Estimation Strategies for the Regression Coefficient Parameter Matrix in Multivariate Multiple Regression

Séverien Nkurunziza* and S. Ejaz Ahmed†

Abstract

We consider some estimation strategies for the parameter matrix in multivariate multiple regression models under a general and natural linear constraint. In the context of two competing models where one model includes all predictors and the other restricts variable coefficients to a candidate linear subspace based on prior information. In this scenario, we suggest some shrinkage estimators for the targeted parameter matrix. The goal of this paper is to examine the relative performances of the suggested estimators in the direction of the subspace and candidate subspace restricted type estimators. We develop a large sample theory for the estimators including derivation of asymptotic bias and asymptotic distributional risk of the suggested estimators. Further, we conduct intensive Monte Carlo simulation studies to appraise the relative performance of the suggested estimators with the classical estimators. Our large-sample and simulation experiments show that the shrinkage estimator overall performs best. The methods are also applied on two real data sets for illustrative purposes. An interesting and unusual feature of this paper is that no assumptions are made concerning the distribution of the error term. Further, it is not assumed that error terms are either identically or independently distributed.

Keywords: Asymptotic distribution risk; Multivariate multiple regression; Quasi-likelihood estimator; Shrinkage strategies; Simulation.

*University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4. Email: severien@uwindsor.ca
†University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4. Email: seahmed@uwindsor.ca
1 Introduction

In a host of scientific research, the basic goal is to assess the simultaneous influence of several covariates on the response variable: the quantity of interest. Multiple regression models provide an extremely powerful methodology to achieve this task. The multivariate multiple regression model (MMRM) generalizes the multiple regression model for the prediction of several response variables from the same set of explanatory variables. Some of the recent advances in multivariate analysis include artificial intelligence and machine learning theory (see for example Izenman, 2008). A common example of multivariate responses occurs in classification and discrimination problems. The mostly used estimation methods are the multivariate least squares estimation and the multivariate least absolute estimation. These estimation methods apply the univariate least squares estimation method and the least absolute estimation method to each individual response variables and lump results in the parameter matrix. However, these methods do not employ the relationship of the response variables in the estimation process. There are many situations where the correlation of the response variables is measurably high. Systolic blood pressure and diastolic blood pressure from a patient would be positively correlated. The correlation coefficient of the number of cavities in the upper jaw and the lower jaw is supposed to be high. In this context, Breiman and Friedman (1997) suggested a shrinkage multivariate least squares estimator through canonical analysis to utilize the relationship of response variables. Further, it finds the optimal extent of that shrinkage. The simulation comparison and real data examples corroborate the contention that the proposed approach is optimal. In this paper, we concentrate on the direction, as opposed to the norm, of the shrinkage. In nonparametric and semi-parametric regression analysis, Ahmed et al. (2006, 2007) and others suggested that we should use subspaces in the problem where they naturally present themselves. We are interested in establishing an optimal estimation strategy for the parameter matrix in multivariate multiple regression models when the parameter is suspected to satisfy certain constraint.

Let $Y$ be a $T \times m$-random matrix, $X$ be a $T \times k$-random matrix, $U$ be a $T \times m$-random matrix, and $\beta$ be a $k \times m$ coefficients parameter matrix. Then, the multivariate linear regression
model is

\[ Y = X\beta + U, \]

(1.1)

where the random matrices \( Y \) and \( X \) are observed and the matrix \( U \) is the unobserved noise. Further, \( \beta \) is the parameter matrix of the interest. Noting that no assumptions are made concerning distribution of the error term or that they are either identically or independently distributed. However, in many practical situations, a more restricted model is appropriate rather than one with some nuisance predictor variables. In some situations, we might need the elements of regression coefficients matrix in the so-called full model to satisfy a set of known linear constraints. Consequently, the model selection process changes the properties of the standard inferential procedures. The regression coefficient obtained after model selection (either by human or by machine) are biased. In other words, bias caused by a misspecified model should not be ignored. These issues are well summarized in Leeb and Potscher (2005). Interestingly, Hand (2006) argues that the incremental improvement in predictive performance that more complex methods sometimes enjoy proves illusory in the face of real problems.

### 1.1 Objective of the paper

Our goal in this paper is to consider inference problem of the parameter matrix \( \beta \) under a very general linear candidate subspace such as

\[ L_1\beta L_2 = d, \]

(1.2)

where \( L_1 \) and \( L_2 \) are respectively \( p \times k \) and \( m \times q \)-known matrix full rank with \( p < k \) and \( q \leq m \), and \( d \) is a \( p \times q \)-known matrix.

A useful compromising method for tackling the uncertainty regarding regression coefficients is to implement the shrinkage estimation strategy. In this paper, we consider the estimation problem for MMRM when there are many potential predictor variables, and some of them may not have influence on the response of interest. With this in mind, several authors as Ahmed et al. (2006, 2007, 2009) and Judge and Mittlehammaer (2004) have reappraised
some of the standard shrinkage estimation strategies for parametric, semiparametric, and nonparametric regression models. However, surprisingly, the MMRM model and consequently the matrix parameter estimation is not considered much in the reviewed literature. The goal of this paper is to analyze and provide guidelines to some of the issues involved in the estimation of the matrix parameters in MMRM subject to be over-parameterizations; that is, some unimportant predictor variables (for example lab variables) are included in the preliminary model. For example, amitriptyline is prescribed by a host of physicians as an antidepressant. However, like any other drug, it may be subject to side effects: abnormal blood pressures, irregular heartbeats, irregular waves on electrocardiogram, and several other effects. These variables can be considered as covariates in regression model to predict Total TCAD plasma level and amount of amitriptyline present in patients who admitted to the hospital after an amitriptyline overdose. However, the preliminary analysis indicates that some predictor variables (lab), for example the gender and age, among others may not be a useful predictor. This study was first reported in Rudorfer (1982). The candidate subspace in (1.2) includes many interesting hypotheses and a variety of applications that can be based upon this general set of linear constraints. Further, this constraint can be applied to tackle a variety of experimental design problems, including profile analysis. The similarity of a given number of profiles can be expressed as a set of linear constraint on $\beta$. We can constrain the population treatment mean profiles further, so that not only are they parallel, but also could require them to be “coincidental” (i.e., identical). Capital asset pricing model (CPAM) as described in Lai and Xing (2008, p. 78) can be viewed as a special case of MMRM. In this model an investment with positive values of intercept parameters is considered to perform better than the market. However, it is not possible to outperform the market. Hence, a natural restriction would be to set the intercept vector to a null vector in the regression model. This restriction is called efficient market hypothesis (see Lai and Xing, 2008, p. 77, 275). The present investigation is a result of diverse application of the MMRM. Here, we laid down the strategies for estimating $\beta$, regression coefficient matrix with and without restriction (1.2). More importantly, we suggest how to combine these two strategies in an optimal way. To achieve this goal, we plan to implement a shrinkage strategy based on James-Stein type estimator.
The remainder of this paper is organized as follows. Section 2 considers estimation strategies. In Section 3, we give the main results, and, more specifically, we also present the shrinkage estimator and show its supremacy over the quasi-likelihood (or LSE). Section 4 presents analysis of two real data sets examples to illustrate the suggested methods, and also presents some simulation studies of the ADR results. Section 5 gives concluding remarks; and, two Appendices give some numerical details and technical results.

