ELSEVIER
Linear Algebra and its Applications 354 (2002) 3–20
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Linear regression analysis using the relative squared error

Bernhard F. Arnold, Peter Stahlecker*

Institut für Statistik und Ökonometrie, Universität Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany

Received 26 January 2000; accepted 4 November 2001

Submitted by H.J. Werner

Abstract

In order to determine estimators and predictors in a generalized linear regression model we apply a suitably defined relative squared error instead of the most frequently used absolute squared error. The general solution of a matrix problem is derived leading to minimax estimators and predictors. Furthermore, we consider an important special case, where an analogon to a well-known relation between estimators and predictors holds and where generalized least squares estimators as well as Kuks–Olman and ridge estimators play a prominent role.

Keywords: Linear affine estimator; Linear affine predictor; Linear regression; Löwner ordering; Minimax principle; Ridge regression

1. Introduction

In this paper we consider the linear regression model

$$y = X\beta + u \tag{1}$$

with

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{2}$$

where $y_1 \in \mathbb{R}^{n_1}$ is the column vector of known observations of the dependent variable, $y_2 \in \mathbb{R}^{n_2}$ is the column vector of the unknown values of the dependent variable,

* Corresponding author.
E-mail address: fw9a009@rrz.uni-hamburg.de (P. Stahlecker).

PII: S0024-3795(01)00572-9

$X_1 \in \mathbb{R}^{n_1 \times k}$ and $X_2 \in \mathbb{R}^{n_2 \times k}$ are the deterministic model matrices of the known values of the k explanatory variables, $\beta \in \mathbb{R}^k$ is the column vector of the unknown regression coefficients, and $u_1 \in \mathbb{R}^{n_1}$ and $u_2 \in \mathbb{R}^{n_2}$ are the column vectors of the unobservable disturbances. Setting $n = n_1 + n_2$ we obtain $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times k}$, and $u \in \mathbb{R}^n$.

Let us first focus on the problem of estimating β by means of a linear function $b = C_e y_1$ with $C_e \in \mathbb{R}^{k \times n_1}$. When $rk(X_1) = k$ the most prominent estimator of β is the ordinary least squares estimator given by $C_e = (X_1' X_1)^{-1} X_1'$. If $rk(X_1) < k$ or if the matrix $X_1' X_1$ is ill-conditioned, i.e., in case of exact or near multicollinearity, a ridge estimator proposed by Hoerl and Kennard [6] may be used. This estimator is defined by

$$C_e = \left(r I_k + X_1' X_1\right)^{-1} X_1', \tag{3}$$

where $I_k \in \mathbb{R}^{k \times k}$ is the identity matrix and r is a suitably selected positive real number.

Kuks and Olman [8,9] suggested to use the minimax principle when there is some prior information available about β represented by a compact nonempty set $\mathfrak{B} \subseteq \mathbb{R}^k$. For this purpose they assume that the expectation $E(u_1)$ of the random disturbances u_1 is zero and that the covariance matrix $V_{11} \in \mathbb{R}^{n_1 \times n_1}$ of u_1 is known and positive definite (p.d.):

$$E(u_1) = 0 \quad \text{and} \quad E(u_1 u_1') = V_{11}. \tag{4}$$

Note that throughout this paper any nonnegative definite (n.n.d.) or p.d. matrix is assumed to be symmetric. Kuks and Olman applied the minimax principle to the weighted scalar mean squared error $E((C_e y_1 - \beta)' B_e (C_e y_1 - \beta))$, where $B_e \in \mathbb{R}^{k \times k}$ is the given n.n.d. matrix of weights. According to this approach a linear estimator $b^* = C_e^* y_1$ of β is called optimal if the inequality

$$\max_{\beta \in \mathfrak{B}} E((C_e^* y_1 - \beta)' B_e (C_e^* y_1 - \beta)) \leqslant \max_{\beta \in \mathfrak{B}} E((C_e y_1 - \beta)' B_e (C_e y_1 - \beta))$$

holds for all $C_e \in \mathbb{R}^{k \times n_1}$. In general, this optimization problem cannot be solved explicitly. For further discussions see, e.g., [4,7,10,12,13,16,18,19]. In the special case of $rk(B_e) = 1$ and of an ellipsoidal information set $\mathfrak{B} = \{\beta \in \mathbb{R}^k | \beta' S \beta \leqslant 1\}$, where $S \in \mathbb{R}^{k \times k}$ is a given p.d. matrix, Kuks and Olman already derived an optimal linear estimator $b^* = C_e^* y_1$ with

$$C_e^* = (S + X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1}, \tag{5}$$

not depending on B_e. Here, no rank condition is imposed on X_1. Following (5) we subsequently call an estimator $b = C_e y_1$ of β a Kuks–Olman estimator if

$$C_e = (W_1 + X_1' W_2 X_1)^{-1} X_1' W_2 \tag{6}$$

with given p.d. matrices $W_1 \in \mathbb{R}^{k \times k}$ and $W_2 \in \mathbb{R}^{n_1 \times n_1}$. Kuks–Olman estimators can also be viewed as general ridge estimators discussed, e.g., in [11,15]. It is noteworthy that Kuks–Olman estimators also appear within the framework of a Bayesian approach, where the knowledge about β is represented by a probability distribution (see, e.g., [15] or [14, p. 270]).

We now consider the problem of predicting the unknown y_2 by a suitable linear function $p = C_p y_1$ with $C_p \in \mathbb{R}^{n_2 \times n_1}$. First of all, any linear estimator $b = C_e y_1$ of β ($C_e \in \mathbb{R}^{k \times n_1}$) can be used to obtain a linear predictor for y_2 defined by

$$C_p = X_2 C_e. \tag{7}$$

For instance, in case of $rk(X_1) = k$, the generalized least squares estimator given by

$$C_e = (X_1' V_{11}^{-1} X_1)^{-1} X_1' V_{11}^{-1} \tag{8}$$

may be inserted into (7), where V_{11} is the p.d. covariance matrix of u_1 (see (4)). Of course, such a linear predictor does not make use of the covariance structure between y_1 and y_2. Goldberger [5] was the first who exploited this relationship. His approach is also presented by Rao and Toutenburg [16] in a more general framework; they assume that the disturbance term u is a random variable with expectation zero and with an n.n.d. covariance matrix $V \in \mathbb{R}^{n \times n}$:

$$E(u) = 0 \quad \text{and} \quad E(uu') = V. \tag{9}$$

Setting

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

and supposing that the p.d. matrix $V_{11} \in \mathbb{R}^{n_1 \times n_1}$ and the matrix $V_{12} = V_{21}' \in \mathbb{R}^{n_1 \times n_2}$ are known, they show that

$$C_p^* = X_2 C_e + V_{21} V_{11}^{-1} (I_{n_1} - X_1 C_e) \tag{10}$$

minimizes the weighted scalar mean squared error under the condition of unbiasedness, where $I_{n_1} \in \mathbb{R}^{n_1 \times n_1}$ is the identity matrix and C_e is defined by (8):

$$E((C_p^* y_1 - y_2)' B_p (C_p^* y_1 - y_2)) \leqslant E((C_p y_1 - y_2)' B_p (C_p y_1 - y_2))$$

for all $C_p \in \mathbb{R}^{n_2 \times n_1}$ with $C_p X_1 = X_2$ and for any arbitrarily selected n.n.d. matrix $B_p \in \mathbb{R}^{n_2 \times n_2}$ of weights. Note that the condition $C_p X_1 = X_2$ is equivalent to $E(C_p y_1 - y_2) = 0$ for all $\beta \in \mathbb{R}^k$, i.e., the condition $C_p X_1 = X_2$ is equivalent to the unbiasedness of the linear predictor $p = C_p y_1$.

