REGULAR ARTICLE

An alternative look at the linear regression model

Oskar Maria Baksalary[1] · Götz Trenkler[2]

Abstract

An alternative look at the linear regression model is taken by proposing an original treatment of a full column rank model (design) matrix. In such a situation, the Moore–Penrose inverse of the matrix can be obtained by utilizing a particular formula which is applicable solely when a matrix to be inverted can be columnwise partitioned into two matrices of disjoint ranges. It turns out that this approach, besides simplifying derivations, provides a novel insight into some of the notions involved in the model and reduces computational costs needed to obtain sought estimators. The paper contains also a numerical example based on astronomical observations of the localization of Polaris, demonstrating usefulness of the proposed approach.

1 Introduction

The problem of curve fitting on the basis of a finite number of observations arises in almost all areas where mathematics is applied and one of the most powerful tools used for this purpose is based on the least squares method. Over the years a rich sample of results occurred in the literature providing an indisputable evidence that the matrix analysis concepts and techniques offer handy means to apply the method. The present paper constitutes a further contribution to this stream of considerations by demonstrating how an expression for the Moore–Penrose inverse of a columnwise partitioned matrix derived in Baksalary and Baksalary (2007, Theorem 1) may be advantageously utilized to deal with the problems originating from linear regression.

✉ Oskar Maria Baksalary
  OBaksalary@gmail.com

1 Faculty of Physics, Adam Mickiewicz University, ul. Uniwersytetu Poznańskiego 2, 61-614 Poznan, Poland

2 Faculty of Statistics, Dortmund University of Technology, Dortmund, Germany

Among the benefits resulting from the proposed approach one may mention: a simplification of derivations, a novel insight into the notions involved in the regression model, and reduction of computational costs necessary to obtain sought estimators. Furthermore, by simplifying inevitable mathematical operations, the proposed approach offers an attractive alternative to the researchers, who are not keen on exploiting more advanced than necessary matrix methods or utilizing software packages which do not provide a comprehensive control over the processed data, as the approach enables to perform calculations almost "by hand" preserving an insight into every step of linear regression.

The aforementioned representation of the Moore–Penrose inverse established in Baksalary and Baksalary (2007, Theorem 1) is recalled in the following lemma.

Lemma 1.1 *Let $\mathbf{A}$ be an $n \times m$, $m \geq 2$, real matrix columnwise partitioned as $\mathbf{A} = (\mathbf{A}_1 : \mathbf{A}_2)$, with $\mathbf{A}_i$ denoting $n \times m_i$, $i = 1, 2$, matrices such that $m_1 + m_2 = m$. Furthermore, let the ranges of $\mathbf{A}_1$ and $\mathbf{A}_2$ be disjoint. Then the Moore–Penrose inverse of $\mathbf{A}$ is of the form*

$$\mathbf{A}^\dagger = \begin{pmatrix} (\mathbf{Q}_2\mathbf{A}_1)^\dagger \\ (\mathbf{Q}_1\mathbf{A}_2)^\dagger \end{pmatrix}, \tag{1}$$

where $\mathbf{Q}_i$, $i = 1, 2$, is the orthogonal projector onto the null space of the transpose of $\mathbf{A}_i$ and $(\mathbf{Q}_i\mathbf{A}_j)^\dagger$, $i = 1, 2$, $i \neq j$, is the Moore–Penrose inverse of $\mathbf{Q}_i\mathbf{A}_j$.

In the next section we briefly discuss particular linear regression models, shading a spotlight on issues in which the Moore–Penrose inverse of a columnwise partitioned matrix naturally emerges. These considerations are followed by Sect. 3, which contains an example demonstrating applicability of the present approach. The data used in the example originated from the observations of the position of Polaris made on December 12, 1983, by S.G. Brewer, which were afterwards used by Pedler (1993) to develop a solution to an (as the author claims) "astronomical problem" aimed at fitting a circle to a set of points. Section 4 provides a number of remarks concerned with the proposed approach.