2 Estimation Strategies

Again, the model of interest is (1.1) $Y = X\beta + U$, where $Y = (Y_1, \ldots, Y_T)'$, $X = (X_1, \ldots, X_T)'$ and $U = (U_1, \ldots, U_T)'$. Note that $Y_i', X_i'$, and $U_i'$ represent the $i^{th}$-row of the matrices $Y$, $X$ and $U$ respectively.

2.1 Quasi-likelihood and Least Square Estimation

Let $\mathcal{F}_i, i = 1, 2, \ldots$ be the $\mathcal{G}$-field generated by $\{X_1, U_1, X_2, U_2, \ldots, X_i, U_i, X_{i+1}\}$ for $i = 1, 2, \ldots$ and let $\mathcal{F}_0$ denote the trivial $\mathcal{G}$-field. Further, let $\mathcal{H}$ be the family of estimating functions

$$\mathcal{H} = \left\{ h : h = \sum_{i=1}^{T} a_i'(Y_i - E(Y_i|\mathcal{F}_{i-1})) \right\}$$

where for each $i = 1, 2, \ldots$ the $m \times 1$ column-vector function $a_i$ is $\mathcal{F}_{i-1}$-measurable and square integrable.

In this paper, we assume that the model in (1.1) satisfies the following regularity conditions.

Assumptions (regularity conditions):

$(\mathcal{A}_1)$ $(U_i, X_i), i = 1, 2, \ldots, T$ is a strongly stationary and ergodic process with $(U_i, X_i)$ and $(U_j, X_j)$ non-correlated for every $i \neq j$;

$(\mathcal{A}_2)$ $E[U_i|\mathcal{F}_{i-1}] = 0$ and $\text{Var}[U_i|\mathcal{F}_{i-1}] = \Sigma; i = 1, 2, \ldots, T$;

$(\mathcal{A}_3)$ For $i = 1, 2, \ldots, T$, $E[||X_i'X_i||] < \infty$ and $E[X_i'X_i] = \Upsilon$, where $\Upsilon$ is a positive definite matrix;
The matrix $X$ is full rank with probability one.

### 2.1.1 Estimation Under Full Parametric Space

Let us denote $\hat{\beta}$ to be the quasi-likelihood estimator (or the ordinary least square estimator (OLS)) of $\beta$.

**Proposition 2.1** Under Assumptions $(\mathcal{A}_1)$-$(\mathcal{A}_4)$, with respect to the family of the estimating equations $\mathcal{H}$, the quasi-likelihood estimator of $\beta$ is $\hat{\beta} = (X'X)^{-1}X'Y$. Furthermore, $\hat{\beta}$ is the ordinary least square estimator (OLSE). □

### 2.1.2 Estimation Under Candidate Parametric Sub-space

Let $\tilde{\beta}$ be the restricted quasi likelihood estimator (RQLE) or be the restricted least square estimator (RLSE) of $\beta$ under the restriction in (1.2).

**Proposition 2.2** Let $J = (X'X)^{-1}L_1'(L_1(X'X)^{-1}L_1')^{-1}$ and let $P = (L_2L_2)^{-1}L_2$. We have $\tilde{\beta} = \hat{\beta} - J(L_1\hat{\beta}L_2 - d)P$. □

**Proof of Proposition 2.2** The restricted estimator of $\beta$ is the solution of the minimization problem $\min_\beta \left\{ ||Y - X\beta||^2 \right\}$ subjected to $L_1\beta L_2 = d$. Then, the Lagrangian function becomes $\mathcal{L}_\lambda(\beta) = \text{trace} \left[ (Y - X\beta)(Y - X\beta)' \right] + \text{trace} \left[ (L_1\beta L_2 - d)\lambda \right]$. Then, by differentiating $\mathcal{L}_\lambda(\beta)$ with respect to $\beta$ and $\lambda$, we get the desired result. □

Under assumed regularity conditions, the strong consistency of $\hat{\beta}$ and $\tilde{\beta}$ can be established.

**Proposition 2.3** Let $W_m(n, \Sigma)$ denote a $n \times n$-random matrix whose distribution is Wishart with parameter $\Sigma$ and degrees of freedom $m$. Also, assume that the model (1.1) and Assumptions $(\mathcal{A}_1)$-$(\mathcal{A}_4)$ hold. Then, $\sqrt{T} \left( \hat{\beta} - \beta \right) \xrightarrow{T \to \infty} \mathcal{N}_{k \times m}(0, \Sigma \otimes \Upsilon)$, and $T \Sigma^{-\frac{1}{2}} \left( \hat{\beta} - \beta \right)' \Upsilon \left( \hat{\beta} - \beta \right) \Sigma^{-\frac{1}{2}} \xrightarrow{T \to \infty} W_k(m, I_k)$. □
2.2 Shrinkage Estimation Strategy (SES)

In this subsection, we show how to improve the predictive performance of the quasi-likelihood estimator (QMLE) by shrinking them towards a restricted estimate. Following Ahmed (2001) and others, we consider two Stein-rule (or shrinkage) estimators of the matrix parameter. The shrinkage estimator (SE) $\hat{\beta}^S$ of $\beta$ is defined as

$$\hat{\beta}^S = \hat{\beta} + \{1 - c \psi^{-1}_T\} (\hat{\beta} - \tilde{\beta}), \quad (2.1)$$

where $c$ is allowed to vary over $[0, 2(pq - 2))$, $pq > 2$, and often is taken as $c = pq - 2$; thus tacitly, we assume that $pq \geq 3$. The weight $(pq - 2)\psi^{-1}_T$ in $\hat{\beta}^S$ is chosen to minimize asymptotic mean squared error (Ahmed (2001)), where

$$\psi_T = \text{trace} \left[ \left( L_2^T \hat{\Sigma} L_2 \right)^{-\frac{1}{2}} L_2^T \xi T L_1 \left( L_1 \hat{Y}^{-1} L_1' \right)^{-1} L_1 \xi T L_2 \left( L_2^T \hat{\Sigma} L_2 \right)^{-\frac{1}{2}} \right], \quad (2.2)$$

with $\hat{Y} = \frac{1}{r} X' X$, and $\hat{\Sigma} = \frac{1}{r} \left( Y - X \hat{\beta} \right)' \left( Y - X \hat{\beta} \right)$. Shrinking towards $\tilde{\beta}$ may be motivated by preliminary testing or empirical Bayes considerations (Judge and Mittlehammer (2004), Ahmed et al. (2007)).

Note that $\psi_T \geq 0$, and hence, for $\psi_T < c \iff 1 - c \psi^{-1}_T < 0$, this causes a possible inversion of sign or over-shrinkage. To avoid this problem, we suggest positive-rule shrinkage estimators (PSE). Formally, the PSE $\left( \hat{\beta}^{S+} \right)$ is defined as

$$\hat{\beta}^{S+} = \hat{\beta} + \{1 - c \psi^{-1}_T\}^+ (\hat{\beta} - \tilde{\beta}), \quad (2.3)$$

where $z^+ = \max(0, z)$. Accordingly, Ahmed (2001) recommended that the shrinkage estimator should be used as a tool for developing the PSE and should not be used as an estimator in its own right. Oman (1997) suggested a similar estimator by shrinking the maximum likelihood estimator towards a restricted estimator.