In case of multicollinearity, e.g., a ridge estimator or a Kuks–Olman estimator $b = C_e y_1$ given by (3) or (6), respectively, may be inserted into the right-hand side of the equation

$$C_p = X_2 C_e + V_{21} V_{11}^{-1} (I_{n_1} - X_1 C_e), \tag{11}$$

(see (10)) in order to obtain a linear predictor $p = C_p y_1$ of y_2. In [2] the minimax principle is directly applied to the weighted scalar mean squared error of a linear predictor. In that paper the maximum is taken over a fuzzy set representing the prior information. This approach, of course, also contains the case of a classical (crisp) information set. It is shown that, in an important special case, an optimal linear predictor is obtained by inserting the matrix C_e of a specific Kuks–Olman estimator $b = C_e y_1$ of β into the right-hand side of equation (11).

The following sections deal with a direct minimax approach to the problems of estimating and predicting in linear regression analysis. This approach is based on a concept of relative rather than absolute squared error and requires no prior information. The general solution of a matrix problem is derived that leads to optimal estimators and predictors. Furthermore, we consider an important special case, where generalized least squares estimators, Kuks–Olman estimators, and the Eq. (11) play a prominent role.

2. Estimating the regression coefficients

Let b be an estimator of the regression coefficient β in the linear model (1), (2), and let $(b-\beta)'A_e(b-\beta)$ be the weighted squared error of b, where $A_e \in \mathbb{R}^{k\times k}$ is a given n.n.d. matrix of weights. Whereas the traditional analysis of biased estimators starts from this loss function and uses the expected squared error of b as a risk function, we are interested in a measure of the relative squared error allowing for a worst case analysis. To focus on the maximum relative squared error might be appropriate for those empirical studies, where replicated experiments are very expensive or even impossible.

Using a simple example, we first explain the idea and specify how it is related to the problem of (near) multicollinearity. Assume $A_e = I_k$, $rk(X_1) = k$ and consider the ordinary least squares estimator of β, i.e., $b = (X_1'X_1)^{-1}X_1'y_1$. The squared error of b is given by

$$(b-\beta)'(b-\beta) = u_1'X_1\left(X_1'X_1\right)^{-2}X_1'u_1.$$

Therefore, the estimation error of b solely depends on the unobservable (and unavoidable) disturbance vector u_1. However, for a worst case analysis, one should not use the squared error loss, since it is unbounded. Rather, the relative estimation error, defined by

$$\frac{(b-\beta)'(b-\beta)}{u_1'u_1} = \frac{u_1'X_1(X_1'X_1)^{-2}X_1'u_1}{u_1'u_1}$$

for any $u_1 \in \mathbb{R}^{n_1}$, $u_1 \neq 0$, makes sense. Then we get

$$\begin{aligned}\max_{u_1\neq 0}\frac{(b-\beta)'(b-\beta)}{u_1'u_1} &= \lambda_{\max}(X_1(X_1'X_1)^{-2}X_1')\\ &= \lambda_{\max}((X_1'X_1)^{-1})\\ &= \frac{1}{\lambda_{\min}(X_1'X_1)},\end{aligned}$$

where $\lambda_{\max}(\cdot),\ \lambda_{\min}(\cdot)$ denote the maximum and minimum eigenvalue of the matrix in the argument, respectively. This formula shows how near multicollinearity affects the least squares estimator, yielding a small value of $\lambda_{\min}(X_1'X_1)$ and therefore a large value of the maximum relative squared error. Note that the ratio

$$\frac{\max\limits_{u_1\neq 0}((b-\beta)'(b-\beta))/(u_1'u_1)}{\min\limits_{u_1\neq 0}((b-\beta)'(b-\beta))/(u_1'u_1)}=\frac{\lambda_{\max}(X_1'X_1)}{\lambda_{\min}(X_1'X_1)}$$

yields the square of the so-called condition number of $X_1'X_1$. Let us now assume $rk(X_1)<k$. In this case we may consider any least squares estimator

$$b=X_1^+y_1+(I_k-X_1^+X_1)h,$$

where X_1^+ denotes the Moore–Penrose inverse of X_1, and $h\in\mathbb{R}^k$ is an arbitrarily chosen vector of constants. We obtain

$$(b-\beta)'(b-\beta)=\left[(X_1^+X_1-I_k)(\beta-h)+X_1^+u_1\right]'\times\left[(X_1^+X_1-I_k)(\beta-h)+X_1^+u_1\right].$$

Because we now have $X_1^+X_1\neq I_k$, the squared estimation error of b does not only depend on the 'noise' u_1 but also on the 'signal' β, more precisely, on $\beta-h$. Again, a worst case analysis should start from a relative rather than an absolute squared error, where a suitably weighted squared signal should enter the denominator.

Generalizing the concept of a relative error we want to allow for an increasing squared error $(b-\beta)'A_e(b-\beta)$ whenever the disturbance vector u_1 is increasing with respect to some suitably chosen pseudo-norm; furthermore, we tolerate a greater value of the weighted squared error of b in case of a larger pseudo-norm of the parameter vector β. Combining both vectors β and u_1 to one column vector $\gamma=\binom{\beta}{u_1}\in\mathbb{R}^{k+n_1}$ we consider the ratio

$$\frac{(b-\beta)'A_e(b-\beta)}{\gamma'T_e\gamma},\tag{12}$$

where $T_e\in\mathbb{R}^{(k+n_1)\times(k+n_1)}$ is a given n.n.d. matrix, $T_e\neq 0$, and where we assume $T_e\gamma\neq 0$, i.e., γ is not an element of the null-space N_e of T_e.

Obviously, (12) meets both of the requirements stated above. Moreover, it may be more appropriate not to focus on the values of β and u_1 themselves, i.e., on the deviations of β and u_1 from the corresponding null vectors, but on the deviations of β and u_1 from given parameters $\beta_0\in\mathbb{R}^k$ and $u_{10}\in\mathbb{R}^{n_1}$, respectively. Here, β_0 might be the result of theoretical or empirical considerations; note that, in the introductory example from above, β_0 may correspond to the vector h. Furthermore, u_{10} might be some presumed specification error which, for instance, may occur when a multiplicative model is transformed into the linear regression model (1), (2). Setting $\gamma_0 = \binom{\beta_0}{u_{10}} \in \mathbb{R}^{k+n_1}$ we obtain

$$\frac{(b-\beta)' A_e (b-\beta)}{(\gamma-\gamma_0)' T_e (\gamma-\gamma_0)} \tag{13}$$

as an analogon to expression (12). When there are no preferences with respect to β_0 and u_{10} these parameters should be set equal to zero, and (12) will be relevant.