2 Particular linear regression models

All matrices occurring in what follows are of real entries and the superscript $'$ stands for a matrix transpose. Let us consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{2}$$

where $\mathbf{y}$ is an $n \times 1$ random vector of observations, $\mathbf{X}$ is an $n \times p$ known model (design) matrix of constants, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\mathbf{u}$ is an $n \times 1$ vector of unknown errors. The entries of the vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)'$ are assumed to have a mean of zero and (unknown) variance σ^2, and each pair $u_i, u_j, i \neq j$, is assumed to be uncorrelated, i.e., the expectation vector and the covariance matrix of

$\mathbf{u}$ are $\mathsf{E}(\mathbf{u}) = \mathbf{0}$ and $\mathsf{Cov}(\mathbf{u}) = \sigma^2 \mathbf{I}_n$, respectively. Customarily, the symbol $\mathbf{I}_n$ stands for the identity matrix of order n. We also assume that the matrix $\mathbf{X}$ is of full column rank. Then, the least squares estimator (LSE) of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

It is worth emphasizing that the assumption that $\mathbf{X}$ is of full column rank plays a crucial role, and is most often made to assure uniqueness of the estimator of $\boldsymbol{\beta}$; see Puntanen et al. (2011, p. 34). It turns out that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}^\dagger$, i.e., the Moore–Penrose inverse of $\mathbf{X}$; see Appendix A.

To calculate $\mathbf{X}^\dagger$, instead of bothering with the inverse of $\mathbf{X}'\mathbf{X}$, we may write the regressor matrix $\mathbf{X}$ in the columnwise partitioned form

$$\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2), \tag{3}$$

where $\mathbf{X}_i$, $i = 1, 2$, denote $n \times p_i$ matrices such that $p_1 + p_2 = p$. Since $\mathbf{X}$ is of full column rank, it follows that

$$\mathcal{R}(\mathbf{X}_1) \cap \mathcal{R}(\mathbf{X}_2) = \{\mathbf{0}\}, \tag{4}$$

where $\mathcal{R}(.)$ stands for the column space (range) of a matrix argument. According to Lemma 1.1, the Moore–Penrose inverse of a matrix of the form (3), such that (4) is satisfied, can be expressed as

$$\mathbf{X}^\dagger = \begin{pmatrix} (\mathbf{Q}_2\mathbf{X}_1)^\dagger \\ (\mathbf{Q}_1\mathbf{X}_2)^\dagger \end{pmatrix}, \tag{5}$$

where $\mathbf{Q}_i = \mathbf{I}_n - \mathbf{X}_i\mathbf{X}_i^\dagger$ is the orthogonal projector onto $\mathcal{N}(\mathbf{X}_i')$, the null space of $\mathbf{X}_i'$, $i = 1, 2$; see Appendix A. Note that the condition (4), under which the representation of the Moore–Penrose inverse (5) is valid, is weaker than the requirement that $\mathbf{X}$ specified in (3) is of full column rank (the assumption which will be extensively exploited in what follows).

Let us consider a simple linear regression model

$$\mathbf{y} = \beta_0\mathbf{1} + \beta_1\mathbf{x} + \mathbf{u},$$

where $\beta_0, \beta_1 \in \mathbb{R}$, $\mathbf{1} = (1, 1, \ldots, 1)'$ is the vector of n ones and $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ is the vector of observations on one regressor variable. The vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)'$ consists of the unknown errors. To obtain the LSE of the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ we may partition $n \times 2$ matrix $\mathbf{X}$ as

$$\mathbf{X} = (\mathbf{1} : \mathbf{x}), \tag{6}$$

where $\mathcal{R}(\mathbf{1}) \cap \mathcal{R}(\mathbf{x}) = \{\mathbf{0}\}$ since we assume that $\mathbf{X}$ is of full column rank. By (5) it follows that

$$\mathbf{X}^\dagger = \begin{pmatrix} (\mathbf{Q_x 1})^\dagger \\ (\mathbf{Q_1 x})^\dagger \end{pmatrix}, \tag{7}$$

where $\mathbf{Q_x 1} = (\mathbf{I}_n - \mathbf{xx}^\dagger)\mathbf{1}$ and $\mathbf{Q_1 x} = (\mathbf{I}_n - \mathbf{11}^\dagger)\mathbf{x}$ are both column vectors. Thus, by the identity (A1) given in Appendix A,

$$(\mathbf{Q_x 1})^\dagger = \frac{\mathbf{1'Q_x}}{\mathbf{1'Q_x 1}} \quad \text{and} \quad (\mathbf{Q_1 x})^\dagger = \frac{\mathbf{x'Q_1}}{\mathbf{x'Q_1 x}}. \tag{8}$$

