OPTIMAL CORRELATION PRESERVING LINEAR PREDICTORS OF FACTOR SCORES IN FACTOR ANALYSIS

Kazumasa Mori* and Hiroshi Kurata**

This paper studies a prediction problem of factor scores with correlation-preserving linear predictors. We deal with three new risk functions that are obtained by modifying some typical risk functions in the literature, and derive optimal correlation-preserving linear predictors with respect to them. A necessary and sufficient condition for an identical equality among the predictors to hold is also derived.

Key words and phrases: Correlation-preserving predictor, factor analysis, factor scores, linear predictor, risk matrix, mean squared error matrix.

1. Introduction

In this paper, we derive some optimal linear predictors of factor scores under a factor analysis model. The model considered here is the following:

\[ x = \Lambda f + \varepsilon, \]

where \( x \) is an \( m \times 1 \) observed score vector, \( \Lambda \) is an \( m \times q \) factor loading matrix of full column rank, \( f \) is a \( q \times 1 \) vector of common factors to be predicted, and \( \varepsilon \) is an \( m \times 1 \) vector of unique factors. Assume as usual that

\[ E(f) = 0, \quad E(\varepsilon) = 0 \quad \text{and} \quad E(f'\varepsilon) = 0, \]

which implies that the variance-covariance matrix \( \Sigma \equiv V(x) \) of \( x \) is of the form

\[ \Sigma = \Lambda \Phi \Lambda' + \Psi. \]

Here, \( \Phi \) and \( \Psi \) are the variance-covariance matrices of \( f \) and \( \varepsilon \), respectively, both of which are assumed to be positive definite. (Although it is common to assume that \( \Psi \) is diagonal and \( \Phi \) is the correlation matrix of \( f \), we do not need this assumption.)

We are concerned with linear prediction of the factor score \( f \) under the assumption that the quantities \( \Lambda, \Phi \) and \( \Psi \) are known. A predictor \( \hat{f} \) of \( f \) is said to be linear if it is of the form

\[ \hat{f} = A'x \quad \text{for some} \quad A : m \times q. \]

Received October 14, 2012. Revised April 21, 2013. Accepted May 12, 2013.

*Office for Educational and Research Data Analysis, The University of Tokyo, Japan.

**Graduate School of Arts and Sciences, The University of Tokyo, Japan.
A typical example is the regression predictor defined by \( \hat{f} = \Phi A' \Sigma^{-1} x \). Although this predictor has several optimal properties and is widely employed, it is often the case that we prefer a predictor that preserves the covariance structure of the factor score, that is, a predictor \( \hat{f} \) whose variance-covariance matrix is identical with that of \( f \): \( V(\hat{f}) = V(f) (= \Phi) \). A predictor having this property is called correlation-preserving. As can be easily seen, a linear predictor \( \hat{f} = A' x \) is correlation-preserving if and only if the matrix \( A \) satisfies the equality \( A' \Sigma A = \Phi \), which is in turn equivalent to that \( A \) is of the form

\[
A = \Sigma^{-1/2} C \Phi^{1/2} \quad \text{for some} \quad C : m \times q \quad \text{such that} \quad C' C = I_q.
\]

Here, for a positive (semi)definite matrix \( B \), we denote by \( B^{1/2} \) the positive (semi)definite matrix such that its square is equal to \( B \). The matrices \( B^{-1/2} \) and \( B^{3/2} \) are the inverse (if it exists) and the third power of \( B^{1/2} \), respectively.

The class of correlation-preserving linear predictors includes \( \hat{f}_G = A'_G x \) with \( A_G = \Sigma^{-1/2} C_G \Phi^{1/2} \), where

\[
C_G = \Sigma^{-1/2} \Lambda^{3/2} (\Phi^{3/2} \Lambda'^{-1} \Phi^{3/2})^{-1/2}.
\]

This predictor is proposed by Green (1969) as a predictor minimizing the following risk function:

\[
r(\hat{f}) = E\{ (\hat{f} - f)' (\hat{f} - f) \}.
\]