In this paper we are going to apply the minimax principle to quantity (13); here, we consider linear affine estimators and we do not make use of any information about γ. Inserting (1) and (2) into $b = C_e y_1 + c_e$ ($C_e \in \mathbb{R}^{k\times n_1}$, $c_e \in \mathbb{R}^k$) and setting $D_e = (C_e X_1 - I_k, C_e) \in \mathbb{R}^{k\times(k+n_1)}$, we get

$$\frac{(D_e\gamma + c_e)' A_e (D_e\gamma + c_e)}{(\gamma-\gamma_0)' T_e (\gamma-\gamma_0)} \tag{14}$$

being equivalent to (13). This leads to the following definition of an optimal linear affine estimator.

Definition 1. Let $A_e \in \mathbb{R}^{k\times k}$ and $T_e \in \mathbb{R}^{(k+n_1)\times(k+n_1)}$, $T_e \neq 0$, be given n.n.d. matrices, and let $\gamma_0 \in \mathbb{R}^{k+n_1}$ be a given vector. Then a linear affine estimator $b^* = C_e^* y_1 + c_e^*$ ($C_e^* \in \mathbb{R}^{k\times n_1}$, $c_e^* \in \mathbb{R}^k$) for β in model (1) and (2) is optimal if

$$\text{(i)} \quad \sup_{\substack{\gamma\in\mathbb{R}^{k+n_1} \\ \gamma-\gamma_0\notin N_e}} \frac{(D_e^*\gamma + c_e^*)' A_e (D_e^*\gamma + c_e^*)}{(\gamma-\gamma_0)' T_e (\gamma-\gamma_0)} < \infty$$

and if the inequality

$$\text{(ii)} \quad \sup_{\substack{\gamma\in\mathbb{R}^{k+n_1} \\ \gamma-\gamma_0\notin N_e}} \frac{(D_e^*\gamma + c_e^*)' A_e (D_e^*\gamma + c_e^*)}{(\gamma-\gamma_0)' T_e (\gamma-\gamma_0)} \leqslant \sup_{\substack{\gamma\in\mathbb{R}^{k+n_1} \\ \gamma-\gamma_0\notin N_e}} \frac{(D_e\gamma + c_e)' A_e (D_e\gamma + c_e)}{(\gamma-\gamma_0)' T_e (\gamma-\gamma_0)}$$

holds for all $C_e \in \mathbb{R}^{k\times n_1}$ and all $c_e \in \mathbb{R}^k$. Here, we have set $D_e^* = (C_e^* X_1 - I_k, C_e^*)$ and $D_e = (C_e X_1 - I_k, C_e)$.

Note that there are situations, where the supremum of expression (14) is infinite for all linear affine estimators, and thus, it is sensible to impose condition (i) on an optimal estimator. To give an example, let $A_e = I_k$ and $T_e = \begin{pmatrix} 0 & 0 \\ 0 & W \end{pmatrix}$ with $W \in \mathbb{R}^{n_1\times n_1}$, p.d. Assuming $rk(X_1) < k$ which implies $C_e X_1 - I_k \neq 0$ we see that the supremum of (14) is infinite for all $C_e \in \mathbb{R}^{k\times n_1}$.

The problem of determining an optimal linear affine estimator can be reduced to find its linear part. To see this we look at the relation

$$
\begin{aligned}
&\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \frac{(D_e\gamma+c_e)'A_e(D_e\gamma+c_e)}{(\gamma-\gamma_0)'T_e(\gamma-\gamma_0)}\\
&\quad=\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \frac{(D_e(\gamma-\gamma_0)+D_e\gamma_0+c_e)'A_e(D_e(\gamma-\gamma_0)+D_e\gamma_0+c_e)}{(\gamma-\gamma_0)'T_e(\gamma-\gamma_0)}\\
&\quad=\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \big\{(\gamma-\gamma_0)'D_e'A_eD_e(\gamma-\gamma_0)+(D_e\gamma_0+c_e)'A_e(D_e\gamma_0+c_e)\\
&\qquad\quad +2(\gamma-\gamma_0)'D_e'A_e(D_e\gamma_0+c_e)\big\}\big/(\gamma-\gamma_0)'T_e(\gamma-\gamma_0) \qquad (15)\\
&\quad=\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \big\{(\gamma-\gamma_0)'D_e'A_eD_e(\gamma-\gamma_0)+(D_e\gamma_0+c_e)'A_e(D_e\gamma_0+c_e)\\
&\qquad\quad +2|(\gamma-\gamma_0)'D_e'A_e(D_e\gamma_0+c_e)|\big\}\big/(\gamma-\gamma_0)'T_e(\gamma-\gamma_0), \qquad (16)
\end{aligned}
$$

where $|\cdot|$ denotes the absolute value. If the last term in the numerator of (15) is negative, we replace γ by $-\gamma+2\gamma_0$ leaving all other terms in (15) unchanged and arrive at (16). Now we are going to minimize (16) with respect to $c_e\in\mathbb{R}^k$, where $C_e\in\mathbb{R}^{k\times n_1}$ is kept fixed, and we conclude that, in determining an optimal linear affine estimator, we can restrict ourselves to those linear affine estimators $b=C_ey_1+c_e$ satisfying

$$A_ec_e=-A_eD_e\gamma_0. \qquad (17)$$

In the following we focus on the special solution

$$c_e=-D_e\gamma_0 \qquad (18)$$

of (17) which does not depend on the matrix A_e of weights. Inserting $D_e=(C_eX_1-I_k, C_e)$ and $\gamma_0=\binom{\beta_0}{u_{10}}(\beta_0\in\mathbb{R}^k, u_{10}\in\mathbb{R}^{n_1})$ into (18) we obtain

$$c_e=-(C_eX_1-I_k)\beta_0-C_eu_{10}. \qquad (19)$$

For linear affine estimators meeting (17) and, in particular, for those linear affine estimators satisfying (18) or (19), we get by (15):

$$
\begin{aligned}
&\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \frac{(D_e\gamma+c_e)'A_e(D_e\gamma+c_e)}{(\gamma-\gamma_0)'T_e(\gamma-\gamma_0)}\\
&\quad=\sup_{\substack{\gamma\in\mathbb{R}^{k+n_1}\\ \gamma-\gamma_0\notin N_e}} \frac{(\gamma-\gamma_0)'D_e'A_eD_e(\gamma-\gamma_0)}{(\gamma-\gamma_0)'T_e(\gamma-\gamma_0)}. \qquad (20)
\end{aligned}
$$