Consequently, the LSE of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^\dagger \mathbf{y} = \begin{pmatrix} (\mathbf{Q_x 1})^\dagger \mathbf{y} \\ (\mathbf{Q_1 x})^\dagger \mathbf{y} \end{pmatrix} = \begin{pmatrix} (\mathbf{1'Q_x 1})^{-1}\mathbf{1'Q_x y} \\ (\mathbf{x'Q_1 x})^{-1}\mathbf{x'Q_1 y} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}. \tag{9}$$

From (8) we obtain

$$(\mathbf{Q_x 1})^\dagger \mathbf{x} = 0, \ (\mathbf{Q_1 x})^\dagger \mathbf{1} = 0, \ (\mathbf{Q_x 1})^\dagger \mathbf{1} = 1, \ (\mathbf{Q_1 x})^\dagger \mathbf{x} = 1,$$

whence

$$\mathsf{E}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \mathsf{E}(\hat{\beta}_0) \\ \mathsf{E}(\hat{\beta}_1) \end{pmatrix} = \begin{pmatrix} (\mathbf{Q_x 1})^\dagger \mathsf{E}(\mathbf{y}) \\ (\mathbf{Q_1 x})^\dagger \mathsf{E}(\mathbf{y}) \end{pmatrix} = \begin{pmatrix} (\mathbf{Q_x 1})^\dagger (\beta_0 \mathbf{1} + \beta_1 \mathbf{x}) \\ (\mathbf{Q_1 x})^\dagger (\beta_0 \mathbf{1} + \beta_1 \mathbf{x}) \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Let us use the symbol $\mathbf{H}$ to denote the so-called hat-matrix, which represents the orthogonal projector onto $\mathcal{R}(\mathbf{X})$, i.e., $\mathbf{H} = \mathbf{XX}^\dagger$. Then, the identities (6)–(8) entail

$$\mathbf{H} = \mathbf{1}(\mathbf{Q_x 1})^\dagger + \mathbf{x}(\mathbf{Q_1 x})^\dagger = \frac{\mathbf{11'Q_x}}{\mathbf{1'Q_x 1}} + \frac{\mathbf{xx'Q_1}}{\mathbf{x'Q_1 x}}, \tag{10}$$

which means that the hat-matrix is a sum of two matrices of rank one. Furthermore, each of the summands involved in (10) is idempotent. This observation leads to the conclusion (see e.g., Rao and Mitra (1971, Theorem 5.1.2)) that the matrices necessarily commute and their product is equal to the zero matrix, i.e.,

$$\mathbf{1}(\mathbf{Q_x 1})^\dagger \mathbf{x}(\mathbf{Q_1 x})^\dagger = \mathbf{0} = \mathbf{x}(\mathbf{Q_1 x})^\dagger \mathbf{1}(\mathbf{Q_x 1})^\dagger.$$

Since $\mathbf{1} \in \mathcal{R}(\mathbf{X})$, it follows that $\mathbf{H1} = \mathbf{1}$. Hence, by denoting $\hat{\mathbf{y}} = \mathbf{Hy}$, we see that $\mathbf{1'\hat{y}} = \mathbf{1'Hy} = \mathbf{1'y}$, which gives $\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$. Furthermore, by putting $\hat{\mathbf{u}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$, we obtain $\mathbf{1'\hat{u}} = \mathbf{1'}(\mathbf{I}_n - \mathbf{H})\mathbf{y} = \mathbf{0}$, so $\sum_{i=1}^n \hat{u}_i = 0$. Another consequence of $\mathbf{1} \in \mathcal{R}(\mathbf{X})$ is the identity

$$\|(\mathbf{I}_n - \mathbf{J})\mathbf{y}\|^2 = \|(\mathbf{H} - \mathbf{J})\mathbf{y}\|^2 + \|(\mathbf{I}_n - \mathbf{H})\mathbf{y}\|^2, \tag{11}$$

with $\mathbf{J} = \mathbf{11}^\dagger$; see Puntanen et al. ([2011](#), Proposition 8.5). Alternatively, the equality ([11](#)) can be expressed as $SST = SSR + SSE$, where SST stands for the total sum of squares, SSR for the regression sum of squares, and SSE for the residual sum of squares. The coefficient of determination defined as

$$R^2 = \frac{SSR}{SST}$$

turns out to be

$$R^2 = \frac{\|(\mathbf{H} - \mathbf{J})\mathbf{y}\|^2}{\|(\mathbf{I}_n - \mathbf{J})\mathbf{y}\|^2} = \frac{\mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y}}{\mathbf{y}'(\mathbf{I}_n - \mathbf{J})\mathbf{y}}. \tag{12}$$

