Ann Inst Stat Math (2016) 68:705–723
DOI 10.1007/s10463-015-0512-2

Best equivariant estimator of regression coefficients in a seemingly unrelated regression model with known correlation matrix

Hiroshi Kurata · Shun Matsuura

Received: 25 February 2014 / Revised: 22 January 2015 / Published online: 12 March 2015

Abstract This paper derives the best equivariant estimator (BEE) of the regression coefficients of a seemingly unrelated regression model with an elliptically symmetric error. Equivariance with respect to the group of location and scale transformations is considered. We assume that the correlation matrix of the error term is known. Since the correlation matrix is a maximal invariant parameter under the group action, the model treated in this paper is generated as exactly one orbit on the parameter space. It is also shown that the BEE can be viewed as a generalized least squares estimator.

Keywords Equivariant estimator · Seemingly unrelated regression model · Group invariance · Maximal invariant · Generalized least squares estimator

1 Introduction

A seemingly unrelated regression (SUR) model is defined to be a set of p different linear regression models with cross-correlation:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \quad \text{with} \quad \mathrm{E}[\boldsymbol{\varepsilon}_i] = \boldsymbol{0}, \quad \mathrm{V}[\boldsymbol{\varepsilon}_i] = \sigma_{ii} \boldsymbol{I}_m \quad (i = 1, \ldots, p) \tag{1}$$

H. Kurata (✉)
Graduate School of Arts and Sciences, The University of Tokyo,
3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
e-mail: kurata@waka.c.u-tokyo.ac.jp

S. Matsuura
Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku,
Yokohama 223-8522, Japan

and

$$\mathrm{E}\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\right] = \sigma_{ij}\boldsymbol{I}_m \quad (i, j = 1, \ldots, p),$$

where $\boldsymbol{y}_i : m \times 1$, $\boldsymbol{X}_i : m \times k_i$ is of rank k_i, $\boldsymbol{\beta}_i : k_i \times 1$ and $\boldsymbol{\varepsilon}_i : m \times 1$. The model was originally formulated by Zellner (1962, 1963), and has been playing an essential role especially in econometric and biometric data analyses. It can be expressed in a compact form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{with} \quad \mathrm{V}[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} \otimes I_m \tag{2}$$

by setting $n = mp$, $k = \sum_{i=1}^{p} k_i$, $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathcal{S}_p^+$,

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{X}_p \end{pmatrix} = \mathrm{diag}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p\} : n \times k,$$

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_p \end{pmatrix} : n \times 1, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_p \end{pmatrix} : k \times 1 \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_p \end{pmatrix} : n \times 1,$$

where $\mathcal{S}_p^+$ denotes the set of $p \times p$ positive definite matrices and $\otimes$ denotes the Kronecker product of matrices.

In this paper, we consider the problem of estimating the coefficient vector $\boldsymbol{\beta}$ with respect to the loss function

$$L\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \boldsymbol{X}' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right) \boldsymbol{X} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right). \tag{3}$$

Let $\mathcal{Y} = \mathbb{R}^n$ be the sample space of $\boldsymbol{y}$ and $\tilde{\Theta} = \mathbb{R}^k \times \mathcal{S}_p^+$ be the parameter space of $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The class of distributions of $\boldsymbol{y}$ is denoted by

$$\mathcal{P} = \left\{P_{\boldsymbol{\beta},\boldsymbol{\Sigma}} | (\boldsymbol{\beta}, \boldsymbol{\Sigma}) \in \tilde{\Theta}\right\}, \tag{4}$$

where $P_{\boldsymbol{\beta},\boldsymbol{\Sigma}}$ is a distribution on $\mathcal{Y}$ with density function (with respect to the Lebesgue measure) of the form

$$f(\boldsymbol{y}|(\boldsymbol{\beta}, \boldsymbol{\Sigma})) = |\boldsymbol{\Sigma} \otimes \boldsymbol{I}_m|^{-1/2}\, h\left((\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right) (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$$

for some $h : [0, \infty) \to [0, \infty)$. The function h does not depend on $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Clearly, $\mathcal{P}$ is a class of elliptically symmetric distributions of mean $\boldsymbol{X}\boldsymbol{\beta}$ and variance–covariance matrix $\boldsymbol{\Sigma} \otimes \boldsymbol{I}_m$. For fundamental properties of elliptically symmetric distributions, see, for example, Chapter 1 of Muirhead (1982).