It remains to minimize the right-hand side of (20) with respect to $C_e\in\mathbb{R}^{k\times n_1}$. Reformulating (20) we set $\eta=\gamma-\gamma_0$ and decompose $\eta=\eta_1+\eta_2$ into its components

$$\eta_1 = \big(T_e^{1/2}\big)^+ T_e^{1/2}\eta = T_e^+ T_e \eta$$

and

$$\eta_2 = \big(I_{k+n_1} - (T_e^{1/2})^+ T_e^{1/2}\big)\eta = (I_{k+n_1} - T_e^+ T_e)\eta,$$

where $I_{k+n_1},\ T_e^+ \in \mathbb{R}^{(k+n_1)\times(k+n_1)}$ is the identity matrix and the Moore–Penrose inverse of T_e, respectively, η_1 is an element of the range R_e of T_e, and η_2 belongs to the null-space N_e of T_e. We obtain

$$\sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{\eta' D_e' A_e D_e \eta}{\eta' T_e \eta} = \sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{(\eta_1+\eta_2)' D_e' A_e D_e (\eta_1+\eta_2)}{\eta_1' T_e \eta_1},$$

and, applying part (i) of Definition 1, conclude that, in determining an optimal linear affine estimator of β, we have to restrict ourselves to those matrices $C_e \in \mathbb{R}^{k+n_1}$ satisfying the condition $\eta_2' D_e' A_e D_e \eta_2 = 0$ for all $\eta_2 \in N_e$. This condition is equivalent to

$$A_e^{1/2} D_e (I_{k+n_1} - T_e^+ T_e) = 0 \tag{21}$$

and to

$$A_e D_e (I_{k+n_1} - T_e^+ T_e) = 0. \tag{22}$$

We get for all $C_e \in \mathbb{R}^{k+n_1}$ satisfying (21):

$$\begin{aligned}
\sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{\eta' D_e' A_e D_e \eta}{\eta' T_e \eta} &= \sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{\eta' T_e^{1/2}\big(T_e^{1/2}\big)^+ D_e' A_e D_e \big(T_e^{1/2}\big)^+ T_e^{1/2}\eta}{\eta' T_e \eta}\\
&\leqslant \lambda_{\max}\big((T_e^{1/2})^+ D_e' A_e D_e (T_e^{1/2})^+\big)\\
&= \lambda_{\max}\big(A_e^{1/2} D_e T_e^+ D_e' A_e^{1/2}\big).
\end{aligned}$$

On the other hand, we obtain

$$\begin{aligned}
&\sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\neq 0}} \frac{\eta'\big(T_e^{1/2}\big)^+ D_e' A_e D_e \big(T_e^{1/2}\big)^+\eta}{\eta'\eta}\\
&\quad= \sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\neq 0}} \frac{\eta_1'\big(T_e^{1/2}\big)^+ D_e' A_e D_e \big(T_e^{1/2}\big)^+\eta_1}{\eta_1'\eta_1 + \eta_2'\eta_2}\\
&\quad\leqslant \sup_{\substack{\eta_1\in R_e\\ \eta_1\neq 0}} \frac{\eta_1'\big(T_e^{1/2}\big)^+ D_e' A_e D_e \big(T_e^{1/2}\big)^+\eta_1}{\eta_1'\eta_1}\\
&\quad= \sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{\eta' T_e^{1/2}\big(T_e^{1/2}\big)^+ D_e' A_e D_e \big(T_e^{1/2}\big)^+ T_e^{1/2}\eta}{\eta' T_e \eta}
\end{aligned}$$

for all $C_e \in \mathbb{R}^{k\times n_1}$. Therefore, the equation

$$\sup_{\substack{\eta\in\mathbb{R}^{k+n_1}\\ \eta\notin N_e}} \frac{\eta' D_e' A_e D_e \eta}{\eta' T_e \eta} = \lambda_{\max}\big(A_e^{1/2} D_e T_e^{+} D_e' A_e^{1/2}\big) \tag{23}$$

holds for all $C_e \in \mathbb{R}^{k\times n_1}$ satisfying (21). Thus, in order to determine an optimal linear affine estimator $b = C_e^* y_1 + c_e^*$ of β, we have to calculate C_e^* by minimizing $\lambda_{\max}(A_e^{1/2} D_e T_e^{+} D_e' A_e^{1/2})$ with respect to all $C_e \in \mathbb{R}^{k\times n_1}$ satisfying condition (21). Then, c_e^* can be taken from (19):

$$c_e^* = -(C_e^* X_1 - I_k)\beta_0 - C_e^* u_{10}. \tag{24}$$

Let us now consider the argument of the $\lambda_{\max}$ operator, i.e., the function $f_e : \mathbb{R}^{k\times n_1} \to \mathbb{R}^{k\times k}_{\text{sym}}$ with

$$\begin{aligned} f_e(C_e) &= A_e^{1/2} D_e T_e^{+} D_e' A_e^{1/2} \\ &= A_e^{1/2}\left[C_e(X_1, I_{n_1}) - (I_k, 0_{k\times n_1})\right] T_e^{+} \\ &\quad \times \left[C_e(X_1, I_{n_1}) - (I_k, 0_{k\times n_1})\right]' A_e^{1/2}, \end{aligned} \tag{25}$$

where $0_{k\times n_1} \in \mathbb{R}^{k\times n_1}$ is the null matrix and $\mathbb{R}^{k\times k}_{\text{sym}}$ denotes the set of all k-dimensional symmetric matrices equipped with the Löwner ordering defined by $A_1 \geqslant A_2$ iff $A_1 - A_2$ is n.n.d. Note that $f_e(\cdot)$ is convex with respect to the Löwner ordering. In Section 4, a general matrix problem is solved which, in particular, leads to a minimizer C_e^* of $f_e(\cdot)$ under condition (21). As $\lambda_{\max}(\cdot)$ is an isotone functional on $\mathbb{R}^{k\times k}_{\text{sym}}$, C_e^* also minimizes $\lambda_{\max}(f_e(\cdot))$ under condition (21) and thus, $b^* = C_e^* y_1 + c_e^*$ is an optimal linear affine estimator for β in model (1) and (2), where c_e^* is calculated by means of (24).

Before turning to the general solution we look at linear affine predictors in the next section.

3. Predicting observations

Let p be a predictor of the unknown observation $y_2 \in \mathbb{R}^{n_2}$ in the linear regression model (1) and (2), and let $(p - y_2)' A_p (p - y_2)$ be the weighted squared error of p, where $A_p \in \mathbb{R}^{n_2\times n_2}$ is a given n.n.d. matrix of weights. Proceeding in analogy to Section 2 we consider this weighted squared error relative to a suitably measured magnitude of the regression coefficient β and of the disturbance term $u = \binom{u_1}{u_2}$, i.e., we look at the ratio

$$\frac{(p - y_2)' A_p (p - y_2)}{\delta' T_p \delta}, \tag{26}$$

where β and u are combined to the column vector $\delta = \binom{\beta}{u} \in \mathbb{R}^{k+n}$, $T_p \in \mathbb{R}^{(k+n)\times(k+n)}$ is a given n.n.d. matrix, $T_p \neq 0$, and where δ is not an element of the null-space N_p of T_p.

Analogously to (13), (26) may be generalized to

$$\frac{(p-y_2)'A_p(p-y_2)}{(\delta-\delta_0)'T_p(\delta-\delta_0)} \tag{27}$$

with a given vector $\delta_0 = \begin{pmatrix}\beta_0\\ u_{10}\\ u_{20}\end{pmatrix} \in \mathbb{R}^{k+n}$.