Clearly, $R^2 \geq 0$. Another observation is that $\mathbf{I}_n - \mathbf{J} - (\mathbf{H} - \mathbf{J}) = \mathbf{I}_n - \mathbf{H}$ is nonnegative definite (as $\mathbf{I}_n - \mathbf{H}$ is the orthogonal projector onto $\mathcal{N}(\mathbf{X}')$). Hence, $\mathbf{H} - \mathbf{J} \overset{\mathsf{L}}{\leq} \mathbf{I}_n - \mathbf{J}$, where the symbol $\overset{\mathsf{L}}{\leq}$ denotes the Löwner partial ordering, from where we conclude that $0 \leq R^2 \leq 1$. The fact that values of R^2 are restricted to the interval $[0, 1]$ is known in the literature (see e.g., Davidson and MacKinnon ([1993](#), p. 14)), but usually it is demonstrated in rather more involved way than in the present paper.

Consider now the general linear model with intercept

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = (\mathbf{1} : \mathbf{X}) \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix} + \mathbf{u},$$

where $\mathbf{X}$ and $\boldsymbol{\beta}$ are of dimensions $n \times (p-1)$ and $(p-1) \times 1$, respectively, and $(\mathbf{1} : \mathbf{X})$ is assumed to be of full column rank. Then, by analogy to ([9](#)), the LSE of $(\beta_0, \boldsymbol{\beta})'$ is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} (\mathbf{Q_X 1})^\dagger \mathbf{y} \\ (\mathbf{Q_1 X})^\dagger \mathbf{y} \end{pmatrix}. \tag{13}$$

On account of ([A1](#)), we obtain $(\mathbf{Q_X 1})^\dagger = (\mathbf{1}'\mathbf{Q_X 1})^{-1}\mathbf{1}'\mathbf{Q_X}$. Hence, $(\mathbf{Q_X 1})^\dagger = (\mathbf{Q_X 1})^\dagger \mathbf{Q_X}$. Similarly, we arrive at $(\mathbf{Q_1 X})^\dagger = (\mathbf{Q_1 X})^\dagger \mathbf{Q_1}$. In consequence, since $\mathsf{E}(\mathbf{y}) = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta}$, and since both, $\mathbf{Q_1 X}$ and $\mathbf{Q_X 1}$, are of full column ranks, we have

$$\begin{pmatrix} \mathsf{E}(\hat{\beta}_0) \\ \mathsf{E}(\hat{\boldsymbol{\beta}}) \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix}.$$

Note that the hat-matrix turns out to be

$$\mathbf{H} = \mathbf{1}(\mathbf{Q_X 1})^\dagger + \mathbf{X}(\mathbf{Q_1 X})^\dagger, \tag{14}$$

a sum of two matrices of which the former is of rank one and the latter is of the same rank as matrix $\mathbf{X}$, i.e., $p - 1$. Similarly as above, both summands which determine

the hat-matrix specified in (14) are commuting idempotents, whose product equals the zero matrix.

The formula (14) (as well as its particular case (10)) can be viewed as an alternative to the representation of $\mathbf{H}$ as a sum of two orthogonal projectors, which reads $\mathbf{H} = \mathbf{J} + \mathbf{P_{CX}}$, where $\mathbf{P_{CX}} = \mathbf{CX}(\mathbf{CX})^{\dagger}$, with $\mathbf{C}$ denoting the so-called centering matrix defined as $\mathbf{C} = \mathbf{I}_n - \mathbf{J}$; see Puntanen et al. (2011, formula (8.108)). It is clear that $\mathbf{JP_{CX}} = \mathbf{0}$.