We assume that the matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ is partially known. More specifically, the correlation coefficients

$$\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2} \ \ (1 \le i, j \le p) \tag{5}$$

are all known. In other words, $\boldsymbol{\Sigma}$ is supposed to be the product of an unknown diagonal matrix with positive diagonal elements and a known correlation matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Delta} \ \text{ with } \ \boldsymbol{\Lambda} = (\rho_{ij}) \ \text{ and } \ \boldsymbol{\Delta} = \text{diag}\left\{\sigma_{11}^{1/2}, \ldots, \sigma_{pp}^{1/2}\right\} : p \times p. \tag{6}$$

In most applications, $\boldsymbol{\Sigma}$ is fully unknown and the model is estimated by the method of generalized least squares (GLS). The method has been one of central topics in estimation of SUR models. Some of the development that have taken place so far are summarized, for example, in the books by Srivastava and Giles (1987) and Kariya and Kurata (2004). Papers by Kariya (1981), Bilodeau (1990), Kurata and Kariya (1996) and Kurata (1999) investigated the finite sample efficiency of some typical GLS estimators (GLSEs) under normal or elliptical distributions. Liu (2002) and Ma and Ye (2010) proposed new estimators based on the idea of covariance adjustment. Fang et al. (1997) proved the minimaxity of the ordinary least squares estimator (OLSE) under another loss function.

In the present model, a partial information on $\boldsymbol{\Sigma}$ is available, and hence we need to take a different approach to incorporate the information into the estimation procedure. Our approach here is based on the equivariance principle [e.g., Chapter 3 of Lehmann and Casella (1998) and Eaton (1989)]. To state it precisely, let $G_i = (0, \infty) \times \mathbb{R}^{k_i}$ $(i = 1, \ldots, p)$ and consider the group $G = G_1 \times \cdots \times G_p$. The group G acts on the spaces $\mathcal{Y}$ and $\tilde{\Theta}$ via the group action

$$\boldsymbol{y}_i \longrightarrow a_i \boldsymbol{y}_i + \boldsymbol{X}_i \boldsymbol{c}_i, \tag{7}$$

$$\boldsymbol{\beta}_i \longrightarrow a_i \boldsymbol{\beta}_i + \boldsymbol{c}_i, \ \sigma_{ij} \longrightarrow a_i a_j \sigma_{ij} \ \ (i, j = 1, \ldots, p), \tag{8}$$

respectively, where $(a_i, \boldsymbol{c}_i) \in G_i$ $(i = 1, \ldots, p)$. It is easy to see that this action leaves the class $\mathcal{P}$ invariant. An estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{y}) = \left(\hat{\boldsymbol{\beta}}_1(\boldsymbol{y})', \ldots, \hat{\boldsymbol{\beta}}_p(\boldsymbol{y})'\right)'$ of $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_p'\right)'$ is called equivariant under G, if $\hat{\boldsymbol{\beta}}(\boldsymbol{y})$ satisfies

$$\hat{\boldsymbol{\beta}}_i(a_1\boldsymbol{y}_1 + \boldsymbol{X}_1\boldsymbol{c}_1, \ldots, a_p\boldsymbol{y}_p + \boldsymbol{X}_p\boldsymbol{c}_p) = a_i\hat{\boldsymbol{\beta}}_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p) + \boldsymbol{c}_i \ \ (i = 1, \ldots, p) \tag{9}$$

for any $\boldsymbol{y} \in \mathcal{Y}$ and $(a_i, \boldsymbol{c}_i) \in G_i$ $(i = 1, \ldots, p)$. The action in (7) and (8) is expressed in matrix form as

$$\boldsymbol{y} \longrightarrow (\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{y} + \boldsymbol{X}\boldsymbol{c} \ \text{ and } \ (\boldsymbol{\beta}, \boldsymbol{\Sigma}) \longrightarrow (\tilde{\boldsymbol{A}}\boldsymbol{\beta} + \boldsymbol{c}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'), \tag{10}$$

where $\boldsymbol{A} = \operatorname{diag}\left\{a_1, \ldots, a_p\right\} : p \times p$, $\tilde{\boldsymbol{A}} = \operatorname{diag}\left\{a_1 \boldsymbol{I}_{k_1}, \ldots, a_p \boldsymbol{I}_{k_p}\right\}$ and $\boldsymbol{c} = \left(\boldsymbol{c}_1', \ldots, \boldsymbol{c}_p'\right)' : k \times 1$. And also, the condition (9) is rewritten as

$$\hat{\boldsymbol{\beta}}((\boldsymbol{A} \otimes \boldsymbol{I}_m)\boldsymbol{y} + \boldsymbol{X}\boldsymbol{c}) = \tilde{\boldsymbol{A}}\hat{\boldsymbol{\beta}}(\boldsymbol{y}) + \boldsymbol{c}. \tag{11}$$