The finite sample distribution theory of these shrinkage estimators is not simple to obtain. This difficulty has been largely overcome by asymptotic methods (Ahmed and Saleh (1999), Ahmed (2001), and others). These asymptotic methods relate primarily to convergence in distribution which may not generally guarantee convergence in quadratic risk. This technicality has been taken care of by the introduction of asymptotic distributional risk (ADR)
(Ahmed (2001)), which, in turn, is based on the concept of a *shrinking neighborhood* of the pivot for which the ADR serves a useful and interpretable role in *asymptotic risk analysis*.

### 3 Main Results

It is well known that, even for normal distribution, the effective domain of risk dominance of PSE or SE over the QMLE is a small neighborhood of the chosen pivot (viz., $L_1\beta L_2 = d$); and as we make the sample size $T$ larger and larger, this domain becomes narrower. To avoid asymptotic degeneracy we consider the following sequence of alternative restrictions

$$K_T : L_1\beta L_2 = d + \frac{\delta}{\sqrt{T}}, \quad T = 1, 2, \ldots$$  \hspace{2cm} (3.1)

where $\delta$ is a nonzero $p \times q$-matrix. Also, we assume that $\|\delta\| < \infty$ and that the regularities conditions ($A_1$)–($A_4$) are satisfied.

Now we generalize the optimality criterion given in Ahmed (2001) in an effort to accommodate the matrix estimation. For an estimator $\hat{\beta}^*$ of $\beta$, we consider a *quadratic loss function* of the form

$$L(\hat{\beta}^*, \beta; W) = \text{trace} \left\{ \left[ \sqrt{T} L_2' (\hat{\beta}^* - \beta) \right] W \left[ \sqrt{T} (\hat{\beta}^* - \beta) L_2 \right] \right\},$$  \hspace{2cm} (3.2)

where $W$ is a $k \times k$-positive semi-definite (p.s.d) matrix. Let $A$ denote $m \times p$-matrix and let $\text{Vec}(A)$ denote the $np$ column vector obtained by stacking together the columns of $A$ one underneath the other. Using the distribution of $\sqrt{T} (\hat{\beta}^* - \beta) L_2$ and taking the expected value both sides of (3.2), we get the expected loss that would be called the *quadratic risk* $R_T^p (\hat{\beta}^*; \beta; W)$. For instance, if

$$\bar{\Sigma}_1(T) \otimes \bar{\Sigma}_2(T) = TE \left[ \text{vec} \left( \left( \hat{\beta}^* - \beta \right) L_2 \right) \left( \text{vec} \left( \left( \hat{\beta}^* - \beta \right) L_2 \right) \right)' \right],$$

with $\bar{\Sigma}_1(T)$ nonsingular matrix, we get $R_T^p (\hat{\beta}^*; \beta; W) = \text{trace} \left( W \bar{\Sigma}_2(T) \right) \text{trace} \left( \bar{\Sigma}_1(T) \right)$. Thus, whenever $\lim_{T \to \infty} \bar{\Sigma}_1(T) = \Sigma_{01}$ and $\lim_{T \to \infty} \bar{\Sigma}_2(T) = \Sigma_{02}$ exists,

$$R_T^p (\hat{\beta}^*; \beta; W) \xrightarrow{T \to \infty} R^p (\hat{\beta}^*; \beta; W) = \text{trace} \left( W \Sigma_{02} \right) \text{trace} (\Sigma_{01}),$$
which is termed the *asymptotic risk*. In our setup, we denote the distribution of \( \sqrt{T} (\hat{\beta}^* - \beta) L_2 \) by \( \tilde{G}_T(u), u \in \mathbb{R}^k \times \mathbb{R}^q \). Suppose that \( \tilde{G}_T \rightarrow \tilde{G} \) (at all points of continuity), as \( T \rightarrow \infty \), and let \( \Sigma_G^{(1)} \otimes \Sigma_G^{(2)} = \int \ldots \int x x' d\tilde{G}(x) \) with \( \Sigma_G^{(1)} \) and \( \Sigma_G^{(2)} \) respectively \( k \times k \) and \( q \times q \)-matrices. Then the ADR of \( \hat{\beta}^* \) is defined as

\[
R^\circ \left( \hat{\beta}^*, \beta; W \right) = \text{trace} \left( W \Sigma_G^{(2)} \right) \text{trace} \left( \Sigma_G^{(1)} \right). \tag{3.3}
\]

The shrinkage estimators are, in general, biased estimators. Nevertheless, the bias is accompanied by reduction in the ADR, and hence, it does not have a serious impact on ADR assessment. In this vein, we define the asymptotic bias as

\[
B_T^0 \left( \hat{\beta}^*, \beta \right) = \mathbb{E} \left[ \sqrt{T} \left( \hat{\beta}^* - \beta \right) L_2 \right], \tag{3.4}
\]

and side by side, the *asymptotic distributional bias* (ADB) as the limit

\[
B_T^0 \left( \hat{\beta}^*, \beta \right) = \int \ldots \int x d\tilde{G}_T(x) \xrightarrow{T \rightarrow \infty} B \left( \hat{\beta}^*, \beta \right) = \int \ldots \int x d\tilde{G}(x). \tag{3.5}
\]

In an effort to derive the asymptotic distributional properties of the suggested estimation strategies, we give some preliminary result in the following proposition. First, let

\[
J_0 = \Upsilon^{-1} L_1' \left( L_1 \Upsilon^{-1} L_1' \right)^{-1}, \text{let } \delta^* = -J_0 \delta \text{ and let } \Upsilon^* = J_0 L_1 \Upsilon^{-1}. \]

Also, let the quantities

\[
\varrho_T = \sqrt{T} \left( \hat{\beta} - \beta \right), \quad \xi_T = \sqrt{T} \left( \hat{\beta}^* - \beta \right), \quad \text{and} \quad \zeta_T = \sqrt{T} \left( \bar{\beta} - \beta \right).
\]

**Proposition 3.1** If Proposition 2.3 holds, then, under local alternative \( K_T \) in (3.1), we have

\[
\left( \varrho_T, \xi_T \right) \xrightarrow{T \rightarrow \infty} \mathcal{N}_{2k \times 2m} \left( (0, J_0 \delta P), \begin{pmatrix} \Sigma \otimes \Upsilon^{-1} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right), \quad \text{and}
\]

\[
\left( \xi_T, \zeta_T \right) \xrightarrow{T \rightarrow \infty} \mathcal{N}_{2k \times 2m} \left( (J_0 \delta P, -J_0 \delta P), \begin{pmatrix} \Omega_{22} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{pmatrix} \right),
\]

with \( \Omega_{22} = P' L_2' \Sigma L_2 P \otimes \Upsilon^* \), \( \Omega_{21} = \Omega_{12}' = P' L_2' \Sigma \otimes \Upsilon^* \), \( \Upsilon_{22} = \Sigma \otimes \Upsilon^{-1} - \Omega_{21} - \Omega_{12} + \Omega_{22} \), \( \Upsilon_{21} = \Upsilon_{12}' = \Omega_{12} - \Omega_{22} \). \( \square \)

