mathbf{Y}_{2k} = (y_{3k}^*, y_{5k})'$, and let $\mathbf{C}_1$ denote the upper 3×5 submatrix of $\mathbf{C}$ corresponding to the (rearranged) response variables in $\mathbf{Y}_{1k}$. The squared sample canonical correlations between $\mathbf{Y}_{1k}$ and $\mathbf{X}_k$ (adjusting for sample means) are $\hat{\rho}_1^2 = 0.96244$, $\hat{\rho}_2^2 = 0.16460$, and $\hat{\rho}_3^2 = 0.02684$. The LR test of $H_0 : \mathrm{rank}(\mathbf{C}_1) \leq 1$ gives $\mathcal{M} = 6.315$ with 8 degrees of freedom, so the hypothesis of reduced rank of one for $\mathbf{C}_1$ is clearly acceptable. We obtain the ML estimate $\widehat{\mathbf{C}}_1 = \widehat{\mathbf{A}}\widehat{\mathbf{B}}$ under this rank one model, with $\widehat{\mathbf{A}} = [2.3808*, \ 1.4350*, \ 2.1792*]'$ and $\widehat{\mathbf{B}} = [1.7153, \ -0.1984, \ 0.1738, \ -0.3951, \ 0.7226]$. The associated reduced rank-one ML estimate $\widehat{\mathbf{C}}_1$ rather accurately represents the corresponding LS estimates of $\mathbf{C}$, displayed previously as the first, third, and fourth rows of $\widetilde{\mathbf{C}}$ in the original ordering of response variables. The associated ML estimate of $\mathbf{C}_2$ (last two rows of $\mathbf{C}$ under rearrangement) is obtained from (2.6) as

$$\widehat{\mathbf{C}}_2 = \begin{bmatrix} -2.4660 & 0.6212 & -1.2741 & 2.4845 & -2.4863 \\ 7.7977 & 0.1343 & 0.5454 & -0.4755 & -0.4024 \end{bmatrix}.$$

The corresponding ML estimate $\widehat{\boldsymbol{\Sigma}}$ of the error covariance matrix $\boldsymbol{\Sigma}$ has diagonal elements, corresponding to the original ordering of the response variables, of 10.6964, 2.8908, 69.5369, 26.3046, 34.6696, with $\log(|\widehat{\boldsymbol{\Sigma}}|) = 13.7554$, which are close to values from $\widetilde{\boldsymbol{\Sigma}}$ under (full-rank) LS estimation and from the earlier 'general' reduced rank 3 model for $\mathbf{C}$. Taking into account the magnitudes of variation of the five predictor variables and (approximate) standard errors of elements of $\widehat{\mathbf{B}}$, the single predictive index variable $x_{1k}^* = \widehat{\mathbf{B}}\mathbf{X}_k$ in the above rank one subset model is composed most prominently of strong contributions from x_{1k}, x_{4k}, and x_{5k}, with much lesser relative weight given to x_{2k} and x_{3k}. Graphical illustration in support of the reduced-rank one model feature can be provided (similar to Figure 1, but not displayed) by scatter plots of each of the response variables against the single predictive index variable $x_{1k}^* = \widehat{\mathbf{B}}\mathbf{X}_k$ determined from the rank one modeling of the coefficient matrix $\mathbf{C}_1$ for the variables y_1, y_2^*, and y_4. Such illustration further confirms that each of y_1, y_2^*, and y_4 has a quite strong linear relationship with the single index x_1^*, whereas response variables y_3^* and y_5 (especially y_3^*) do not show as strong of relationship with x_1^* because they are also influenced by other factors within the set of predictor variables.