If we write (7) and (8) (or 10) as

$$\boldsymbol{y} \to g\boldsymbol{y} \ \text{ and } \ (\boldsymbol{\beta}, \boldsymbol{\Sigma}) \to g(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \equiv (g\boldsymbol{\beta}, g\boldsymbol{\Sigma}) \ \text{ with } g \in G, \tag{12}$$

the condition (9) is expressed as

$$\hat{\boldsymbol{\beta}}(g\boldsymbol{y}) = g\hat{\boldsymbol{\beta}}(\boldsymbol{y}) \quad \text{for all } \boldsymbol{y} \in \mathcal{Y}, g \in G. \tag{13}$$

A statistic $U(\boldsymbol{y})$ is called invariant if it satisfies $U(g\boldsymbol{y}) = U(\boldsymbol{y}) \quad$ for all $g \in G$ and $\boldsymbol{y} \in \mathcal{Y}$. If an invariant statistic $U(\boldsymbol{y})$ satisfies the following condition:

$$U(\boldsymbol{y}) = U(\boldsymbol{y}_*) \Rightarrow \exists g \in G \text{ such that } \boldsymbol{y} = g\boldsymbol{y}_*, \tag{14}$$

then it is called a maximal invariant. A maximal invariant parameter is defined in exactly the same way: it is a function $\tau(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ on $\tilde{\Theta}$ that is invariant under the action of G and satisfies the condition described in (14) [with U and $\boldsymbol{y}$ replaced by τ and $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, respectively].

Let $\hat{\boldsymbol{\beta}}_*$ be an equivariant estimator. If the risk function of $\hat{\boldsymbol{\beta}}_*$ is uniformly smaller than or equal to that of any other equivariant estimator $\hat{\boldsymbol{\beta}}$, that is,

$$R\left(\hat{\boldsymbol{\beta}}_*, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) \leq R\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right)$$

holds uniformly for $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, then $\hat{\boldsymbol{\beta}}_*$ is called a best equivariant estimator (BEE) of $\boldsymbol{\beta}$, where $R\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) \equiv \mathrm{E}_{\boldsymbol{\beta},\boldsymbol{\Sigma}}\left[L\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right)\right]$ and $\mathrm{E}_{\boldsymbol{\beta},\boldsymbol{\Sigma}}$ denotes the expectation with respect to $P_{\boldsymbol{\beta},\boldsymbol{\Sigma}}$. Since the maximum likelihood estimator (MLE) is equivariant [see, for example, Theorem 3.2 of Eaton (1989)], the BEE is better than the MLE, unless they coincide.

It is important to view the assumption of known correlation coefficients from the group invariance theoretic standpoints. To do so, let $\mathcal{C}_p$ be the set of $p \times p$ positive definite matrices whose diagonal elements are all ones (i.e., the set of correlation matrices), and define

$$\tau : \tilde{\Theta} \longrightarrow \mathcal{C}_p : (\boldsymbol{\beta}, \boldsymbol{\Sigma}) \longrightarrow \boldsymbol{\Lambda} \ \text{ with } \ \boldsymbol{\Sigma} = \boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{\Delta},$$

where $\boldsymbol{\Sigma}$ is decomposed as (6). Then, as can be easily seen, the mapping $\tau(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \boldsymbol{\Lambda}$ is a maximal invariant parameter under G. Hence our problem can be viewed as that of equivariant estimation with a known maximal invariant parameter, where the parameter space $\tilde{\Theta}$ is reduced to

$$\Theta = \left\{(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \in \tilde{\Theta} | \tau(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \boldsymbol{\Lambda}\right\}$$

with known $\boldsymbol{\Lambda} = (\rho_{ij})$. This means that the space Θ consists of exactly one orbit that contains $(\mathbf{0}, \boldsymbol{\Lambda})$. A simple calculation shows that $\Theta = \{(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \in \tilde{\Theta} \mid \boldsymbol{\beta} \in \mathbb{R}^k, \boldsymbol{\Sigma} = (a_i a_j \rho_{ij}),\ a_1, \ldots, a_p > 0\}$.