The above risk function can be expressed as the trace of the usual mean squared error risk matrix:

\[
r(\hat{f}) = \text{tr}[R(\hat{f})] \quad \text{with} \quad R(\hat{f}) = E\{ (\hat{f} - f)' (\hat{f} - f)' \} : q \times q.
\]

Hence it is interesting to ask whether the predictor \( \hat{f}_G \) in (1.5) also minimizes the risk matrix \( R(\hat{f}) \) in terms of positive semidefiniteness. Here, a correlation-preserving linear predictor \( \hat{f}_* \) is said to minimize \( R(\hat{f}) \) in terms of positive semidefiniteness, if for any correlation-preserving linear predictor \( \hat{f} \), the difference \( R(\hat{f}) - R(\hat{f}_*) \) is positive semidefinite, which will be denoted in the sequel as

\[
R(\hat{f}) \geq R(\hat{f}_*).
\]

The inequality (1.7) is much stronger than the optimality with respect to \( r(\hat{f}) \) in (1.6). In fact, if \( \hat{f}_* \) minimizes the risk matrix \( R(\hat{f}) \) in terms of positive semidefiniteness, then it minimizes not only the trace of \( R(\hat{f}) \) in (1.6), but also the determinant of \( R(\hat{f}) \) and the following weighted risk functions associated with \( R(\hat{f}) \),

\[
r_* (\hat{f}) = E\{ (\hat{f} - f)' W (\hat{f} - f) \} \quad ( = \text{tr}[W R(\hat{f})])
\]

where \( W \) is an arbitrary \( q \times q \) positive semidefinite matrix. Ten Berge et al. (1999) derived the optimal correlation-preserving linear predictor that minimizes
the determinant. Interestingly, the predictor derived is different from $\hat{f}_G$ and is given by

$$C = \Sigma^{-1/2}\Lambda\Phi^{1/2}(\Phi^{1/2}\Lambda'\Sigma^{-1}\Lambda\Phi^{1/2})^{-1/2}. \quad (1.9)$$

This result suggests that there does not exist a predictor minimizing $R(\hat{f})$ in terms of positive semidefiniteness.

On the other hand, Anderson and Rubin (1956) and McDonald (1981) approached the problem through a risk function of the form

$$r_{**}(\hat{f}) = \mathbb{E}\{(x - \Lambda\hat{f})'\Psi^{-1}(x - \Lambda\hat{f})\}, \quad (1.10)$$

and showed that the predictor satisfying

$$C = \Sigma^{1/2}\Psi^{-1/2}\Phi^{1/2}(\Phi^{1/2}\Lambda'\Psi^{-1}\Sigma\Psi^{-1}\Lambda\Phi^{1/2})^{-1/2}$$

is optimal. Hence it is also interesting to find out whether it remains optimal with respect to the corresponding risk matrix

$$E\{\Psi^{-1/2}(x - \Lambda\hat{f})(x - \Lambda\hat{f})'\Psi^{-1/2}\}. \quad (1.11)$$

Minimization of the above risk matrices among linear predictors is discussed, for example, by Krijnen et al. (1996) and Neudecker (2004). However, as far as correlation-preserving linear predictors are concerned, few significant results have been derived until recently, perhaps due to the mathematical difficulty of this problem.

In this paper, we deal with three other risk matrices that are more tractable compared with the matrices in (1.6) and (1.11), and their corresponding one-dimensional risk functions. The three matrices are modifications of (1.6) and (1.11). We derive correlation-preserving linear predictors that are optimal with respect to them. The main theorems are given in Section 2. Section 3 is devoted to deriving a necessary and sufficient condition for the predictors derived in Section 2 to be identical with each other. A numerical example is given in Section 4.