3 Applications

As in Pedler (1993, Sect. 6), we consider now the linear regression model with 4 regressors

$$
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & a_1 & b_1 \\ 1 & 0 & a_2 & b_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & a_n & b_n \\ 0 & 1 & b_1 & -a_1 \\ 0 & 1 & b_2 & -a_2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & b_n & -a_n \end{pmatrix} \begin{pmatrix} p \\ q \\ u \\ v \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}.
\tag{15}
$$

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$, $\mathbf{b} = (b_1, b_2, \ldots, b_n)'$, $\boldsymbol{\beta} = (p, q, u, v)'$, $\mathbf{c} = (c_1, c_2, \ldots, c_n)'$, $\mathbf{d} = (d_1, d_2, \ldots, d_n)'$. Furthermore, let $\mathbf{X}$ be a $2n \times 4$ matrix of the form

$$
\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{a} & \mathbf{b} \\ \mathbf{0} & \mathbf{1} & \mathbf{b} & -\mathbf{a} \end{pmatrix}.
\tag{16}
$$

Then (15) can be written as

$$
\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}.
$$

As the matrix $\mathbf{X}$ is of full column rank, we can determine the LSE of $\boldsymbol{\beta}$ by applying the representation derived in Baksalary and Trenkler (2021, Example 1) as a consequence of Baksalary and Baksalary (2007, Theorem 1). In order to take advantage of this result, let

$$
A = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} \left(a_i^2 + b_i^2\right)}, \quad B = \frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} \left(a_i^2 + b_i^2\right)},
\tag{17}
$$

$$N = n - A \sum_{i=1}^{n} a_i - B \sum_{i=1}^{n} b_i, \ M = \sum_{i=1}^{n} (a_i^2 + b_i^2) - \frac{1}{n} \left[\left(\sum_{i=1}^{n} a_i \right)^2 + \left(\sum_{i=1}^{n} b_i \right)^2 \right],$$

(18)

$$\mathbf{e} = \begin{pmatrix} \mathbf{1} - A\mathbf{a} - B\mathbf{b} \\ -A\mathbf{b} + B\mathbf{a} \end{pmatrix}, \ \mathbf{f} = \begin{pmatrix} A\mathbf{b} - B\mathbf{a} \\ \mathbf{1} - A\mathbf{a} - B\mathbf{b} \end{pmatrix}, \ \mathbf{g} = \begin{pmatrix} \mathbf{a} - \overline{a}\mathbf{1} \\ \mathbf{b} - \overline{b}\mathbf{1} \end{pmatrix}, \ \mathbf{h} = \begin{pmatrix} \mathbf{b} - \overline{b}\mathbf{1} \\ -\mathbf{a} + \overline{a}\mathbf{1} \end{pmatrix},$$

(19)

where $\overline{a} = \frac{1}{n} \sum_{i=1}^{n} a_i$ and $\overline{b} = \frac{1}{n} \sum_{i=1}^{n} b_i$. Then, on account of Baksalary and Trenkler (2021, formula (8)), we obtain a very handy representation of the Moore–Penrose inverse of the model matrix (16), namely

$$\mathbf{X}^{\dagger} = \begin{pmatrix} \begin{pmatrix} N^{-1} & 0 \\ 0 & N^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{e}' \\ \mathbf{f}' \end{pmatrix} \\ \begin{pmatrix} M^{-1} & 0 \\ 0 & M^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{g}' \\ \mathbf{h}' \end{pmatrix} \end{pmatrix}.$$

(20)

Hence, analogously to (13), the LSE of the parameter vector $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{u} \\ \hat{v} \end{pmatrix} = \mathbf{X}^{\dagger}\mathbf{z} = \begin{pmatrix} N^{-1}\mathbf{e}'\mathbf{z} \\ N^{-1}\mathbf{f}'\mathbf{z} \\ M^{-1}\mathbf{g}'\mathbf{z} \\ M^{-1}\mathbf{h}'\mathbf{z} \end{pmatrix}.$$

(21)

Let us now demonstrate the usefulness of the expressions (20) and (21) by applying them to a set of real data.

Example 3.1 As mentioned in Introduction, Pedler (1993) considered "astronomical problem" of fitting a circle to a set of points and solved it from first principles. The considerations in Pedler (1993) contain also an example which exploits the data collected by S.G. Brewer from the observations of the position of Polaris made on December 12, 1983. The data are given in Table 1.

The data provided in Table 1 enable to calculate the scalars A, B, M, N as well as the vectors $\mathbf{e}, \mathbf{f}, \mathbf{g}, \mathbf{h}$ defined in (17)–(19). Hence, we obtain the vector inner products of the four vectors and the vector $\mathbf{z}$. It should be emphasized that these straightforward calculations involve only scalars and vectors and require relatively low computational cost; for details on advantages of the algorithm to calculate the Moore–Penrose inverse which takes into account columnwise partitioning into range disjoint matrices see Baksalary and Trenkler (2021). The outcomes of these computations are provided in Table 2.