In the literature, Kariya (1989) established a general theory of equivariant estimation with a known maximal invariant parameter, where an ancillary statistic appears as a maximal invariant statistic. The framework he gave enables us to treat various "curved" models (such as a model with a known variational coefficient) in a unified way. See, for example, Chapter 3 of Giri (1996). The present research is also descended from Kariya's pioneering work.

In general, the risk function of an equivariant estimator is constant on each orbit in the parameter space [e.g., Theorem 6.3 of Eaton (1989)] which means that it is a function of a maximal invariant parameter. Hence in our case, for each equivariant estimator $\hat{\boldsymbol{\beta}}$, the risk function $R\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right)$ is constant on Θ. Thus, when evaluating the risk function, we can let $\boldsymbol{\beta}$ be the zero vector and $\boldsymbol{\Sigma}$ be the (known) correlation matrix $\boldsymbol{\Lambda}$:

$$\begin{aligned} R\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) &= \mathrm{E}_{\boldsymbol{\beta}, \boldsymbol{\Sigma}}\left[L\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right)\right] = \mathrm{E}_{\mathbf{0}, \boldsymbol{\Lambda}}\left[L\left(\hat{\boldsymbol{\beta}}, (\mathbf{0}, \boldsymbol{\Lambda})\right)\right] \\ &= \mathrm{E}_{\mathbf{0}, \boldsymbol{\Lambda}}\left[\hat{\boldsymbol{\beta}}' \boldsymbol{X}' \left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right) \boldsymbol{X} \hat{\boldsymbol{\beta}}\right]. \end{aligned} \tag{15}$$

This paper is organized as follows. In the next section, we derive a maximal invariant statistic and describe the class of equivariant estimators. Section 3 is devoted to deriving a general expression of the BEE. The result is applied to a two-equation model with normal error. In Sect. 4, the BEE is examined through the theory of GLS estimation. A numerical example is also given.

2 Reduction by equivariance

In this section, a maximal invariant statistic is derived and a characterization of an equivariant estimator is given.

To do so, let

$$\boldsymbol{b}_i = \left(\boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i' \boldsymbol{y}_i \text{ and } \boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{b}_i = \left\{\boldsymbol{I}_m - \boldsymbol{X}_i \left(\boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i'\right\} \boldsymbol{y}_i$$

be the ordinary least squares estimator (OLSE) and the residual vector calculated equation-wise from the ith equation.

Lemma 1 *Let* $\boldsymbol{u} = U(\boldsymbol{y})$ *be defined as*

$$\boldsymbol{u} = U(\boldsymbol{y}) = \left(\boldsymbol{u}_1', \ldots, \boldsymbol{u}_p'\right)' : n \times 1, \tag{16}$$

where

$$\boldsymbol{u}_i = \begin{cases} \boldsymbol{e}_i/\|\boldsymbol{e}_i\| & (\text{if } \boldsymbol{e}_i \neq \boldsymbol{0}) \\ \boldsymbol{0} & (\text{if } \boldsymbol{e}_i = \boldsymbol{0}) \end{cases} \quad (i = 1, \ldots, p).$$

Then the statistic $\boldsymbol{u} = U(\boldsymbol{y})$ is a maximal invariant under G.

Proof First we show the invariance of U, that is, $U(g\boldsymbol{y}) = U(\boldsymbol{y})$ for any $g \in G$. The statistic $U(\boldsymbol{y})$ depends on $\boldsymbol{y}$ only through $\boldsymbol{e}_i$s, and the action $\boldsymbol{y} \to g\boldsymbol{y}$ in (7) induces the one

$$\boldsymbol{e}_i \longrightarrow a_i \boldsymbol{e}_i \quad (i = 1, \ldots, p)$$

on the spaces of $\boldsymbol{e}_i$s. From this, the quantities $\boldsymbol{u}_i$s are invariant, and hence $U(\boldsymbol{y})$ is invariant.