Considering the class of linear affine predictors of y_2 we insert (1) and (2) into $p = C_p y_1 + c_p$ ($C_p \in \mathbb{R}^{n_2\times n_1}$, $c_p \in \mathbb{R}^{n_2}$) and write (27) equivalently as

$$\frac{(D_p\delta + c_p)'A_p(D_p\delta + c_p)}{(\delta-\delta_0)'T_p(\delta-\delta_0)}, \tag{28}$$

where $D_p = (C_pX_1 - X_2, C_p, -I_{n_2}) \in \mathbb{R}^{n_2\times(k+n)}$ with the identity matrix $I_{n_2} \in \mathbb{R}^{n_2\times n_2}$. In order to determine an optimal linear affine predictor we apply the minimax principle to quantity (28) and arrive at the following definition.

Definition 2. Let $A_p \in \mathbb{R}^{n_2\times n_2}$ and $T_p \in \mathbb{R}^{(k+n)\times(k+n)}$, $T_p \neq 0$, be given n.n.d. matrices, and let $\delta_0 \in \mathbb{R}^{k+n}$ be a given vector. Then a linear affine predictor $p^* = C_p^* y_1 + c_p^*$ ($C_p^* \in \mathbb{R}^{n_2\times n_1}$, $c_p^* \in \mathbb{R}^{n_2}$) for y_2 in model (1) and (2) is optimal if

(i) $$\sup_{\substack{\delta\in\mathbb{R}^{k+n}\\ \delta-\delta_0\notin N_p}} \frac{(D_p^*\delta + c_p^*)'A_p(D_p^*\delta + c_p^*)}{(\delta-\delta_0)'T_p(\delta-\delta_0)} < \infty$$

and if the inequality

(ii) $$\sup_{\substack{\delta\in\mathbb{R}^{k+n}\\ \delta-\delta_0\notin N_p}} \frac{(D_p^*\delta + c_p^*)'A_p(D_p^*\delta + c_p^*)}{(\delta-\delta_0)'T_p(\delta-\delta_0)} \leqslant \sup_{\substack{\delta\in\mathbb{R}^{k+n}\\ \delta-\delta_0\notin N_p}} \frac{(D_p\delta + c_p)'A_p(D_p\delta + c_p)}{(\delta-\delta_0)'T_p(\delta-\delta_0)}$$

holds for all $C_p \in \mathbb{R}^{n_2\times n_1}$ and all $c_p \in \mathbb{R}^{n_2}$. Here, we have set $D_p^* = (C_p^*X_1 - X_2, C_p^*, -I_{n_2})$ and $D_p = (C_pX_1 - X_2, C_p, -I_{n_2})$.

Obviously, our considerations following Definition 1 in Section 2 may directly be transferred to the prediction problem of this section.

First, in determining an optimal linear affine predictor, we can restrict ourselves to those predictors $p = C_p y_1 + c_p$ satisfying

$$A_p c_p = -A_p D_p \delta_0. \tag{29}$$

As in Section 2 we focus on the special solution

$$c_p = -D_p\delta_0 \tag{30}$$

of (29) which can equivalently be written as

$$c_p = -(C_pX_1 - X_2)\beta_0 - C_p u_{10} + u_{20}. \tag{31}$$

Second, it remains to minimize

$$\sup_{\substack{\delta\in\mathbb{R}^{k+n}\\ \delta-\delta_0\notin N_p}} \frac{(\delta-\delta_0)'D_p'A_pD_p(\delta-\delta_0)}{(\delta-\delta_0)'T_p(\delta-\delta_0)}$$

with respect to $C_p \in \mathbb{R}^{n_2\times n_1}$, and, as in Section 2, we conclude that this is equivalent to minimize

$$\lambda_{\max}\big(A_p^{1/2}D_pT_p^+D_p'A_p^{1/2}\big) \tag{32}$$

with respect to $C_p \in \mathbb{R}^{n_2\times n_1}$ under the condition

$$A_p^{1/2}D_p(I_{k+n}-T_p^+T_p)=0, \tag{33}$$

where $D_p=(C_pX_1-X_2, C_p, -I_{n_2})$ and where $I_{k+n}, T_p^+ \in \mathbb{R}^{(k+n)\times(k+n)}$ is the identity matrix and the Moore–Penrose inverse of T_p, respectively.

Third, defining $f_p : \mathbb{R}^{n_2\times n_1} \to \mathbb{R}^{n_2\times n_2}_{\text{sym}}$ with

$$\begin{aligned} f_p(C_p) &= A_p^{1/2}D_pT_p^+D_p'A_p^{1/2}\\ &= A_p^{1/2}\big[C_p(X_1, I_{n_1}, 0_{n_1\times n_2})-(X_2, 0_{n_2\times n_1}, I_{n_2})\big]\\ &\quad\times T_p^+\big[C_p(X_1, I_{n_1}, 0_{n_1\times n_2})-(X_2, 0_{n_2\times n_1}, I_{n_2})\big]'A_p^{1/2}, \end{aligned} \tag{34}$$

we see that any $C_p^* \in \mathbb{R}^{n_2\times n_1}$, which, under condition (33), minimizes $f_p(\cdot)$ with respect to the Löwner ordering on the set $\mathbb{R}^{n_2\times n_2}_{\text{sym}}$ of all n_2-dimensional symmetric matrices, also minimizes (32) under condition (33). Thus, such a minimizer C_p^* of $f_p(\cdot)$ under condition (33) leads to an optimal linear affine predictor $p^*=C_p^*y_1+c_p^*$ for y_2 in model (1) and (2), where c_p^* is given by (31):

$$c_p^* = -(C_p^*X_1-X_2)\beta_0 - C_p^*u_{10}+u_{20}. \tag{35}$$

4. Solving a general matrix problem

Looking at the problems of minimizing $f_e(\cdot)$ (see (25)) under condition (21) and of minimizing $f_p(\cdot)$ (see (34)) under condition (33) we realize that both problems are essentially of the same kind and can be summarized to minimizing the function $f : \mathbb{R}^{m\times r} \to \mathbb{R}^{l\times l}_{\text{sym}}$,

$$f(C)=A(CZ-F)T^+(CZ-F)'A' \tag{36}$$

under the condition

$$A(CZ-F)(I_s-T^+T)=0, \tag{37}$$

where $\mathbb{R}^{l\times l}_{\text{sym}}$ is the set of all l-dimensional symmetric matrices equipped with the Löwner ordering, and where $I_s \in \mathbb{R}^{s\times s}$ is the identity matrix. Furthermore, the matrices $A \in \mathbb{R}^{l\times m}$, $Z \in \mathbb{R}^{r\times s}$, $F \in \mathbb{R}^{m\times s}$, and $T \in \mathbb{R}^{s\times s}$, n.n.d., are assumed to be given. Note that similar problems arise when estimators for the regression coefficient are restricted to prior informations (see, e.g., [17]).