The proof follows by combining Proposition 2.2 and Proposition 2.3, along with some algebraic computations.
Corollary 3.1 If Proposition 2.3 holds, then, under local alternative $K_T$ in (3.1), we have

$$
(\rho_T L_2, \xi_T L_2) \xrightarrow{T \to \infty} (\theta, \xi) \sim \mathcal{N}_{2k \times 2q} \left( (0, J_0 \delta), \begin{pmatrix} L_2 T \Sigma L_2 \otimes \Upsilon_1 & L_2 T \Sigma L_2 \otimes \Upsilon_2 \\ L_2 T \Sigma L_2 \otimes \Upsilon_3 & L_2 T \Sigma L_2 \otimes \Upsilon_4 \end{pmatrix} \right)
$$

and

$$
(\xi_T L_2, \zeta_T L_2) \xrightarrow{T \to \infty} (\xi, \zeta) \sim \mathcal{N}_{2k \times 2q} \left( (J_0 \delta, -J_0 \delta), \begin{pmatrix} L_2 T \Sigma L_2 \otimes \Upsilon_1 \delta & 0 \\ 0 & \Omega_{33} \end{pmatrix} \right),
$$

with $\Omega_{33} = L_2 T \Sigma L_2 \otimes (\Upsilon_1 - \Upsilon_2)$. □

Two key results for the study of ADR and ADB of the shrinkage estimators are given in Proposition 2.3 and Corollary 3.1. Indeed, from these results, we use the results on the (normal distribution) parametric model, and thereby give the main results of this subsection. To simplify the notation, let

$$
\Delta = \text{trace} \left[ \delta' \left( L_1 T \Upsilon_1^{-1} L_1' \right)^{-1} \delta \left( L_2 T \Sigma L_2 \right)^{-1} \right],
$$

and let $H_{\nu}(x; \Delta) = \mathbb{P} \{ \chi^2_{\nu}(\Delta) \leq x \}, x \in \mathbb{R}^+.$

Theorem 3.1 Assume that Proposition 3.1 holds. Then, the ADB functions of the estimators are given as follows:

$$
B \left( \tilde{\beta}, \beta \right) = 0, \quad B \left( \tilde{\beta}^s, \beta \right) = -\delta^*, \quad B \left( \tilde{\beta}^s, \beta \right) = -\delta^* \left( pq - 2 \right) \mathbb{E} \left\{ \chi^2_{pq+2}(\Delta) \right\}
$$

$$
B \left( \tilde{\beta}^{s+}, \beta \right) = -\delta^* \left[ H_{pq+2}(pq+2; \Delta) + (pq-2) \mathbb{E} \left\{ \chi^2_{pq+2}(\Delta) I \left( \chi^2_{pq+2}(\Delta) > (pq-2) \right) \right\} \right]. \quad (3.6)
$$

□

The proof of this theorem is outlined in the Appendix. It is noticed that in the ADB of $\tilde{\beta}$, $\tilde{\beta}^s$ and $\tilde{\beta}^{s+}$, the component $\delta$ is common and they differ only by scalar factors. Accordingly, it suffices to compare the scalar factors $\Delta$ only. It is clear that bias of the $\tilde{\beta}$ is an unbounded function of $\Delta$. Interestingly, the ADB of both $\tilde{\beta}^s$ and $\tilde{\beta}^{s+}$ are bounded in $\Delta$. Noting that since $\mathbb{E} \left\{ \chi^2_{pq+2}(\Delta) \right\}$ is a decreasing log-convex function of $\Delta$ the ADB of $\tilde{\beta}^s$ starts from the origin at $\Delta = 0$, increases to a maximum, and then decreases towards 0. The characteristic $\tilde{\beta}^{s+}$ is similar to $\tilde{\beta}^s$. The bias curve of $\tilde{\beta}^{s+}$ remains below the curve of SE for all values of $\Delta$. 

9
Theorem 3.2 Assume that Proposition 3.1 holds. Then, the ADR functions of the estimators are given as follows:

\[
\begin{align*}
R(\tilde{\beta}, \beta; W) &= \text{trace}(L'_2 \Sigma L_2)\text{trace}(W'Y^{-1}), \\
R(\hat{\beta}, \beta; W) &= \text{trace}(L'_2 \Sigma L_2)\text{trace}(W'Y^{-1}) - \text{trace}(L'_2 \Sigma L_2)\text{trace}(WY^*) \\
&\quad + \text{trace}(\delta^*W\delta^*), \\
R(\tilde{\beta}^S, \beta; W) &= \text{ADR}(\tilde{\beta}) + \text{trace}(\delta^*W\delta^*) \left((pq)^2 - 4\right)E(\chi_{pq+4}^{-4}(\Delta)) \\
&\quad - (qp - 2)\text{trace}(L'_2 \Sigma L_2)\text{trace}(WY^*) \left\{2E(\chi_{pq+2}^{-2}(\Delta)) - (pq - 2)E(\chi_{pq+2}^{-4}(\Delta))\right\}, \\
R(\tilde{\beta}^{S+}, \beta; W) &= \text{ADR}(\tilde{\beta}^S) + (pq - 2)\text{trace}(L'_2 \Sigma L_2)\text{trace}(WY^*) \\
&\quad \times [2E(\chi_{pq+2}^{-2}(\Delta)I(\chi_{pq+2}^{-2}(\Delta) \leq (pq - 2))] \\
&\quad - (pq - 2)E(\chi_{pq+2}^{-4}(\Delta)I(\chi_{pq+2}^{-4}(\Delta) \leq (pq - 2))] \\
&\quad - \text{trace}(L'_2 \Sigma L_2)\text{trace}(WY^{-1})H_{pq+2}(pq - 2; \Delta) \\
&\quad + \text{trace}(\delta^*W\delta^*) \left\{2H_{pq+2}(pq - 2; \Delta) - H_{pq+4}(pq - 2; \Delta)\right\} \\
&\quad - (pq - 2)\text{trace}(\delta^*W\delta^*) \left[2E(\chi_{pq+2}^{-2}(\Delta))I(\chi_{pq+2}^{-2}(\Delta) \leq (pq - 2))] \\
&\quad - 2E(\chi_{pq+4}^{-2}(\Delta)I(\chi_{pq+4}^{-2}(\Delta) \leq (pq - 2))] \\
&\quad + (pq - 2)E(\chi_{pq+4}^{-4}(\Delta)I(\chi_{pq+4}^{-4}(\Delta) \leq (pq - 2))] \right\}. \tag{3.7}
\end{align*}
\]

The proof of this theorem is outlined in the Appendix. For a suitable choice of the matrix \(W\), ADR dominance of the estimators are similar to those under normal theory and can be summarized as follows.