2. Predictors minimizing risk matrix in terms of eigenvalues

In this section, we introduce three risk matrices and derive optimal correlation-preserving linear predictors with respect to them. One of the matrices treated here is given by

$$R_1(\hat{f}) = \mathbb{E}\{\Phi^{-1/2}(\hat{f} - \hat{f})(\hat{f} - f)'\Phi^{-1/2}\}. \quad (2.1)$$

By taking the trace of $R_1(\hat{f})$, we have the corresponding risk function

$$r_1(\hat{f}) = \mathbb{E}\{(\hat{f} - f)'\Phi^{-1}(\hat{f} - f)\}, \quad (2.2)$$

which is the one in (1.6) with $W = \Phi^{-1}$, that is, the mean squared error standardized by $\Phi = V(f)$. 
We begin with introducing the notion of “minimizing the risk matrix in terms of eigenvalues”: A correlation-preserving linear predictor \( \tilde{f}_1 \) is said to minimize the risk matrix \( R_1 \) in terms of eigenvalues, if for each correlation-preserving linear predictor \( \hat{f} \), there exists an orthogonal matrix \( U \) such that
\[
R_1(\hat{f}) \geq UR_1(\tilde{f}_1)U'
\]
holds, where throughout this paper inequalities for symmetric matrices are understood in terms of positive semidefiniteness. This optimality criterion is weaker than the optimality in terms of positive semidefiniteness introduced in (1.7). In fact, if \( \tilde{f}_1 \) is optimal in terms of positive semidefiniteness, then it minimizes the risk matrix in terms of eigenvalues. An implication of the inequality (2.3) is that, for any correlation-preserving linear predictor \( \hat{f} \), its risk matrix is bounded below by a positive semidefinite matrix whose eigenvalues are the same as those of \( R_1(\tilde{f}_1) \). In this sense, the inequality clarifies the eigenvalues of matrices that form a lower bound for \( R_1(\hat{f}) \).

The inequality (2.3) also implies that \( \tilde{f}_1 \) minimizes the risk function \( r_1 \). That is, for any correlation-preserving linear predictor \( \hat{f} \), the predictor \( \tilde{f}_1 \) satisfies
\[
r_1(\hat{f}) = E\{(\hat{f} - f)'\Phi^{-1}(\hat{f} - f)\} \geq E\{(\tilde{f}_1 - f)'\Phi^{-1}(\tilde{f}_1 - f)\} = r_1(\tilde{f}_1).
\]
This result is obtained by taking the trace of (2.3) and by noting that the orthogonal matrix \( U \) vanishes.

We proceed in line with Neudecker (2004). Recall the following well-known matrix inequality: For any \( q \times q \) matrix \( A \), there exists an orthogonal matrix \( U \) satisfying
\[
(A + A')/2 \leq U(A'A)^{1/2}U',
\]
(see, for example, Equation (8.20) of Zhang (2011)). By using the above inequality, we can see that for a given \( m \times q \) matrix \( V \) of full column rank and for any \( m \times q \) matrix \( C \) satisfying \( C'C = I_q \), there exists an orthogonal matrix \( U \) such that
\[
V'C + C'V \leq 2U(V'C'CV)^{1/2}U'.
\]
Since the matrix \( CC' \) is an orthogonal projection matrix, it holds that \( CC' \leq I_m \), from which we have \( V'C'CV \leq V'V \) and hence \( (V'C'CV)^{1/2} \leq (V'V)^{1/2} \). (See, for example, Theorem 7.9 of Zhang (2011).) Therefore
\[
V'C + C'V \leq 2U(V'V)^{1/2}U'.
\]
The equality is attained by \( C = V(V'V)^{-1/2} \) with the corresponding orthogonal matrix \( U = I_q \).

Arguing in a similar manner, we also have the following inequality: For each \( V : m \times q \) of full column rank and \( C : m \times q \) such that \( C'C = I_q \), there exists an orthogonal matrix \( U \) such that
\[
VC' + CV' \leq 2U(V'V')^{1/2}U'.
\]
The equality is attained by $C = V(V'V)^{-1/2}$ with $U = I_q$, since

$$V(V'V)^{-1/2}V' = (VV')^{1/2}.$$  

**Theorem 1.** The correlation-preserving linear predictor $\tilde{f}_1 = A'_1x$ with $A_1 = \Sigma^{-1/2}C_1\Phi^{1/2}$ and

$$C_1 = \Sigma^{-1/2}\Lambda\Phi^{1/2}((\Phi^{1/2}\Lambda^{-1}\Phi^{1/2})^{-1}$$

minimizes the risk matrix $R_1$ in (2.1) in terms of eigenvalues.