In the light of (21), we arrive at the components of the estimator of $\hat{\boldsymbol{\beta}}$, which are given in Table 3. As expected, the values coincide with the ones given in Pedler (1993, Table 2).

Table 1 At the ith observation at time t_i, (x_i, y_i) denotes (observed) cartesian coordinates of the Polaris position, whereas $a_i = \cos(2\pi t_i/T)$ and $b_i = \sin(2\pi t_i/T)$, where $T = 23h\ 56m\ 04s$ is the length of the sidereal day. The values are recalled from Pedler (Pedler 1993, Table 1)

i	x_i	y_i	a_i	b_i
1	2.7	1.9	−0.1829	− 0.9831
2	2.6	2.0	−0.0963	− 0.9953
3	2.4	2.05	0.0348	− 0.9994
4	2.2	2.15	0.1653	− 0.9862
5	2.0	2.2	0.2720	− 0.9623
6	1.7	2.3	0.4935	− 0.8698
7	1.4	2.3	0.6204	− 0.7843
8	1.2	2.3	0.7023	− 0.7119
9	1.0	2.2	0.8026	− 0.5965
10	0.8	2.1	0.8942	− 0.4476

Table 2 Values of the scalars calculated on account of the data provided in Table 1

A	B	M	N	$e'z$	$f'z$	$g'z$	$h'z$
0.371	−0.834	1.677	1.677	2.649	1.174	−2.271	−1.453

Table 3 Least squares estimators obtained on account of (21)

$\hat{p}$	$\hat{q}$	$\hat{u}$	$\hat{v}$
1.579	0.700	−1.354	−0.867

Table 4 Values of SST, SSR, SSE, and R^2 obtained on account of (11) and (12)

SST	SSR	SSE	R^2
0.170	0.072	0.098	0.426

The values of the total sum of squares SST, regression sum of squares SSR, residual sum of squares SSE, and the coefficient of determination R^2 are provided in Table 4.

4 Supplementary remarks

In the linear regression models considered in Sect. 2 it was assumed that the model matrices are of full column ranks and that the vector **1** is one of the columns. Such assumptions are well justified as they correspond to several common situations. However, the present approach enables to generalize the considerations by weakening the assumption that the model matrix is of full column rank to the requirement that it can be columnwise partitioned into two range disjoint matrices and by relaxing the assumption that one of the columns is the vector **1**. To demonstrate this fact, let us assume that the model matrix **X** in (2) is partitioned in accordance with (3), i.e.,

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}, \tag{22}$$

where vectors $\boldsymbol{\beta}_i$ are of orders $p_i \times 1$, $i = 1, 2$. Provided that the condition (4) holds, by (5) we conclude that the LSE of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$ is given by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} (\mathbf{Q}_2\mathbf{X}_1)^\dagger\mathbf{y} \\ (\mathbf{Q}_1\mathbf{X}_2)^\dagger\mathbf{y} \end{pmatrix}. \tag{23}$$

Visibly, the expression (13) is obtained from (23) by taking $\mathbf{X}_1 = \mathbf{1}$ and $\mathbf{X}_2 = \mathbf{X}$. Furthermore, from (3) and (5) we obtain

$$\mathbf{H} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{Q}_2\mathbf{X}_1)^\dagger\mathbf{X}_1'\mathbf{Q}_2 + \mathbf{X}_2(\mathbf{X}_2'\mathbf{Q}_1\mathbf{X}_2)^\dagger\mathbf{X}_2'\mathbf{Q}_1, \tag{24}$$

which under $\mathbf{X}_1 = \mathbf{1}$ and $\mathbf{X}_2 = \mathbf{X}$ leads to the formula (14). Note that the expression (24) is given in Puntanen et al. (2011, Proposition 16.1) along with its equivalent counterparts, one of which is (4).