(i) \[R(\tilde{\beta}^S, \beta; W) < \text{trace}(L'_2 \Sigma L_2)\text{trace}(W'Y^{-1}) \quad \text{for all} \quad \Delta \in [0, \infty).\]

In other words, shrinkage estimator provides greater prediction accuracy than the QMLE/LSE, a gold standard. Note that the ADR function of SE is monotone in \(\Delta\), the smallest value is achieved at \(\Delta = 0\) and the largest is trace\((L'_2 \Sigma L_2)\text{trace}(W'Y^{-1})\). Hence, \(\tilde{\beta}^S\) outperforms \(\tilde{\beta}\) in the entire parameter space induced by the non-centrality parameter \(\Delta\).

(ii) \(\tilde{\beta}^{S+}\) asymptotically superior to \(\tilde{\beta}^S\) for \(\Delta \in [0, \infty)\). Therefore, \(\tilde{\beta}^{S+}\) is also superior to the classical QMLE/LSE.

The simulation results in the next section are also pointing in that direction.
4 Simulation Study and Data Analysis

4.1 Simulation Study

We conduct a Monte Carlo simulation study to evaluate the performance of the suggested method in small and moderate sample sizes. However, to save space, we report the results corresponding to the case where the noise follows multivariate normal distribution with zero mean vector and $\sigma^2I_m (\sigma = 10)$. Also, for the explanatory variable matrix, we consider the cases, and $X$ is either random or non-random.

To appraise the relative performance of the estimators, we apply the notion of “relative efficiency”. The relative efficiency of the estimator $\hat{\beta}^*$ with respect to $\hat{\beta}$ is defined by $\text{RE}(\hat{\beta}^*) = \frac{\text{risk}(\hat{\beta})}{\text{risk}(\hat{\beta}^*)}$. Thus, a relative efficiency larger than one indicates the degree of superiority of the estimator over $\hat{\beta}$.

Case I: $X$ is random  The random matrix $X$ is generated from multivariate normal distribution and then kept fixed with respect to the replications. The sample sizes under consideration are $T = 15, T = 30, T = 40,$ and $T = 50,$ and 1000 replications have been performed. First, we consider $m = 5, k = 5, p = 4,$ and $q = 1$. The vectors/matrix of the candidate sub-space is selected as: $L_2 = (1, 1, 1, 1, 1), d = (1.99, -0.2, -3.41, 0.76)$ and the matrices $L_1$ and $\beta$ are given by relation (B.1) in the Appendix B.

Next, we consider $m = 6, k = 7, p = 6, q = 1$, with $d = (1.99, -0.2, -2.91, 0.26, 2.35, -2.8)$, $L_2 = (1, 1, 1, 1, 1, 1, 1)$, and the matrices $L_1$ and $\beta$ are given in relation (B.2) in the Appendix B.

Case II: $X$ is fixed  In this scenario, we consider $T = r \times k, X = I_k \otimes e_r$ where $e_r$ denotes $r$-column vector with unity entries. Also, we choose $L_2 = e_6, d = 0$ and

$$L_1 = \begin{bmatrix} I_6 & \vdots & -e_6 \end{bmatrix}. \quad (4.1)$$

The restriction $L_1 \beta L_2 = 0$ relates whether or not the population treatment mean profiles are identical. The matrix parameter $\beta$ is chosen the same as in (B.2) for simulation purposes.

The results of the simulation studies are given in Figures 1-3 and can be summarized as follows:
(a) $T = 15, m = 5, k = 5, p = 4$

(b) $T = 30, m = 5, k = 5, p = 4$

(c) $T = 40, m = 5, k = 5, p = 4$

(d) $T = 50, m = 5, k = 5, p = 4$

Figure 1: The efficiency of the estimators relative to QMLE
Figure 2: The efficiency of the estimators relative to QMLE
Figure 3: The efficiency of the estimators relative to QMLE)
(a) The shrinkage estimation strategy is relatively stable and dominates uniformly the QMLE, as theoretically expected.

(b) The risk dominance of the shrinkage estimators over the QMLE is more pronounced as the subspace parameter dimensions are higher.

(c) The efficiency of $\tilde{\beta}$ converges to 0 as the true value of the parameter matrix is far away from the pivot. Nevertheless, around the pivot $\tilde{\beta}$ outperforms over the shrinkage estimators.

Now, we can safely conclude that our simulation study findings strongly corroborate with our theoretical results presented in Section 2.

4.2 Data Analysis

In this subsection, we illustrate the application of the suggested method on two real data sets. 

**Example 1**  The data set is given in Jiang et.al (2008), and it is about the *Microcystis aeruginosa*. In this study, Plackett-Burman design of 10 environmental factors (that play a role of covariates) and two responses which are cell dry weight ($Y_1$) and microcystins concentration ($Y_2$) of *Microcystis aeruginosa* are considered. The 10 co-variates are $NaNO_3$, $K_2HPO_4$, Ferric ammonium citrate, $MgSO_4 \cdot 7H_2O$, $CaCl_2 \cdot 2H_2O$, Trace metal $I^a$, Trace metal $II^b$, Initial medium pH, Temperature, and Light intensity. Jiang et.al (2008) performed two separate multiple regressions analysis, assuming $Y_1$ and $Y_2$ are unrelated. Based on their analysis they concluded that $NaNO_3$, $K_2HPO_4$, iron, light intensity, and temperature have significant effects (at 10% level of significance) on $Y_1$ whereas only $NaNO_3$, iron, and light intensity have significant effects on $Y_2$. However, our preliminary analysis suggest that the two responses $Y_1$ and $Y_2$ are correlated. Indeed, the correlation coefficient between the two responses is 0.61, and at 5 % significance level, the correlation is statistically significant (p-value=0.0014). It would be more appropriate to use a MMRM. The full parametric space model for this data is $Y = X\beta + U$, where $Y = [Y_1, Y_2]$ is a random $24 \times 2$-matrix, and $\beta$ is a $10 \times 2$-matrix of constants, and $X$ represents the design matrix from Table 2 given in Jiang et al. (2008) where the
value -1 is taken as 0. Thus, for this data set, \( T = 24, m = 2, \) and \( k = 10. \) In order to illustrate constrained and shrinkage estimation strategies, one can consider the sub-model by including \( \text{NaNO}_3, \text{K}_2\text{HPO}_4, \) iron, light intensity, and temperature in the model, while the others five are suspected to not have a significant effect on the responses. Hence, to improve the overall estimation of \( \beta, \) one can suggest the following candidate sub-space:

\[
L_1 \beta L_2 = 0_{5\times 2}, \quad \text{with} \quad L_1 = \begin{pmatrix} 0_{5\times 3} & I_5 & 0_{5\times 2} \end{pmatrix}, \quad \text{and} \quad L_2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}. \quad (4.2)
\]