**Proof.** Let $C$ be any $m \times q$ matrix such that $C'C = I_q$ and consider the corresponding linear correlation-preserving predictor $\hat{f} = A'x$ with $A = \Sigma^{-1/2}C\Phi^{1/2}$. Then the risk matrix $R_1$ of $\hat{f}$ is evaluated as

$$R_1(\hat{f}) = 2I_q - (V'C + C'V) \quad \text{with} \quad V = \Sigma^{-1/2}\Lambda\Phi^{1/2}.$$  

Applying (2.6), we see that there exists an orthogonal matrix $U$ such that

$$R_1(\tilde{f}) \geq 2I_q - 2U(V'V)^{1/2}U' = U\{2I_q - 2(V'V)^{1/2}\}U'.$$

On the other hand, since the matrix $C_1$ defined in (2.8) is expressed as $C_1 = \Sigma^{-1/2}\Lambda\Phi^{1/2}$, it holds that

$$R_1(\tilde{f}_1) = 2I_q - 2(V'V)^{1/2}.$$  

Thus we have

$$R_1(\tilde{f}) \geq UR_1(\tilde{f}_1)U'.$$

This completes the proof. □

Note that the predictor $\tilde{f}_1$ is the same as the one in (1.9) that minimizes the determinant of the risk matrix in (1.6).

**Corollary 1.** The predictor $\tilde{f}_1$ in (2.8) minimizes the risk function $r_1$ in (2.2) among the class of correlation-preserving linear predictors.

Similar results can also be obtained for two other risk functions. The risk functions considered here are of the form

$$r(\hat{f}) = E\{(x - \Lambda\hat{f})'W(x - \Lambda\hat{f})\} \quad \text{with} \quad W : m \times m \text{ positive semidefinite},$$

which is the trace of the following risk matrix

$$R(\hat{f}) = E\{W^{1/2}(x - \Lambda\hat{f})(x - \Lambda\hat{f})'W^{1/2}\}. $$
The first one is the risk matrix with \( W^{-1} = \Sigma(= V(x)) \):

\[
R_2(\hat{f}) = E\{\Sigma^{-1/2}(x - \Lambda\hat{f})(x - \Lambda\hat{f})'\Sigma^{-1/2} \}: m \times m.
\]

(2.11)

Taking the trace of \( R_2 \), we obtain the risk function

\[
r_2(\hat{f}) = E\{(x - \Lambda\hat{f})'(\Sigma^{-1}(x - \Lambda\hat{f})\}\} \quad (= \text{tr}[R_2(\hat{f})]),
\]

(2.12)

which is McDonald’s (1981) risk function (1.10) with \( \Psi = V(\varepsilon) \) replaced by \( \Sigma = V(x) \). The risks \( R_2 \) and \( r_2 \) are obtained by standardizing the mean squared error matrix with \( \Sigma \). Since all the predictors considered here are functions of \( x \), it may be reasonable to use \( \Sigma = V(x) \) as a standardizing matrix.

Also interesting is the risk function obtained by standardizing the mean squared error matrix with \( \Lambda\Phi\Lambda' = V(\Lambda f) \), the variance-covariance matrix of \( \Lambda f \). Since the predictors \( \hat{f} \) treated here are correlation-preserving, we have

\[
V(\Lambda\hat{f}) = \Lambda\Phi\Lambda' = V(\Lambda f)
\]

for all \( \hat{f} \), and hence the risk function below can be viewed as a counterpart of \( r_2 \) above. Since the standardizing matrix is singular, we need to use the Moore-Penrose inverse to define

\[
r_3(\hat{f}) = E\{(x - \Lambda\hat{f})'(\Lambda\Phi\Lambda')^{+}(x - \Lambda\hat{f})\},
\]

(2.13)

where \( W = (\Lambda\Phi\Lambda')^{+} \) and \( B^{+} \) denotes the Moore-Penrose inverse of \( B \). Let

\[
L = \Lambda\Phi^{1/2}.
\]