Another evidence of the applicability of Lemma 1.1 in statistical estimation theory was provided in Baksalary and Trenkler (2021), by deriving an original representation for the best linear unbiased estimator (BLUE) of $\mathbf{X}\boldsymbol{\beta}$ under the generalized version of the (consistent) linear model (2) with $\mathsf{Cov}(\mathbf{u}) = \sigma^2\mathbf{V}$, where $\mathbf{V}$ denotes a known $n \times n$ positive semidefinite matrix. It was shown in Baksalary and Trenkler (2021, Example 4) that $\mathbf{Gy}$ with

$$\mathbf{G} = \mathbf{X}(\mathbf{Q}_{\mathbf{VQ_X}}\mathbf{X})^\dagger = \mathbf{X}(\mathbf{X}'\mathbf{Q}_{\mathbf{VQ_X}}\mathbf{X})^\dagger\mathbf{X}'\mathbf{Q}_{\mathbf{VQ_X}} \tag{25}$$

is BLUE of $\mathbf{X}\boldsymbol{\beta}$.

Analogously, we can derive representations for BLUE of both, $\mathbf{X}_1\boldsymbol{\beta}_1$ and $\mathbf{X}_2\boldsymbol{\beta}_2$ under the model (22) when $\mathsf{Cov}(\mathbf{u}) = \sigma^2\mathbf{V}$. From Puntanen et al. (2011, formula (10.5)) it follows that $\mathbf{G}_1\mathbf{y}$ is BLUE of an (estimable) parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$ if $\mathbf{G}_1$ satisfies the equation

$$\mathbf{G}_1(\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{VQ_X}) = (\mathbf{X}_1 : \mathbf{0} : \mathbf{0}). \tag{26}$$

On account of (4), which is a necessary and sufficient condition for $\mathbf{X}_1\boldsymbol{\beta}_1$ to be estimable, we arrive at $\mathcal{R}(\mathbf{X}_1) \cap \mathcal{R}(\mathbf{X}_2 : \mathbf{VQ_X}) = \{\mathbf{0}\}$. In consequence, we can utilize the representation of the Moore–Penrose inverse provided in Lemma 1.1, which leads to the conclusion that one of the solutions of (26) is of the form

$$\mathbf{G}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{Q}_{(\mathbf{X}_2:\mathbf{VQ_X})}\mathbf{X}_1)^\dagger\mathbf{X}_1'\mathbf{Q}_{(\mathbf{X}_2:\mathbf{VQ_X})}.$$

Hence,

$$\mathrm{BLUE}(\mathbf{X}_1\boldsymbol{\beta}_1) = \mathbf{X}_1(\mathbf{X}_1'\mathbf{Q}_{(\mathbf{X}_2:\mathbf{VQ_X})}\mathbf{X}_1)^\dagger\mathbf{X}_1'\mathbf{Q}_{(\mathbf{X}_2:\mathbf{VQ_X})}\mathbf{y}.$$

Similarly, by interchanging the subscripts "1" and "2" in (26), we obtain

$$\mathrm{BLUE}(\mathbf{X}_2\boldsymbol{\beta}_2) = \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{Q}_{(\mathbf{X}_1:\mathbf{VQ_X})}\mathbf{X}_2\right)^\dagger\mathbf{X}_2'\mathbf{Q}_{(\mathbf{X}_1:\mathbf{VQ_X})}\mathbf{y}.$$

The paper is concluded with some remarks concerned with advantages of utilizing the representation of the Moore–Penrose inverse provided in Lemma 1.1 from the computational point of view. In comparison to the methods of determining the inverse based on the singular value decomposition (SVD), which are exploited in several popular software packages (e.g., MATLAB, MATHEMATICA or R), an algorithm based on the representation (1) seems to have three main advantages, each leading to a dropping of computational costs. The first one is that it reduces sizes of matrices to be Moore–Penrose inverted—instead of the inverse of an $n \times m$ matrix $\mathbf{A}$, we need to compute two inverses of matrices $\mathbf{Q}_2\mathbf{A}_1$ and $\mathbf{Q}_1\mathbf{A}_2$ of orders $n \times m_1$ and $n \times m_2$, respectively, where $m_1 + m_2 = m$; as several software tools impose limits on sizes of matrices which can be stored, one can encounter a situation in which the inverse of $\mathbf{A}$ may exceed the limit, while the inverses of $\mathbf{Q}_2\mathbf{A}_1$ and $\mathbf{Q}_1\mathbf{A}_2$ are still manageable. The second benefit is that the algorithm allows computing both block entries occurring in the inverse (almost) simultaneously—in the light of Baksalary and Trenkler (2021, formula (6)), the two entries involved in the representation (1) are linked by the identity