In estimating β we have

$$\begin{aligned} &l = m = k, \quad r = n_1, \quad s = k + n_1, \quad A = A_e^{1/2}, \\ &T = T_e, \quad C = C_e, \quad Z = (X_1, I_{n_1}) \quad \text{and} \quad F = (I_k, 0_{k\times n_1}). \end{aligned} \tag{38}$$

In the prediction problem we set

$$\begin{aligned} &l = m = n_2, \quad r = n_1, \quad s = k + n, \quad A = A_p^{1/2}, \quad T = T_p, \\ &C = C_p, \quad Z = (X_1, I_{n_1}, 0_{n_1\times n_2}) \quad \text{and} \quad F = (X_2, 0_{n_2\times n_1}, I_{n_2}). \end{aligned} \tag{39}$$

We see from (38) and (39) that, in both problems, the rank $rk(Z)$ of Z is maximal and equal to $r = n_1$.

The following proposition gives all the (Löwner-) minimizers of $f(\cdot)$, defined by (36), under condition (37). Here, $^-$ denotes any generalized inverse of the corresponding matrix.

Proposition 1. *Let $T \in \mathbb{R}^{s\times s}$, n.n.d., $A \in \mathbb{R}^{l\times m}$, $Z \in \mathbb{R}^{r\times s}$, and $F \in \mathbb{R}^{m\times s}$ be given matrices. Furthermore, we set $P = I_s - T^+T$ and $L = I_r - ZP(ZP)^-$.*

(i) *If $AFP(ZP)^-ZP \neq AFP$, then there exists no $C \in \mathbb{R}^{m\times r}$ satisfying condition* (37).

(ii) *If*

$$AFP(ZP)^-ZP = AFP, \tag{40}$$

then $C \in \mathbb{R}^{m\times r}$ minimizes $f(\cdot)$, given by (36), *under condition* (37), *iff it can be written as*

$$\begin{aligned} C = {} & C^* + (I_m - A^-A)(N - C^*) \\ & + A^-AN(I_r - LZT^+Z'L'(LZT^+Z'L')^-)L \end{aligned} \tag{41}$$

with some $N \in \mathbb{R}^{m\times r}$ and with

$$C^* = FP(ZP)^- + F(I_s - P(ZP)^-Z)T^+Z'L'(LZT^+Z'L')^-L. \tag{42}$$

If, in addition, $rk(Z) = r$, then (41) *reduces to*

$$C = C^* + (I_m - A^-A)(N - C^*), \tag{43}$$

and C^* is the unique minimizer of $f(\cdot)$ under condition (37) *whenever $rk(Z) = r$ and $rk(A) = m$.*

Note that no rank assumptions are required for (41).

Proof. Using the notations of the proposition we see that condition (37) is equivalent to $ACZP = AFP$ which has a solution in C iff the equation

$$AFP(ZP)^-ZP = AFP$$

holds (see (40)). Thus, part (i) of the proposition is shown.

In order to prove (ii) we assume the validity of (40) and see that (37) can equivalently be written as

$$C = A^-AFP(ZP)^- + M - A^-AMZP(ZP)^-, \tag{44}$$

where $M \in \mathbb{R}^{m\times r}$ is arbitrarily selected. Inserting (44) into (36) we obtain the function $g : \mathbb{R}^{m\times r} \to \mathbb{R}^{l\times l}_{\text{sym}}$ defined by

$$\begin{aligned} g(M) = f(C) = {} & A(MLZ + FP(ZP)^-Z - F)T^+ \\ & \times (MLZ + FP(ZP)^-Z - F)'A' \end{aligned} \tag{45}$$

which is to be minimized on $\mathbb{R}^{m\times r}$. As $g(\cdot)$ is convex, precisely these $M \in \mathbb{R}^{m\times r}$ with $\nabla_M g(M)(H) = 0$ for all $H \in \mathbb{R}^{m\times r}$ are the minimizers of $g(\cdot)$ and thus, applying (44), lead to the solutions of our original optimization problem. By direct calculation we see that $\nabla_M g(M)(H) = 0$ for all $H \in \mathbb{R}^{m\times r}$ is equivalent to the equation

$$AMLZT^+Z'L' = AF(I_s - P(ZP)^-Z)T^+Z'L' \tag{46}$$

having a solution in M iff

$$\begin{aligned} & AF(I_s - P(ZP)^-Z)T^+Z'L'(LZT^+Z'L')^-LZT^+Z'L' \\ & \quad = AF(I_s - P(ZP)^-Z)T^+Z'L'. \end{aligned} \tag{47}$$

Because of $T^+Z'L'(LZT^+Z'L')^-LZT^+Z'L' = T^+Z'L'$, Eq. (47) holds, and we conclude by (46) that all $M \in \mathbb{R}^{m\times r}$ with

$$\begin{aligned} M = {} & A^-AF(I_s - P(ZP)^-Z)T^+Z'L'(LZT^+Z'L')^- \\ & + N - A^-ANLZT^+Z'L'(LZT^+Z'L')^- \end{aligned} \tag{48}$$

are minimizers of $g(\cdot)$, where $N \in \mathbb{R}^{m\times r}$ is arbitrarily selected. Inserting (48) into (44) we realize after some rearrangements that precisely these $C \in \mathbb{R}^{m\times r}$ with

$$\begin{aligned} C = {} & C^* + (I_m - A^-A)(N - C^*) \\ & + A^-AN(I_r - LZT^+Z'L'(LZT^+Z'L')^-)L, \end{aligned}$$

where $N \in \mathbb{R}^{m\times r}$ is arbitrarily selected and where we have set

$$C^* = FP(ZP)^- + F(I_s - P(ZP)^-Z)T^+Z'L'\left(LZT^+Z'L'\right)^- L,$$

minimize $f(\cdot)$ under condition (37), i.e., solve our original minimization problem (see (41) and (42)).

In case of $rk(Z) = r$, Eq. (41) can be simplified considerably. First of all, because of $LZP = 0$, we obtain $LZT^+ = LZ(P + T^+)$. Furthermore, the equation

$$P + T^+ = I_s - T^+T + T^+ = I_s + T^+(I_s - T)$$

holds, we conclude that all eigenvalues of the symmetric matrix $P + T^+$ are positive, and thus, $P + T^+$ is p.d. Assuming now $rk(Z) = r$ we see that $Z(P + T^+)Z'$ is invertible, and we get

$$\begin{aligned}&\left(I_r - LZT^+Z'L'\left(LZT^+Z'L'\right)^-\right)L\\&\quad = (I_r - LZ(P + T^+)Z'L'(LZ(P + T^+)Z'L')^-)\\&\qquad \times LZ(P + T^+)Z'(Z(P + T^+)Z')^{-1} = 0.\end{aligned}$$

Thus, in case of $rk(Z) = r$, (41) reduces to (43). Furthermore, if $rk(Z) = r$ and $rk(A) = m$, then C^* is the unique minimizer of $f(\cdot)$ under condition (37). □

5. Applications

We now are going to apply the proposition of the preceding section to relative squared error estimation and prediction in linear regression. As the affine parts of the estimators and predictors can easily be determined by (24) and (35), respectively, we confine ourselves to their linear components. Using (38) and (39), C_e^* and C_p^* calculated by (42) lead to an optimal linear affine estimator and predictor, respectively, whenever Eq. (40) holds. Since C^* does not depend on A, we focus on this special solution of our minimization problem. Here, for each generalized inverse occurring in the proposition, we take the Moore–Penrose inverse.