To see that it also satisfies (14), suppose $\boldsymbol{u} = U(\boldsymbol{y}) = U(\boldsymbol{y}^*) = \boldsymbol{u}^*$ for some $\boldsymbol{y}, \boldsymbol{y}^* \in \mathcal{Y}$. Then for each $i = 1, \ldots, p$, either of the following two equalities holds:

$$\boldsymbol{e}_i/\|\boldsymbol{e}_i\| = \boldsymbol{e}_i^*/\|\boldsymbol{e}_i^*\| \text{ or } \boldsymbol{e}_i = \boldsymbol{e}_i^* = \boldsymbol{0},$$

where $\boldsymbol{e}_i^* = \left\{\boldsymbol{I}_m - \boldsymbol{X}_i\left(\boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i'\right\}\boldsymbol{y}_i^*$ and $\boldsymbol{y}_i^*$ is the ith subvector of $\boldsymbol{y}^*$. For the former case, we have $\boldsymbol{e}_i = a_i\boldsymbol{e}_i^*$ with $a_i = \|\boldsymbol{e}_i\|/\|\boldsymbol{e}_i^*\|$, which in turn implies that $\boldsymbol{y}_i = a_i\boldsymbol{y}_i^* + \boldsymbol{X}_i\boldsymbol{c}_i$ for some $\boldsymbol{c}_i \in \mathbb{R}^{k_i}$. For the latter case, both $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{c}_i$ and $\boldsymbol{y}_i^* = \boldsymbol{X}_i\boldsymbol{c}_i^*$ hold for some $\boldsymbol{c}_i, \boldsymbol{c}_i^* \in \mathbb{R}^{k_i}$. Hence we have $\boldsymbol{y}_i = \boldsymbol{y}_i^* + \boldsymbol{X}_i(\boldsymbol{c}_i - \boldsymbol{c}_i^*)$. Thus, in both cases, we see the existence of $g \in G$ such that $\boldsymbol{y} = g\boldsymbol{y}^*$. Hence $\boldsymbol{u} = U(\boldsymbol{y})$ is a maximal invariant. □

Since the sets $\mathcal{N}_i = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{e}_i = \boldsymbol{0}\}$ are of Lebesgue measure zero, we have

$$\boldsymbol{u}_1 = \boldsymbol{e}_1/\|\boldsymbol{e}_1\|, \ldots, \boldsymbol{u}_p = \boldsymbol{e}_p/\|\boldsymbol{e}_p\| \text{ a.e.}$$

Hence, in what follows, we let the sample space be

$$\mathcal{Y} = \mathbb{R}^n - \mathcal{N} \text{ with } \mathcal{N} = \cup_{i=1}^p \mathcal{N}_i \tag{17}$$

without essential loss of generality, and correspondingly let the range of U be

$$\mathcal{U} = \{\boldsymbol{u} = U(\boldsymbol{y}) | \boldsymbol{y} \in \mathcal{Y}\}.$$

Next we derive a characterization of an equivariant estimator. Let

$$\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$$

be the OLSE calculated from the whole model, which is a collection of the equation-wise OLSEs, that is $\boldsymbol{b} = (\boldsymbol{b}_1', \ldots, \boldsymbol{b}_p')'$. And also, let

$$\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{e}) = \mathrm{diag}\{\|\boldsymbol{e}_1\|\boldsymbol{I}_{k_1}, \ldots, \|\boldsymbol{e}_p\|\boldsymbol{I}_{k_p}\} : k \times k. \tag{18}$$

Lemma 2 *An estimator* $\hat{\boldsymbol{\beta}} : \mathcal{Y} \to \mathbb{R}^k$ *is an equivariant estimator under G if and only if there exists a measurable function* $\boldsymbol{d} : \mathcal{U} \to \mathbb{R}^k$ *such that*

$$\hat{\boldsymbol{\beta}}(\boldsymbol{y}) = \boldsymbol{b} + \boldsymbol{R}(\boldsymbol{e})\boldsymbol{d}(\boldsymbol{u}). \tag{19}$$

Proof Suppose that $\hat{\boldsymbol{\beta}}(\boldsymbol{y})$ is equivariant. Then (11) holds. Substitute $(a_i, \boldsymbol{c}_i) = \left(\|\boldsymbol{e}_i\|^{-1}, -\|\boldsymbol{e}_i\|^{-1}\boldsymbol{b}_i\right) \in G_i$ into (9). Then we have

$$\hat{\boldsymbol{\beta}}_i\left((\boldsymbol{y}_1 - \boldsymbol{X}_1\boldsymbol{b}_1)/\|\boldsymbol{e}_1\|, \ldots, (\boldsymbol{y}_p - \boldsymbol{X}_p\boldsymbol{b}_p)/\|\boldsymbol{e}_p\|\right) = \|\boldsymbol{e}_i\|^{-1}\hat{\boldsymbol{\beta}}_i(\boldsymbol{y}) - \|\boldsymbol{e}_i\|^{-1}\boldsymbol{b}_i$$
$$(i = 1, \ldots, p).$$