Then it is easy to observe that \( L(L'L)^{-2}L' = (\Lambda\Phi\Lambda')^{+} \) and \( (L'L)^{-1}L' = (\Lambda\Phi^{1/2})^{+} \). Hence the matrix-version of \( r_3 \) is obtained as

\[
R_3(\hat{f}) = E\{(\Lambda\Phi^{1/2})^{+}(x - \Lambda\hat{f})(x - \Lambda\hat{f})'(\Lambda\Phi^{1/2})^{+}\}'
\]

\[
= E\{(L'L)^{-1}L'(x - \Lambda\hat{f})(x - \Lambda\hat{f})'L(L'L)^{-1}\} : q \times q,
\]

(2.14)

which clearly satisfies

\[
r_3(\hat{f}) = \text{tr}[R_3(\hat{f})].
\]

Let \( \hat{f} \) be a correlation-preserving linear predictor \( \hat{f} = A'x \) with \( A = \Sigma^{-1/2}C\Phi^{1/2} \). Then the two matrices reduce to

\[
R_2(\hat{f}) = I_m - \Sigma^{-1/2}\Lambda A'\Sigma^{-1/2} - \Sigma^{1/2}A\Lambda'\Sigma^{-1/2} + \Sigma^{-1/2}\Lambda\Phi A'\Sigma^{-1/2}
\]

\[
= I_m - \Sigma^{-1/2}\Lambda\Phi^{1/2}C - C\Phi^{1/2}\Lambda'\Sigma^{-1/2} + \Sigma^{-1/2}\Lambda\Phi A'\Sigma^{-1/2}
\]

\[
= \overline{R}_2(\hat{f}) + \Sigma^{-1/2}\Lambda\Phi A'\Sigma^{-1/2}
\]

(2.15)

with

\[
\overline{R}_2(\hat{f}) = I_m - \Sigma^{-1/2}\Lambda\Phi^{1/2}C - C\Phi^{1/2}\Lambda'\Sigma^{-1/2}
\]

(2.16)
and

\[ R_3(\hat{f}) = I_q - \Phi^{-1/2} A' \Sigma L(L'L)^{-1} - (L'L)^{-1} L' \Sigma A \Phi^{-1/2} \]
\[ + (L'L)^{-1} L' \Sigma L(L'L)^{-1} \]
\[ = I_q - C' \Sigma^{1/2} L(L'L)^{-1} - (L'L)^{-1} L' \Sigma^{1/2} C \]
\[ + (L'L)^{-1} L' \Sigma L(L'L)^{-1} \]
\[ = \overline{R}_3(\hat{f}) + (L'L)^{-1} L' \Sigma L(L'L)^{-1} \]

with

\[ \overline{R}_3(\hat{f}) = I_q - C' \Sigma^{1/2} L(L'L)^{-1} - (L'L)^{-1} L' \Sigma^{1/2} C, \]

respectively.

Since the matrices \( \Sigma^{-1/2} \Lambda \Phi \Lambda' \Sigma^{-1/2} \) and \( (L'L)^{-1} L' \Sigma L(L'L)^{-1} \) are positive semidefinite and do not depend on \( C \), inequalities between two correlation-preserving linear predictors with respect to \( R_2 \) and \( R_3 \) are equivalent to those with respect to \( \overline{R}_2 \) and \( \overline{R}_3 \), respectively. By viewing \( \overline{R}_2 \) and \( \overline{R}_3 \) as

\[ \overline{R}_2(\hat{f}) = I_m - (V_2 C' + CV_2') \quad \text{with} \quad V_2 = \Sigma^{-1/2} \Lambda \Phi^{1/2} \]

and

\[ \overline{R}_3(\hat{f}) = I_q - (V_3' C + C' V_3) \quad \text{with} \quad V_3 = \Sigma^{1/2} L(L'L)^{-1} \]

and by applying (2.6) and (2.7), respectively, we obtain the following results:

The correlation-preserving linear predictors \( \hat{f}_i = A_i' x \) with \( A_i = \Sigma^{-1/2} C_i \Phi^{1/2} \) \((i = 2, 3)\) with

\[ C_2 = V_2 (V_2' V_2)^{-1/2} = \Sigma^{-1/2} \Lambda \Phi^{1/2} (\Phi^{1/2} \Lambda' \Sigma^{-1} \Lambda \Phi^{1/2})^{-1/2} \]

and

\[ C_3 = V_3 (V_3' V_3)^{-1/2} \]
\[ = \Sigma^{1/2} \Lambda (\Lambda' \Lambda)^{-1} \Phi^{-1/2} (\Phi^{-1/2} (\Lambda' \Lambda)^{-1} \Lambda' \Sigma \Lambda (\Lambda' \Lambda)^{-1} \Phi^{-1/2})^{-1/2} \]

minimize \( \overline{R}_i \) \((i = 2, 3)\) in terms of eigenvalues, respectively. That is, for each correlation-preserving linear predictor \( \hat{f} \), there exists an orthogonal matrix \( U_i \) \((i = 2, 3)\) such that

\[ \overline{R}_i(\hat{f}) \geq U_i \overline{R}_i(\widetilde{f}_i) U_i' \quad (i = 2, 3). \]

Here we should note that

\[ \widetilde{f}_1 = \widetilde{f}_2. \]

By taking the trace of (2.23), we obtain the following optimality result:
Theorem 2. The predictors $\tilde{f}_2$ in (2.21) and $\tilde{f}_3$ in (2.22) minimize the risk functions $r_2$ in (2.12) and $r_3$ in (2.13) among the class of correlation-preserving linear predictors, respectively.

Proof. We prove the result for only $\tilde{f}_2$, since the proof for $\tilde{f}_3$ is essentially the same. By taking the trace of (2.23), we see that, for each correlation-preserving linear predictor $\hat{f}$,

$$\text{tr}[R_2(\hat{f})] \geq \text{tr}[R_2(\tilde{f}_2)],$$

which is equivalent to

$$r_2(\hat{f}) = \text{tr}[R_2(\hat{f})] = \text{tr}[R_2(\tilde{f}_2)] + \text{tr}[(\Sigma^{-1/2}\Lambda\Lambda'\Sigma^{-1/2})] \geq \text{tr}[R_2(\tilde{f}_2)] + \text{tr}[(\Sigma^{-1/2}\Lambda\Lambda'\Sigma^{-1/2})] = \text{tr}[R_2(\tilde{f}_2)] = r_2(\tilde{f}_2).$$

Hence the proposition is proved for the predictor $\tilde{f}_2$. This completes the proof. □

3. Identical equality between the two predictors

In this section, we derive a condition for the two predictors ($\tilde{f}_1 = \tilde{f}_2$ and $\tilde{f}_3$) obtained in the previous section to be identical with each other.

Theorem 3. The two correlation-preserving linear predictors $\tilde{f}_1$ and $\tilde{f}_3$ coincide with each other if and only if the following equality holds:

$$\left(\Lambda'\Psi^{-1}\Lambda\right)^{-1} = \left(\Lambda'\Lambda\right)^{-1}\Lambda'\Psi\Lambda\left(\Lambda'\Lambda\right)^{-1}.$$  

(3.1)

Before proving the theorem, it would be helpful to introduce the following matrix result, which is well-known in the context of linear model estimation.

Lemma 1. Let $X$ be an $m \times q$ matrix of full column rank and $\Delta$ be an arbitrary $m \times m$ positive definite matrix. Then the following three statements are equivalent:

(i) $(X'\Delta^{-1}X)^{-1} = (X'X)^{-1}X'\Delta X(X'X)^{-1},$
(ii) $(X'\Delta^{-1}X)^{-1}X'\Delta^{-1} = (X'X)^{-1}X',$
(iii) $X'\Delta Z = 0$ holds for an $m \times (m-q)$ matrix $Z$ of full rank that satisfies $X'Z = 0.$