$$\left(\mathbf{Q}_i\mathbf{A}_j\right)^{\dagger} = \mathbf{A}_j^{\dagger}\left[\mathbf{I}_n - \mathbf{A}_i\left(\mathbf{Q}_j\mathbf{A}_i\right)^{\dagger}\right], \ i, j = 1, 2, i \neq j,$$

which means that one of the Moore–Penrose inverses involved in the representation can be derived from the knowledge of the other. The third advantage of the algorithm is that it can be executed iteratively in each subsequent step to matrices of smaller order—from Baksalary and Trenkler (2021, Lemma 1) it follows that when $\mathbf{A}$ is of full column rank, then $\mathbf{Q}_2\mathbf{A}_1$ and $\mathbf{Q}_1\mathbf{A}_2$ are of full column ranks as well, which means that the inverses $(\mathbf{Q}_2\mathbf{A}_1)^{\dagger}$ and $(\mathbf{Q}_1\mathbf{A}_2)^{\dagger}$ can be computed by applying the same algorithm; the procedure might be carried out iteratively till the matrices to be inverted are all reduced to (row) vectors.

Appendix A: Moore–Penrose inverse of a matrix

Let $\mathbf{S}$ be an $n \times m$ matrix. Then there exists a unique matrix $\mathbf{S}^{\dagger}$ such that

$$\mathbf{S}\mathbf{S}^{\dagger}\mathbf{S} = \mathbf{S}, \ \mathbf{S}^{\dagger}\mathbf{S}\mathbf{S}^{\dagger} = \mathbf{S}^{\dagger}, \ (\mathbf{S}\mathbf{S}^{\dagger})' = \mathbf{S}\mathbf{S}^{\dagger}, \ (\mathbf{S}^{\dagger}\mathbf{S})' = \mathbf{S}^{\dagger}\mathbf{S}.$$

The matrix $\mathbf{S}^{\dagger}$ is called the Moore–Penrose inverse of $\mathbf{S}$.

It can be verified that

$$\mathbf{S}^{\dagger} = (\mathbf{S}'\mathbf{S})^{\dagger}\mathbf{S}', \tag{A1}$$

which takes the form $\mathbf{S}^{\dagger} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, when $\mathbf{S}$ is of full column rank. Another relevant property of the Moore–Penrose inverse is that it offers a handy way to represent orthogonal projectors in $\mathbb{R}^n$ (symmetric idempotent matrices of order n). To be precise, an $n \times n$ matrix $\mathbf{P}$ is an orthogonal projector if and only if it is expressible as $\mathbf{S}\mathbf{S}^{\dagger}$ for some $n \times m$ matrix $\mathbf{S}$. Then, $\mathbf{S}\mathbf{S}^{\dagger}$ is the orthogonal projector onto $\mathcal{R}(\mathbf{S})$ and, consequently, $\mathbf{I}_n - \mathbf{S}\mathbf{S}^{\dagger}$ is the orthogonal projector onto the orthogonal complement of $\mathcal{R}(\mathbf{S})$, which coincides with $\mathcal{N}(\mathbf{S}')$. Similarly, $\mathbf{S}^{\dagger}\mathbf{S}$ and $\mathbf{I}_m - \mathbf{S}^{\dagger}\mathbf{S}$ are the orthogonal projectors onto $\mathcal{R}(\mathbf{S}')$ and $\mathcal{N}(\mathbf{S})$, respectively, where $\mathcal{R}(\mathbf{S}') \overset{\perp}{\oplus} \mathcal{N}(\mathbf{S}) = \mathbb{R}^m$. An important feature is that there is a one-to-one correspondence between the orthogonal projector and the subspace onto which it projects.

References

Baksalary JK, Baksalary OM (2007) Particular formulae for the Moore–Penrose inverse of a columnwise partitioned matrix. Linear Algebra Appl 421:16–23

Baksalary OM, Trenkler G (2021) On formulae for the Moore–Penrose inverse of a columnwise partitioned matrix. Appl Math Comput 403:125913

Davidson R, MacKinnon JG (1993) Estimation and inference in econometrics. Oxford University Press, New York

Pedler PJ (1993) Fitting a circle to numerical data: an illustration of statistical modelling. Int J Math Educ Sci Technol 24:131–143

Puntanen S, Styan GPH, Isotalo J (2011) Matrix tricks for linear statistical models—our personal top twenty. Springer, Berlin

Rao CR, Mitra SK (1971) Generalized inverse of matrices and its applications. Wiley, New York