Throughout this section we consider block-diagonal matrices

$$T_e = \begin{pmatrix} T_0 & 0 \\ 0 & W_{11}^{-1} \end{pmatrix}$$

and

$$T_p = \begin{pmatrix} T_0 & 0 \\ 0 & W^{-1} \end{pmatrix},$$

where $T_0 \in \mathbb{R}^{k\times k}$, n.n.d., and

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix},$$

p.d., with $W_{11} \in \mathbb{R}^{n_1\times n_1}$, $W_{12} = W_{21}' \in \mathbb{R}^{n_1\times n_2}$, $W_{22} \in \mathbb{R}^{n_2\times n_2}$ are given. Setting

$$Q = I_k - T_0^+T_0, \tag{49}$$

we see by direct calculation that, in the estimation problem, condition (40) is equivalent to

$$A_eQ(X_1Q)^+X_1Q = A_eQ, \tag{50}$$

whereas, in the prediction problem, (40) can equivalently be written as

$$A_pX_2Q(X_1Q)^+X_1Q = A_pX_2Q. \tag{51}$$

Obviously, (51) is closely related to (50): assuming that (50) holds without the prefactor A_e, i.e.,

$$Q(X_1Q)^+X_1Q = Q, \tag{52}$$

then (51) is valid for all matrices A_p and X_2. In the following subsections we suppose that (52) holds.

5.1. A relation between estimators and predictors

Inserting (38) and (39) into (42) and keeping in mind that we take the Moore–Penrose inverses as special generalized inverses, we arrive at the formulas

$$\begin{aligned} C_e^* = {} & Q(X_1Q)^+ + \big[(I_k - Q(X_1Q)^+X_1)T_0^+X_1' - Q(X_1Q)^+W_{11}\big] \\ & \times \big[(I_{n_1} - X_1Q(X_1Q)^+)(X_1T_0^+X_1' + W_{11})(I_{n_1} - X_1Q\,(X_1Q)^+)\big]^+ \end{aligned} \tag{53}$$

and

$$\begin{aligned} C_p^* = {} & X_2Q(X_1Q)^+ \\ & + \big[X_2(I_k - Q(X_1Q)^+X_1)T_0^+X_1' - X_2Q(X_1Q)^+W_{11} + W_{21}\big] \\ & \times \big[(I_{n_1} - X_1Q(X_1Q)^+)(X_1T_0^+X_1' + W_{11})(I_{n_1} - X_1Q(X_1Q)^+)\big]^+, \end{aligned} \tag{54}$$

respectively, which are related by the equation

$$C_p^* = X_2C_e^* + W_{21}W_{11}^{-1}(I_{n_1} - X_1C_e^*). \tag{55}$$

Note that (55) is formally equivalent to a well-known relation between estimators and predictors in linear regression presented and discussed, e.g., in [16].

In the following two subsections we look at the most extreme situations concerning T_0: we are going to consider the cases of $T_0 = 0$ and of T_0 p.d. The last subsection deals with an 'intermediate' T_0. Since an optimal predictor can be determined by (55) we restrict ourselves to calculating C_e^*.

5.2. Generalized least squares estimators

Here we set $T_0 = 0$ and obtain $Q = I_k$ (see (49)). Condition (52) holds iff $X_1^+X_1 = I_k$, i.e., iff $rk(X_1) = k$. Assuming $rk(X_1) = k$ and applying (53) we obtain

$$C_e^* = X_1^+ - X_1^+W_{11}\big[(I_{n_1} - X_1X_1^+)W_{11}(I_{n_1} - X_1X_1^+)\big]^+$$

which, by standard arguments (see, e.g., [1, pp. 90–91]), can be written as a generalized least squares estimator:

$$C_e^* = (X_1'W_{11}^{-1}X_1)^{-1}X_1'W_{11}^{-1}.$$

5.3. Kuks–Olman and ridge estimators

In this subsection we assume T_0 p.d. By (49) we get $Q = 0$. Thus, condition (52) trivially holds, and (53) leads to

$$C_e^* = T_0^{-1} X_1' (X_1 T_0^{-1} X_1' + W_{11})^{-1}.$$

Using the inversion formula we obtain

$$C_e^* = (T_0 + X_1' W_{11}^{-1} X_1)^{-1} X_1' W_{11}^{-1},$$

which can be interpreted as a Kuks–Olman or as a general ridge estimator. This special case was already considered in [3].

5.4. Combining generalized least squares and ridge estimators

In order to obtain an optimal estimator in Section 5.2 ($T_0 = 0$) we have to assume that the model matrix X_1 has full column rank, whereas in Section 5.3 (T_0 p.d.) no rank assumption is imposed on X_1, i.e., in Section 5.3 we allow for multicollinearity. In the present subsection we are going to use

$$T_0 = \begin{pmatrix} 0_{k_1 \times k_1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & T_2 \end{pmatrix} \in \mathbb{R}^{k \times k}$$

with a given p.d. matrix $T_2 \in \mathbb{R}^{k_2 \times k_2}$ ($k = k_1 + k_2$). Writing $X_1 = (X_{11}, X_{12})$ as a partitioned matrix with $X_{11} \in R^{n_1 \times k_1}$ and $X_{12} \in R^{n_1 \times k_2}$ we see that condition (52) is equivalent to $X_{11}^+ X_{11} = I_{k_1}$ and to $rk(X_{11}) = k_1$. Thus, in order to obtain an optimal estimator for β we have to assume that X_{11} does not contain any multicollinearity which, when existing, has to be concentrated in X_{12}.

We now suppose $rk(X_{11}) = k_1$, whereas no rank assumptions are made concerning X_{12}. Applying (53) and setting

$$\widetilde{W}_{11} = W_{11} + X_{12} T_2^{-1} X_{12}' \tag{56}$$

we get

$$C_e^* = \begin{pmatrix} C_{e1}^* \\ C_{e2}^* \end{pmatrix} \quad \text{with } C_{e1}^* \in \mathbb{R}^{k_1 \times n_1},\ C_{e2}^* \in \mathbb{R}^{k_2 \times n_1},$$

$$C_{e1}^* = X_{11}^+ - X_{11}^+ \widetilde{W}_{11} \big[(I_{n_1} - X_{11} X_{11}^+) \widetilde{W}_{11} (I_{n_1} - X_{11} X_{11}^+) \big]^+ \tag{57}$$

and

$$C_{e2}^* = T_2^{-1} X_{12}' \big[(I_{n_1} - X_{11} X_{11}^+) \widetilde{W}_{11} (I_{n_1} - X_{11} X_{11}^+) \big]^+. \tag{58}$$