To see the essence of the above lemma, let us consider briefly a general linear model

$$y = X\beta + \varepsilon \quad \text{with} \quad E(\varepsilon) = 0 \quad \text{and} \quad V(\varepsilon) = \Delta,$$

with $y : m \times 1$, $X : m \times q$ of full rank, $\beta : q \times 1$ and $\Delta$: positive definite. The above lemma states that the condition (iii) is necessary and sufficient for the best linear unbiased estimator $(X'\Delta^{-1}X)^{-1}X'\Delta^{-1}y$ of $\beta$ to be identical with the ordinary least squares estimator $(X'X)^{-1}X'y$. Applying this observation
to the model (1.1) would facilitate the understanding of Theorem 3. For further
detail, see, for example, Puntanen and Styan’s (1989) survey paper based on
approximately 100 references.

Proof of Theorem 3. We first prove that the identical equality \( \tilde{f}_1 = \tilde{f}_3 \)
is equivalent to the equality

\[
(3.2) \quad (\Lambda'\Sigma^{-1}\Lambda)^{-1} = (\Lambda'\Lambda)^{-1}\Lambda'\Sigma\Lambda(\Lambda'\Lambda)^{-1}.
\]

To do so, assume that \( C_3 = C_1 \):

\[
(3.3) \quad \Sigma^{1/2}\Lambda(\Lambda'\Lambda)^{-1}\Phi^{-1/2}[\Phi^{-1/2}(\Lambda'\Lambda)^{-1}\Lambda'\Sigma\Lambda(\Lambda'\Lambda)^{-1}\Phi^{-1/2}]^{-1/2} = \Sigma^{-1/2}\Lambda\Phi^{1/2}(\Phi^{1/2}\Lambda'\Sigma^{-1}\Lambda\Phi^{1/2})^{-1/2}.
\]

Premultiplying both sides of (3.3) by \( \Phi^{1/2}\Lambda'\Sigma^{-1/2} \), we have

\[
[\Phi^{-1/2}(\Lambda'\Lambda)^{-1}\Lambda'\Sigma\Lambda(\Lambda'\Lambda)^{-1}\Phi^{-1/2}]^{-1} = (\Phi^{1/2}\Lambda'\Sigma^{-1}\Lambda\Phi^{1/2})^{1/2},
\]
or equivalently,

\[
[\Phi^{-1/2}(\Lambda'\Lambda)^{-1}\Lambda'\Sigma\Lambda(\Lambda'\Lambda)^{-1}\Phi^{-1/2}]^{-1} = \Phi^{1/2}\Lambda'\Sigma^{-1}\Lambda\Phi^{1/2},
\]

which implies (3.2).

Conversely, suppose that (3.2) holds. This implies

\[
(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Lambda'\Sigma^{-1} = (\Lambda'\Lambda)^{-1}\Lambda'.
\]

Therefore, we can rewrite the matrix \( C_3 \) as

\[
C_3 = \Sigma^{1/2}[\Sigma^{-1}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}]\Phi^{-1/2}[\Phi^{-1/2}(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2}]^{-1/2} = \Sigma^{-1/2}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2}[\Phi^{-1/2}(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2}]^{-1/2}.
\]

Let \( X = \Phi^{-1/2}(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2} \), then \( C_1 = \Sigma^{-1/2}\Lambda\Phi^{1/2}X^{1/2} \) and \( C_3 = \Sigma^{-1/2}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2}X^{-1/2} \), from which it follows that

\[
C_1 - C_3 = \Sigma^{-1/2}\Lambda\Phi^{1/2}X^{1/2} - \Sigma^{-1/2}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Phi^{-1/2}X^{-1/2} = \Sigma^{-1/2}\Lambda\Phi^{1/2}(X^{1/2} - XX^{-1/2}) = 0.
\]

Next we show that (3.2) is equivalent to (3.1). Let \( Z \) be an \( m \times (m - q) \)
matrix of full rank such that \( \Lambda'Z = 0 \). Then we have

\[
(3.4) \quad \Lambda'\Sigma Z = \Lambda'\Phi\Lambda' + \Psi Z = \Lambda'\Psi Z.
\]