**Example 2** The data for this example is given in Andrews and Herzberg (1985, pp.357-360) and described in Izenman (2008, pp. 193). This data consists of 8 measurements on each of four variates on 13 different types of root-stocks of apple trees. The 4 variates are trunk girth in mm (\( Y_1 \)); extension growth (cm) (\( Y_2 \)) at 4 years after planting; trunk girth (mm) (\( Y_3 \)) at 15 years after planting; and weight (lb) of tree above ground (\( Y_4 \)) at 15 years after planting. The design matrix is a \((4)\)-matrix of 0s and 1s depending upon which tree is derived from which root-stock. The so-called design restriction in this case is so \( L_1 \beta L_2 = 0, \) where \( L_1 = (I_{12}, -e_{12}), \) and \( L_2 = e_4. \) In words, the constraint under consideration corresponds to the case where the population treatment mean profiles are identical. Thus, for this data set, we have, \( T = 104, m = 4, \) and \( k = 13, \) with the design matrix \( X = I_{13} \otimes (1, 1, 1, 1, 1, 1, 1, 1)^t. \) Again, we use the natural restriction to improve the estimation of the remaining parameters of \( \beta, \) and suggest an estimation strategy along with the constrained one. As given in Table 2, the correlation matrix coefficients along with the p-values for testing the non-correlation show that the four responses are highly correlated. Accordingly, four separate regressions will not be appropriate. Applying the suggested methods, we calculate the point estimates of \( \beta \) and the predictive squared error in order to measure the performance of \( \hat{\beta}^{S^+} \) over \( \hat{\beta}. \) In summary, Table 1 shows that the predictive squared error of \( \hat{\beta}^{S^+} \) is smaller than that of \( \hat{\beta}. \) We can safely, conclude that on the whole, the \( \hat{\beta}^{S^+} \) did much better than the usual estimator. This strongly indicates that that we should use the subspaces in our estimation problems where they are naturally available. Oman (1997) demonstrated that the shrinkage estimator performed better than "curds and whey" of Brieman and Freiedman (1997. Hence, this paper reappraises and reinforces the superior performance of the shrinkage estimator.
Table 1: Predictive mean squared error for root-stocks of apple trees data set

<table>
<thead>
<tr>
<th>Response</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\beta}^{S_+}$</th>
<th>% of predictive error reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>27.822</td>
<td>14.513</td>
<td>0.478</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>11.111</td>
<td>2.406</td>
<td>0.783</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>18.554</td>
<td>5.269</td>
<td>0.716</td>
</tr>
<tr>
<td>$Y_4$</td>
<td>8.224</td>
<td>5.174</td>
<td>0.371</td>
</tr>
<tr>
<td>Average</td>
<td>18.093</td>
<td>8.230</td>
<td>0.545</td>
</tr>
</tbody>
</table>

5 Concluding Remarks

Our objective in this paper is to examine the relative performance of the full model QMLE, restricted estimator, and shrinkage estimators in the context of multivariate multiple regression models when we have prior knowledge of a candidate subspace. We examined the risk properties of the estimators in terms of asymptotic and Monte Carlo simulation risk results. We conclude both analytically and numerically that the risk improvement of the restricted QMLE over other estimators is substantial near the restriction. However, the improvement starts diminishing as restriction moves further and further away from assumed subspaces. Thus, the performance of the restricted estimator heavily depends on the quality of the non sample information regarding the subspace. On the other hand, the shrinkage estimators with data based weights outperform the full model QMLE $\hat{\beta}$ in the entire parameter space. The weight of the shrinkage estimator has intuitive appealing property. In summary, we show that large gains of suggested shrinkage approach over ordinary least square or QMLE. The continuing use of least squares seems inexplicable, and the suggested shrinkage estimator demonstrates this convincingly when many parameters are involved. Importantly, the shrinkage estimator is computationally elementary, under-demanding, and can be easily implemented. Further, the suggested approach is free from any tuning parameters, and calculations are not iterative. Finally, the simulation results and the real data example support the contention that the suggested method is superior to classical estimation and the method entails no distributional assumption.
A Appendix

Proof of Proposition 2.1  First, note that the sequence of estimating functions
\[ \left\{ \sum_{i=1}^{T} a_i' (Y_i - \mathbb{E}(Y_i|\mathcal{F}_{i-1})) \right\}, \quad T \geq 1 \]
is an \( \mathcal{F}_T \)-martingale. Further, with respect to the family of estimating functions \( \mathcal{H} \), the quasi-likelihood estimator is the solution of the equation
\[ \sum_{i=1}^{T} a_i' (Y_i - \mathbb{E}(Y_i|\mathcal{F}_{i-1})) = 0 \] (A.1)
where
\[ a_i = \frac{\partial \mathbb{E}(Y_i|\mathcal{F}_{i-1})}{\partial \beta} \left[ \text{Var}(Y_i|\mathcal{F}_{i-1}) \right]^{-1}, \quad i = 1, 2, \ldots, T. \]

Further, under Assumption (A1), we have
\[ \mathbb{E}[Y_i|\mathcal{F}_{i-1}] = X_i \beta, \quad \text{and} \quad \text{Var}[Y_i|\mathcal{F}_{i-1}] = \Sigma. \] (A.2)

Therefore, the relation (A.1) becomes
\[ \sum_{i=1}^{T} X_i' (Y_i - X_i \beta) = 0 \] (A.3)
and then, by using the Assumption (A4), we have
\[ \hat{\beta} = \left( \sum_{i=1}^{T} X_i' X_i \right)^{-1} \sum_{i=1}^{T} X_i' Y_i \iff \hat{\beta} = (X'X)^{-1} X'Y. \] (A.4)

Moreover, from (A.3), it is noticed that the established quasi-likelihood is also the solution of the least square equation
\[ \min_{\beta} \left\{ \| Y - X \beta \|^2 \right\} \]
where the norm \( \| . \| \) is defined through the inner product \( \langle x, y \rangle = \text{trace}(xy') \) where \( x \) and \( y \) are matrices.

Proof of Theorem 3.1  From Corollary 3.1, we get directly the two first statements
\[ B \left( \hat{\beta}, \beta \right) = \mathbb{E}(\rho) = 0, \quad \text{and} \quad B \left( \tilde{\beta}, \beta \right) = \mathbb{E}(\zeta) = -\delta^*. \]

For the third statement, we have,
\[ B \left( \hat{\beta}^S, \beta \right) = \mathbb{E} \left[ \zeta + (1 - (pq - 2)\phi^{-1}) \xi \right] = -\delta^* + \mathbb{E} \left[ (1 - (pq - 2)\phi^{-1}) \xi \right]. \] (A.5)
Further, since $\Upsilon\frac{1}{2}\Upsilon^*\Upsilon^{\frac{1}{2}}$ is a symmetric and idempotent matrix, there exists an orthogonal matrix $Q$ such that

$$Q\Upsilon\frac{1}{2}\Upsilon^*\Upsilon^{\frac{1}{2}}Q' = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}.$$ (A.6)

Moreover, let $\Lambda = (L_2^\prime \Sigma L_2)$ and let $V = \Lambda^{-\frac{1}{2}}\xi\Upsilon^{\frac{1}{2}}Q$. One can verify that

$$\text{Vec}(V) = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \mathcal{N}_{kq} \left( \begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{pq} & 0 \\ 0 & 0 \end{pmatrix} \right),$$ (A.7)

with

$$\mu_1 = [I_{pq}, 0] \text{ E}(\text{Vec}(V)) = ([I_p, 0] \otimes I_q) \left( Q\Upsilon^{\frac{1}{2}} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec}(J\delta).$$ (A.8)