To interprete (57) and (58) we simplify the term $C_{e2}^* + T_2^{-1} X_{12}' \widetilde{W}_{11}^{-1} X_{11} C_{e1}^*$ and obtain after some rearrangements

$$C_{e2}^* + T_2^{-1} X_{12}' \widetilde{W}_{11}^{-1} X_{11} C_{e1}^*$$
$$= T_2^{-1} X_{12}' \widetilde{W}_{11}^{-1} \Big(X_{11} X_{11}^+ + (I_{n_1} - X_{11} X_{11}^+) \widetilde{W}_{11}$$

$$\times\left[(I_{n_1}-X_{11}X_{11}^{+})\widetilde{W}_{11}(I_{n_1}-X_{11}X_{11}^{+})\right]^{+}\Big)$$
$$=T_2^{-1}X_{12}'\widetilde{W}_{11}^{-1}(P_1+P_2) \tag{59}$$

with $P_1 = X_{11}X_{11}^{+}$ and

$$P_2=(I_{n_1}-X_{11}X_{11}^{+})\widetilde{W}_{11}(I_{n_1}-X_{11}X_{11}^{+})\times\left[(I_{n_1}-X_{11}X_{11}^{+})\widetilde{W}_{11}(I_{n_1}-X_{11}X_{11}^{+})\right]^{+}.$$

As P_1 and P_2 are projection matrices satisfying $P_1P_2 = P_2P_1 = 0$, $P_1 + P_2$ is also a projection matrix. Furthermore, $rk(P_1 + P_2) = n_1$, $P_1 + P_2 = I_{n_1}$ and thus, using (59), we get

$$C_{e2}^{*}+T_2^{-1}X_{12}'\widetilde{W}_{11}^{-1}X_{11}C_{e1}^{*}=T_2^{-1}X_{12}'\widetilde{W}_{11}^{-1}$$

and

$$C_{e2}^{*}=T_2^{-1}X_{12}'\widetilde{W}_{11}^{-1}\left[I_{n_1}-X_{11}C_{e1}^{*}\right]. \tag{60}$$

Writing (57) as a generalized least squares estimator, looking at (56), and applying the inversion formula to the term $\widetilde{W}_{11}^{-1}$ in (60) we arrive at the expressions

$$C_{e1}^{*}=\left(X_{11}'\widetilde{W}_{11}^{-1}X_{11}\right)^{-1}X_{11}'\widetilde{W}_{11}^{-1} \tag{61}$$

and

$$C_{e2}^{*}=\left(T_2+X_{12}'W_{11}^{-1}X_{12}\right)^{-1}X_{12}'W_{11}^{-1}\left[I_{n_1}-X_{11}C_{e1}^{*}\right]. \tag{62}$$

For the sake of simplicity we assume in the following that the affine part of the estimator vanishes. Thus, $b^* = C_e^* y_1 = \binom{C_{e1}^*}{C_{e2}^*} y_1$ is an optimal estimator for $\beta = \binom{\beta_1}{\beta_2}$ in our linear regression model $y_1 = X_{11}\beta_1 + X_{12}\beta_2 + u_1 (\beta_1 \in \mathbb{R}^{k_1}, \beta_2 \in \mathbb{R}^{k_2})$ with a full column rank matrix X_{11}. We see from (61) that β_1 is estimated by means of a generalized least squares estimator, where the sum $X_{12}\beta_2 + u_1$ is taken as an aggregated error term. In consequence of this augmentation of u_1 the matrix W_{11} is transformed into $\widetilde{W}_{11}$ (see (56)). Note that X_{12} may contain some multicollinearity. Once the estimator $C_{e1}^* y_1$ for β_1 is determined, the second part β_2 of the regression coefficient β is estimated by a Kuks–Olman or ridge estimator, where the 'residual' term $y_1 - X_{11}C_{e1}^* y_1$ is used instead of the 'full' observation y_1.

Thus, our optimal estimator is calculated by a two-step procedure: first the 'good' part β_1 of the regression coefficient is estimated and then the 'bad' one, β_2.

6. Concluding remarks

In this paper a general relative squared error approach is given to the interconnected estimation and prediction problems in linear regression. It should be noted that this approach does not include any stochastic aspects and that it can also be viewed as some kind of a 'signal-to-noise' ratio approach. Furthermore, looking at (12) or (13) and at (26) or (27) we see that our concept is not restricted to linear regression analysis but might also be applied to more general models.

Acknowledgements

The authors wish to thank an anonymous referee for his helpful comments and suggestions.

References

[1] A. Albert, Regression and the Moore–Penrose Pseudoinverse, Academic Press, New York, 1972.
[2] B.F. Arnold, P. Stahlecker, Prediction in linear regression combining crisp data and fuzzy prior information, Statist. Decisions 16 (1998) 19–33.
[3] B.F. Arnold, P. Stahlecker, Another view of the Kuks–Olman estimator, J. Statist. Plann. Inference 89 (2000) 169–174.
[4] N. Gaffke, B. Heiligers, Bayes, admissible, and linear minimax estimators in linear models with restricted parameter space, Statistics 20 (1989) 487–508.
[5] A.S. Goldberger, Best linear unbiased prediction in the generalized linear regression model, J. Amer. Statist. Assoc. 57 (1962) 369–375.
[6] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation of nonorthogonal problems, Technometrics 12 (1970) 55–67.
[7] K. Hoffmann, Characterization of minimax linear estimators in linear regression, Math. Operationsforsch. Statist., Ser. Statist. 10 (1979) 19–26.
[8] J. Kuks, W. Olman, Minimax linear estimation of regression coefficients I, Iswestija Akad. Nauk Estonskoj SSR 20 (1971) 480–482 (in Russian).
[9] J. Kuks, W. Olman, Minimax linear estimation of regression coefficients II, Iswestija Akad. Nauk Estonskoj SSR 21 (1972) 66–72 (in Russian).
[10] H. Läuter, A minimax linear estimator for linear parameters under restrictions in form of inequalities, Math. Operationsforsch. Statist., Ser. Statist. 6 (1975) 689–695.
[11] A. Markiewicz, Characterization of general ridge estimators, Statist. Probab. Lett. 27 (1996) 145–148.
[12] J. Pilz, Minimax linear regression estimation with symmetric parameter restriction, J. Statist. Plann. Inference 13 (1986) 297–318.
[13] J. Pilz, Bayesian Estimation and Experimental Design in Linear Regression Models, Wiley, New York, 1991.
[14] F. Pukelsheim, Optimal Design of Experiments, Wiley, New York, 1993.
[15] C.R. Rao, Estimation of parameters in a linear model, Ann. Statist. 4 (1976) 1023–1037.
[16] C.R. Rao, H. Toutenburg, Linear Models, Least Squares and Alternatives, Springer, Heidelberg, 1995.
[17] P. Schönfeld, H.J. Werner, Beiträge zur Theorie und Anwendung linearer Mo-delle, in: W. Krelle (Ed.), Ökonomische Prognose-, Entscheidungs- und Gleichgewichtsmo-delle, VCH Verlagsgesellschaft, Weinheim, 1986, pp. 251–262 (in German).
[18] P. Stahlecker, A priori Information und Minimax-Schätzung im Linearen Regressionsmodell, Mathematical Systems in Economics, Athenäum, Frankfurt, 1987 (in German).
[19] P. Stahlecker, G. Trenkler, Minimax estimation in linear regression with singular covariance structure and convex polyhedral constraints, J. Statist. Plann. Inference 36 (1993) 185–196.