Hence clearly \( \Lambda'\Sigma Z = 0 \) is equivalent to \( \Lambda'\Psi Z = 0 \). Thus, from (iii) of Lemma 1
(with \( X = \Lambda \)), the equivalence between (3.2) and (3.1) follows. This completes
the proof. \( \square \)
4. A numerical example

Theorem 1 states that for each correlation-preserving linear predictor \( \hat{f} \), there exists an orthogonal matrix \( U \) such that

\[
R_1(\hat{f}) \geq U R_1(\tilde{f}_1) U'.
\]

(4.1)

Hence any predictor whose corresponding orthogonal matrix is the identity matrix (i.e., \( U = I_q \)) is dominated by \( \tilde{f}_1 \) with respect to \( R_1 \) in terms of positive semidefiniteness. Such predictors include \( \tilde{f}_3 \) discussed in Section 2. In fact, in our numerical study, the risk matrix of \( \tilde{f}_3 \) is greater than that of \( \tilde{f}_1 \) for every case we have examined.

On the other hand, some predictors have their corresponding non-identical orthogonal matrices (\( U \neq I_q \)), and hence are not dominated by \( \tilde{f}_1 \) in the usual sense. To see this more clearly, let

\[
\Lambda = \begin{pmatrix}
-3 & -2 & 0 & -3 \\
2 & 1 & 1 & 1 \\
1 & -1 & -2 & 1 \\
1 & -2 & -2 & 1 \\
2 & 3 & 2 & 2 \\
0 & 0 & 0 & 3 \\
0 & 0 & 1 & 2 \\
2 & 0 & -2 & -1
\end{pmatrix} : 8 \times 4, \quad \Phi = \begin{pmatrix}
1 & -0.2 & -0.6 & 0.6 \\
-0.2 & 1 & 0.5 & -0.1 \\
-0.6 & 0.5 & 1 & -0.6 \\
0.6 & -0.1 & -0.6 & 1
\end{pmatrix} : 4 \times 4
\]

and \( \Psi = \text{diag}(4, 2, 1, 2, 3, 3, 2, 3) : 8 \times 8 \), where \( \text{diag}(\cdot) \) denotes the diagonal matrix with the diagonal elements being the numbers indicated. Let \( \tilde{f}_1 \) be the predictor in Theorem 1 and let \( \hat{f}_s = A'x \) with \( A = \Sigma^{-1/2} C \Phi^{1/2} \) be a correlation-preserving linear predictor such that the column vectors of \( C : 8 \times 4 \) coincide with the normalized eigenvectors corresponding to the first through fourth largest eigenvalues of \( \Sigma \). Then we can see that the eigenvalues of \( R_1(\hat{f}_s) - R_1(\tilde{f}_1) \) are given by \((3.69, 2.18, 0.40, -0.03)\), which contain a negative number. This clearly means that \( R_1(\hat{f}_s) - R_1(\tilde{f}_1) \) is not positive semidefinite.

The eigenvalues of \( R_1(\hat{f}_s) \) and \( R_1(\tilde{f}_1) \) are given by \((3.82, 2.79, 1.20, 0.14)\) and \((1.22, 0.41, 0.07, 0.03)\), respectively. Since, for each \( i \) \((i = 1, \ldots, 4)\), the \( i \)-th largest eigenvalue of \( R_1(\hat{f}_s) \) is greater than that of \( R_1(\tilde{f}_1) \), we can easily see that \( U = QP' \) makes (4.1) hold, where \( P \) and \( Q \) are the orthogonal matrices such that \( P' R_1(\tilde{f}_1) P \) and \( Q' R_1(\hat{f}_s) Q \) are diagonal matrices whose diagonal elements are the descendingly ordered eigenvalues of \( R_1(\tilde{f}_1) \) and \( R_1(\hat{f}_s) \), respectively. In fact, \( R_1(\hat{f}_s) - U R_1(\tilde{f}_1) U' \) has \((3.70, 2.07, 0.42, 0.04)\) as its eigenvalues, and is positive semidefinite.

Acknowledgements

The authors express their sincere gratitude to two anonymous reviewers whose constructive comments greatly improved the quality of this manuscript.
References