Therefore,

$$\varphi = \text{trace} \left( \Lambda^{-1}\xi_T^\prime (\Upsilon\Upsilon^*\Upsilon) \xi \right) = V_1^\prime V_1.$$ (A.11)

Further, by using (A.5), we get

$$\text{Vec} \left( B \left( \hat{\beta}^S, \beta \right) \right) = -\text{Vec}(\delta^*) + \left( \Upsilon^{-\frac{1}{2}}Q' \otimes \Lambda^{\frac{1}{2}} \right) \text{E} \left[ \left( 1 - (pq - 2) (V_1^\prime V_1)^{-1} \right) \text{Vec}(V) \right] (A.9)$$

Also, using (A.7) and Theorem 2 in Judge and Bock (1978), we have

$$\text{E} \left[ \left( 1 - (pq - 2) (V_1^\prime V_1)^{-1} \right) \text{Vec}(V) \right] = \text{E} \left[ \left( 1 - \frac{(pq - 2)}{\chi_{pq+2}^2} \right) \left( \Upsilon^{-\frac{1}{2}}Q' \otimes \Lambda^{\frac{1}{2}} \right) \left( Q\Upsilon^{\frac{1}{2}} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec}(J\delta) \right], \quad (A.10)$$

where from (A.8),

$$\mu_1^\prime \mu_1 = \text{trace} \left( \delta^* \Upsilon \Upsilon^* \Upsilon \delta^* \Lambda^{-1} \right) = \Delta,$$ (A.11)

Therefore, by combining (A.6) (A.13), (A.10) and (A.11), we get

$$\text{Vec} \left( B \left( \hat{\beta}^S, \beta \right) \right) = -\text{Vec}(\delta^*) + \text{E} \left[ \left( 1 - \frac{(pq - 2)}{\chi_{pq+2}^2} \right) \left( \Upsilon^{-\frac{1}{2}}Q' \otimes \Lambda^{\frac{1}{2}} \right) \left( Q\Upsilon^{\frac{1}{2}} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec}(J\delta) \right],$$ (A.9)
and by elementary algebraic computations, we get
\[ \text{Vec} \left( B \left( \hat{\beta}^2, \beta \right) \right) = - \text{Vec}(\delta^*) + E \left[ 1 - \frac{(pq-2)}{\chi_{pq+2}^2(\Delta)} \right] \text{Vec}(\delta^*) \]
\[ = -(pq-2)E \left[ \frac{1}{\chi_{pq+2}^2(\Delta)} \right] \text{Vec}(\delta^*) , \]
that proves the third statement.

To prove the last statement, it suffices to replace the relation (A.10) by
\[ E \left[ \left( 1 - (pq-2) (V_1'V_1)^{-1} \right)^+ \text{Vec}(V) \right] = E \left[ \left( 1 - \frac{(pq-2)}{\chi_{pq+2}^2(\mu_1')\mu_1} \right)^+ \right] \left( Y^{-\frac{1}{2}}Q' \otimes \Lambda^\frac{1}{2} \right) \]
\[ \times \left( \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} \otimes I_q \right) \left( QY^\frac{1}{2} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec}(J\delta) . \]
The rest of the proof follows from algebraic computations. \qed

**Proposition A.1** Let \( c \) be a real number and assume that Proposition 3.1 holds. Then

(i)
\[ E \left\{ \text{trace} \left[ \left( 1 - c\psi^{-1} \right)^2 \xi'W\xi \right] \right\} = E \left[ \left( 1 - c\chi_{pq+2}^{-2}(\Delta) \right)^2 \right] \text{trace}(WY^*) \text{trace}(\Lambda) \]
\[ + E \left[ \left( 1 - c\chi_{pq+4}^{-2}(\Delta) \right)^2 \right] \text{trace}(\delta'J'WJ\delta) ; \]

(ii)
\[ E \left[ (1 - c\psi^{-1}) \eta'W\xi \right] = -E \left[ (1 - c\chi_{pq+2}^{-2}(\Delta)) \right] \delta'^*W\delta^* , \]
where \( \Delta \) is given in (A.11). \qed

**Proof**

(i) From the transformations in the proof of Theorem 3.1, we have
\[ E \left\{ \text{trace} \left[ \xi'W \left( 1 - c\psi^{-1} \right)^2 \xi \right] \right\} = E \left[ (1 - c (V_1'V_1)^{-1})^2 \text{trace} \left( QY^{-\frac{1}{2}}WY^{-\frac{1}{2}}Q'VAV' \right) \right] , \]
and then, the right side of this relation is

\[
E \left[ \left( 1 - c (V_i' V_1) \right)^2 (\text{Vec}(V))' \left( Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q' \otimes \Lambda \right) \text{Vec}(V) \right].
\]

Further, let

\[
(Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q) = G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}.
\]  

(A.12)

Also, one can verify that

\[
(\text{Vec}(V))' \left( \Lambda \otimes Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q \right) \text{Vec}(V) = V_i' G_{11} V_1,
\]

and then,

\[
E \left[ \left( 1 - c (V_i' V_1) \right)^2 (\text{Vec}(V))' \left( Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q' \otimes \Lambda \right) \text{Vec}(V) \right] = E \left[ \left( 1 - c (V_i' V_1) \right)^2 V_i' G_{11} V_1 \right],
\]

and using Theorem 2 in Judge and Bock (1978), we have

\[
E \left[ \left( 1 - c (V_i' V_1) \right)^2 V_i' G_{11} V_1 \right] = E \left[ \left( 1 - c \chi_{pq}^{-2} \left( \text{trace} \left( \mu_1' \mu_1 \right) \right) \right)^2 \text{trace} (G_{11}) \right] + E \left[ \left( 1 - c \chi_{pq+2}^{-2} \left( \text{trace} \left( \mu_1' \mu_1 \right) \right) \right)^2 \left( \mu_1' G_{11} \mu_1 \right) \right].
\]  

(A.13)

where \( \mu_1 \) is given by (A.8). Further, note that

\[
G_{11} = ([I_p, 0] \otimes I_q) G \begin{pmatrix} I_p \\ 0 \end{pmatrix} \otimes I_q = ([I_p, 0] \otimes I_q) \left( Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q' \otimes \Lambda \right) \times \begin{pmatrix} I_p \\ 0 \end{pmatrix} \otimes I_q
\]

(A.14)

and then, combining this relation with (A.6) and (A.8), we get

\[
\mu_1' G_{11} \mu_1 = \text{Vec}(J\delta)' \left( \chi_2^2 Q' \otimes \Lambda^{-\frac{1}{2}} \right) \left( Q Y^{-\frac{1}{2}} \chi_2 Y^{-\frac{1}{2}} Q' \otimes I_q \right) \left( Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q' \otimes \Lambda \right) \times \left( Q Y^{-\frac{1}{2}} \chi_2 Y^{-\frac{1}{2}} Q' \otimes I_q \right) \left( Q Y^{-\frac{1}{2}} \otimes \Lambda^{-\frac{1}{2}} \right) \text{Vec}(J\delta),
\]

21
that gives
\[ \mu_1' G_{11} \mu_1 = \text{trace} (\delta' J' W J \delta). \] (A.15)

Further, using (A.6), (A.12) and (A.14), we have
\[ \text{trace} (G_{11}) = \text{trace} \left[ \left( Q Y^{-\frac{1}{2}} W Y^{-\frac{1}{2}} Q' \otimes \Lambda \right) \left( Q Y^{-\frac{1}{2}} Y^{-\frac{1}{2}} Q' \otimes I_q \right) \right] = \text{trace} (W Y^*) \text{trace} (\Lambda). \] (A.16)

Therefore, the statement in (i) follows by combining (A.13), (A.15), and (A.16).