Hence we see that $\hat{\boldsymbol{\beta}}_i(\boldsymbol{y})$ must be of the form

$$\hat{\boldsymbol{\beta}}_i(\boldsymbol{y}) = \boldsymbol{b}_i + \|\boldsymbol{e}_i\|\, \boldsymbol{d}_i(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p) \text{ for some } \boldsymbol{d}_i : \mathcal{U} \to \mathbb{R}^{k_i} \quad (i = 1, \ldots, p) \tag{20}$$

on $\mathcal{Y}$. Rewriting this in matrix form yields (19).

Conversely, an estimator $\hat{\boldsymbol{\beta}}$ of the above form clearly satisfies (9), which means that the equality (19) gives a characterization of equivariant estimator of $\boldsymbol{\beta}$. □

If we set $\boldsymbol{d}(\boldsymbol{u}) = \boldsymbol{0}$, then $\hat{\boldsymbol{\beta}}(\boldsymbol{y}) = \boldsymbol{b}$, which means that the class of equivariant estimators contains the OLSE. The BEE derived below dominates $\boldsymbol{b}$ unless they coincide.

3 Best equivariant estimator

This section is devoted to deriving a BEE of $\boldsymbol{\beta}$. To do so, let

$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{u}) = (t_{ij}) : p \times p \text{ with } t_{ij} = t_{ij}(\boldsymbol{u}) = \mathrm{E}_{\boldsymbol{0},\boldsymbol{\Lambda}}[\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\| \mid \boldsymbol{u}] \tag{21}$$

be the conditional expectation of $\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\|$ given $\boldsymbol{u}$, which is an ancillary statistic since the maximal invariant parameter is known. Throughout this section, all expectations are taken under $P_{\boldsymbol{0},\boldsymbol{\Lambda}}$, and hence we omit the suffix and write just $\mathrm{E}[\cdot]$.

Also let

$$\boldsymbol{H} = \boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X} \in \mathcal{S}_k^+,$$
$$\boldsymbol{S} = \boldsymbol{H}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left\{\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\right\}$$
$$= \boldsymbol{H}\left\{\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \left(\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\right\} : k \times n, \tag{22}$$

and decompose $\boldsymbol{S}$ into p^2 blocks:

$$\boldsymbol{S} = [\boldsymbol{S}_{ij}] = \begin{bmatrix} \boldsymbol{S}_{11} & \ldots & \boldsymbol{S}_{1p} \\ \vdots & & \vdots \\ \boldsymbol{S}_{p1} & \ldots & \boldsymbol{S}_{pp} \end{bmatrix} \text{ with } \boldsymbol{S}_{ij} : k_i \times m.$$

For matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ of the same size, $\boldsymbol{A} \circ \boldsymbol{B}$ denotes the Hadamard (element-wise) product: $\boldsymbol{A} \circ \boldsymbol{B} = (a_{ij} b_{ij})$.

Theorem 3 *Suppose that the distribution of* $\boldsymbol{y}$ *is in the class* $\mathcal{P}$. *If the correlation matrix* $\boldsymbol{\Lambda}$ *is known, then the estimator*

$$\hat{\boldsymbol{\beta}}_*(\boldsymbol{y}) = \boldsymbol{b} + \boldsymbol{R}(\boldsymbol{e})\boldsymbol{d}_*(\boldsymbol{u}) \;\; with \;\; \boldsymbol{d}_*(\boldsymbol{u}) = -\boldsymbol{D}(\boldsymbol{u})^{-1}\boldsymbol{F}(\boldsymbol{u})\boldsymbol{u} \tag{23}$$

is a BEE of $\boldsymbol{\beta}$, *where*

$$\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{u}) = \boldsymbol{X}'(\boldsymbol{\Lambda}^{-1} \circ \boldsymbol{T}(\boldsymbol{u}) \otimes \boldsymbol{I}_m)\boldsymbol{X} : k \times k \tag{24}$$

and

$$\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{u}) = \begin{bmatrix} t_{11}(\boldsymbol{u})\boldsymbol{S}_{11} & \ldots & t_{1p}(\boldsymbol{u})\boldsymbol{S}_{1p} \\ \vdots & & \vdots \\ t_{p1}(\boldsymbol{u})\boldsymbol{S}_{p1} & \ldots & t_{pp}(\boldsymbol{u})\boldsymbol{S}_{pp} \end{bmatrix} : k \times n. \tag{25}$$