6. Discussion and Extension

The model considered in this paper might be viewed as complementary to the original reduced-rank model of Anderson (1951), wherein for the regression of $\mathbf{Y}_k$ the set of *predictor* variables $\mathbf{X}_k$ was separated into a subset $\mathbf{X}_{1k}$ whose coefficient matrix was taken to be of reduced rank and another subset $\mathbf{X}_{2k}$ with full-rank coefficient matrix. This division into separate sets could be based on knowledge of the subject matter. In the current model the role is reversed in that the set of response variables is divided into one subset $\mathbf{Y}_{1k}$ having reduced-rank coefficient matrix in its regression on $\mathbf{X}_k$ and so being influenced by only a small number of predictive variables constructed as linear combinations of $\mathbf{X}_k$, and another subset $\mathbf{Y}_{2k}$ having full-rank coefficient matrix with separate predictors that are linearly independent of those influencing $\mathbf{Y}_{1k}$.

An extended version of the original reduced-rank model was considered by Velu (1991), where each subset $\mathbf{X}_{1k}$ and $\mathbf{X}_{2k}$ of the predictors has coefficient matrix of reduced rank and hence each subset can be represented by a smaller number of linear combinations to describe the relationship with $\mathbf{Y}_k$. As an analogous extension of the current 'partially' reduced-rank model introduced in Section 1, one could also easily envision situations where the two subsets $\mathbf{Y}_{1k}$ and $\mathbf{Y}_{2k}$ of response variables each have separate reduced rank structures, instead of only $\mathbf{Y}_{1k}$ having reduced rank. As a motivating example, in an economic system it might be postulated that a certain subset of the endogenous (response) economic variables are influenced by only a few indices of the exogenous (predictor) economic variables in the system, while another subset of the endogenous variables is influenced only by a few different indices of the exogenous variables. Hence, each subset is influenced by a small number of different (linearly independent) linear combinations of $\mathbf{X}_k$. Such a model adds flexibility to the current model and may be useful in various applications. It would be interesting to explore the statistical procedures of ML estimation and LR testing for ranks associated with models of this form. In particular, we anticipate that simultaneous ML estimation of the two separate components of reduced-rank structure will involve a feature, which we might refer to as "seemingly unrelated reduced-rank regression", analogous to the aspect that occurs in simultaneous estimation of "seemingly unrelated (full-rank) regression" systems.

Tarpey of Wright State University for initially bringing the topic discussed in this paper to their attention.

References

Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series. *J. Amer. Statist. Assoc.* **83**, 849-856.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22**, 327-351.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd edition. Wiley, New York.

Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions. *Ann. Statist.* **27**, 1141-1154.

Breiman, L. and Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 3-54.

Bura, E. and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *J. Roy. Statist. Soc. Ser. B* **63**, 393-410.

Cook, R. D. and Setodji, C. M. (2003). A model-free test for reduced rank in multivariate regression. *J. Amer. Statist. Assoc.* **98**, 340-351.

Gudmundsson, G. (1977). Multivariate analysis of economic variables. *Appl. Statist.* **26**, 48-59.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal.* **5**, 248-264.

Klein, L. R., Ball, R. J., Hazlewood, A. and Vandome, P. (1961). *An Econometric Model of the United Kingdom.* Blackwell, Oxford.

Reinsel, G. C. (1999). On multivariate linear regression shrinkage and reduced-rank procedures. *J. Statist. Plann. Inference* **81**, 311-321.

Reinsel, G. C. and Velu, R. P. (1998). *Multivariate Reduced-Rank Regression, Theory and Applications.* Springer-Verlag, New York.

Skager, B., MacGregor, J. and Kiparissides, C. (1992). Multivariate data analysis applied to lowdensity polyethylene reactors. *Chemometr. Intell. Lab. Syst.* **14**, 341-356.

Velu, R. P. (1991). Reduced rank models with two sets of regressors. *Appl. Statist.* **40**, 159-170.

Department of Statistics, University of Wisconsin-Madison.

E-mail: reinsel@stat.wisc.edu

Whitman School of Management, 721 University Avenue, Syracuse University, Syracuse, NY 13244.

E-mail: rpvelu@syr.edu