(ii) Since \( \eta \) and \( \xi \) are independent, we have
\[ E \left[ (1 - c \psi^{-1}) \eta' W \xi \right] = (E(\eta))' W E \left[ (1 - c \psi^{-1}) \xi \right] = - (\delta^*)' W E \left[ (1 - c \psi^{-1}) \xi \right], \]
and then, following the same steps as in proof of Theorem 3.1, we have
\[ E \left[ (1 - c \psi^{-1}) W \xi \right] = E \left[ (1 - c \mu^{-2} (\Lambda)) \right] \delta^*, \]
that completes the proof. \( \square \)

**Proof Theorem 3.2** The ADR of \( \hat{\beta} \) and \( \tilde{\beta} \) follows directly from Corollary 3.1. To establish
\[ R \left( \tilde{\beta}^S, \beta; W \right), \]
note that
\[ R \left( \tilde{\beta}^S, \beta; W \right) = E \left\{ \text{trace} \left[ \left( \eta + (1 - c \psi^{-1}) \xi \right)' W \left( \eta + (1 - c \psi^{-1}) \xi \right) \right] \right\}, \]
and then,
\[ R \left( \tilde{\beta}^S, \beta; W \right) = R \left( \tilde{\beta}, \beta; W \right) + 2E \left\{ \text{trace} \left[ \eta' W (1 - c \psi^{-1}) \xi \right] \right\} + E \left\{ \text{trace} \left[ \xi' W (1 - c \psi^{-1})^2 \xi \right] \right\}. \] (A.17)

Further, using Proposition A.1, we get
\[ R \left( \tilde{\beta}^S, \beta; W \right) = R \left( \tilde{\beta}, \beta; W \right) - 2E \left[ (1 - c \chi_{pq+4}^2 (\Delta)) \right] \text{trace} \left( \delta^* W \delta^* \right) + E \left[ (1 - c \chi_{pq+2}^2 (\Delta))^2 \right] \text{trace} (W Y^*) \text{trace} (\Lambda) + E \left[ (1 - c \chi_{pq+2}^2 (\Delta)) \right] \text{trace} \left( \delta^* W \delta^* \right). \]
Then, by some computations and using the identity

\[ E(\chi_{pq+4}^{-2}) = E(\chi_{pq+2}^{-2}) - 2E(\chi_{pq+4}^{-4}), \]  

(A.18)

we get

\[ R(\hat{\beta}^S, \beta; W) = ADR(\hat{\beta}) + \text{trace}(\delta^\prime W\delta^*) \left( (c + 2)^2 - 4 \right) E(\chi_{pq+4}^{-4}(\Delta)) \]

\[-c \text{trace}(L_1^\prime \Sigma L_2) \text{trace}(W\Psi^*) \left\{ 2E(\chi_{pq+2}^{-2}(\Delta)) - c E(\chi_{pq+2}^{-4}(\Delta)) \right\}, \]

and then, replacing \(c\) by \(pq - 2\), we get the desired result. Furthermore, by following the same steps, we have

\[ R(\hat{\beta}^{S+}, \beta; W) = R(\hat{\beta}, \beta; W) - 2E \left[ \left( 1 - c\chi_{pq+2}^{-2}(\Delta) \right) I(\chi_{pq+2}^{-2}(\Delta) > c) \right] \text{trace}(\delta^\prime W\delta^*) \]

\[ + E \left[ \left( 1 - c\chi_{pq+2}^{-2}(\Delta) \right)^2 I(\chi_{pq+2}^{-2}(\Delta) > c) \right] \text{trace}(W\Psi^*) \text{trace}(\Lambda) \]

\[ + E \left[ \left( 1 - c\chi_{pq+4}^{-2}(\Delta) \right)^2 I(\chi_{pq+4}^{-2}(\Delta) > c) \right] \text{trace}(\delta^\prime W\delta^*). \]

and rest of the proof follows from algebraic computations and using the identity (A.18).

□

B Some details about simulation study and data analysis

For the simulation results presented by Figure 1, the following are matrices used as \(L_1\) and \(\beta\):

\[
L_1 = \begin{pmatrix}
1 & -1 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 1 & -1
\end{pmatrix}, \quad \text{with } \beta = \begin{pmatrix}
0.2 & 0.1 & 3.0 & 1.0 & 0.09 \\
0.4 & 0.2 & 0.5 & 0.2 & 1.10 \\
0.1 & 0.3 & 0.1 & 0.8 & 1.30 \\
2.0 & 0.9 & 0.6 & 2.0 & 0.51 \\
0.7 & 1.0 & 1.5 & 1.2 & 0.85
\end{pmatrix}. \quad \text{(B.1)}
\]
Further, for the simulation results presented by Figure 2, the following are matrices used as $L_1$ and $\beta$;

$$L_1 = \begin{pmatrix}
1 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

and

$$\beta = \begin{pmatrix}
0.2 & 0.0 & 3.0 & 1.0 & 0.09 & 0 \\
0.4 & 0.1 & 0.5 & 0.2 & 1.10 & 0 \\
0.1 & 0.2 & 0.1 & 0.8 & 1.30 & 0 \\
2.0 & 0.3 & 0.6 & 2.0 & 0.51 & 0 \\
0.7 & 0.9 & 1.5 & 1.2 & 0.85 & 0 \\
0.0 & 1.0 & 0.3 & 0.0 & 0.50 & 1 \\
0.9 & 0.0 & 0.7 & 1.0 & 2.00 & 1
\end{pmatrix} \text{(B.2)}$$

<table>
<thead>
<tr>
<th></th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1.000 (NA)</td>
<td>0.918 (0.000)</td>
<td>0.788 (0.000)</td>
<td>0.755 (0.000)</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.918 (0.000)</td>
<td>1.000 (NA)</td>
<td>0.725 (0.000)</td>
<td>0.772 (0.000)</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>0.788 (0.000)</td>
<td>0.725 (0.000)</td>
<td>1.000 (NA)</td>
<td>0.916 (0.000)</td>
</tr>
<tr>
<td>$Y_4$</td>
<td>0.755 (0.000)</td>
<td>0.772 (0.000)</td>
<td>0.916 (0.000)</td>
<td>1.000 (NA)</td>
</tr>
</tbody>
</table>

Table 2: Correlation coefficient (p-value) matrix for root-stocks of apple trees data set

References