The BEE is unique on $\mathcal{Y}$ *(and hence unique a.e. on* $\mathbb{R}^n$*).*

Proof Let $\hat{\boldsymbol{\beta}}(\boldsymbol{y}) = \boldsymbol{b} + \boldsymbol{R}(\boldsymbol{e})\boldsymbol{d}(\boldsymbol{u})$ be an equivariant estimator. Substituting it for $\hat{\boldsymbol{\beta}}$ in (15) yields

$$R\left(\hat{\boldsymbol{\beta}}, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) = \mathrm{E}_{\boldsymbol{0},\boldsymbol{\Lambda}}[(\boldsymbol{b} + \boldsymbol{R}\boldsymbol{d})'\boldsymbol{H}(\boldsymbol{b} + \boldsymbol{R}\boldsymbol{d})] \;\; \text{with} \;\; \boldsymbol{H} = \boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}, \tag{26}$$

where $\boldsymbol{d} = \boldsymbol{d}(\boldsymbol{u})$. Since the distribution of $\boldsymbol{y}$ is elliptically symmetric, the conditional mean of $\boldsymbol{b}$ given $\boldsymbol{e}$ is obtained as

$$\begin{aligned} \mathrm{E}[\boldsymbol{b}|\boldsymbol{e}] &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left\{\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\right\}\boldsymbol{e} \\ &\equiv \boldsymbol{B}\boldsymbol{e} \;\; (\text{say}), \end{aligned} \tag{27}$$

which can be easily shown by using, for example, the matrix identity (36). Then, by replacing $\boldsymbol{b} + \boldsymbol{R}\boldsymbol{d}$ in (26) with $(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{e}) + (\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})$ and conditioning on $\boldsymbol{e}$, we obtain

$$\begin{aligned} \mathrm{E}\left[(\boldsymbol{b} + \boldsymbol{R}\boldsymbol{d})'\boldsymbol{H}(\boldsymbol{b} + \boldsymbol{R}\boldsymbol{d})\right] = \mathrm{E}\left[(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{e})'\boldsymbol{H}(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{e})\right] \\ + \mathrm{E}\left[(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})'\boldsymbol{H}(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})\right], \end{aligned}$$

where the cross-terms vanish since

$$\begin{aligned} \mathrm{E}\left[(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{e})'\boldsymbol{H}(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})\right] &= \mathrm{E}\left[\mathrm{E}[(\boldsymbol{b} - \boldsymbol{B}\boldsymbol{e})'\boldsymbol{H}(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})|\boldsymbol{e}]\right] \\ &= \mathrm{E}\left[(\mathrm{E}[\boldsymbol{b}|\boldsymbol{e}] - \boldsymbol{B}\boldsymbol{e})'\boldsymbol{H}(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})\right] \\ &= \boldsymbol{0}. \end{aligned}$$

Since the first term of the right hand side of the above equality does not depend on $\boldsymbol{d}$, it suffices to minimize the second term. Decompose $\boldsymbol{d} : k \times 1$ as $\boldsymbol{d} = (\boldsymbol{d}'_1, \ldots, \boldsymbol{d}'_p)'$ with $\boldsymbol{d}_i : k_i \times 1$. Since the (i, j)th block of $\boldsymbol{H}$ in (22) is of the form $\rho^{ij} \boldsymbol{X}'_i \boldsymbol{X}_j : k_i \times k_j$ with $\boldsymbol{\Lambda}^{-1} = (\rho^{ij})$, we obtain

$$\begin{aligned}\mathrm{E}[\boldsymbol{d}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{d}] &= \sum_{i,j=1}^{p} \mathrm{E}\left[\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\|\rho^{ij}\boldsymbol{d}'_i\boldsymbol{X}'_i\boldsymbol{X}_j\boldsymbol{d}_j\right] = \sum_{i,j=1}^{p} \mathrm{E}\left[t_{ij}(\boldsymbol{u})\rho^{ij}\boldsymbol{d}'_i\boldsymbol{X}'_i\boldsymbol{X}_j\boldsymbol{d}_j\right] \\ &= \mathrm{E}\left[\boldsymbol{d}'\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \circ \boldsymbol{T} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\boldsymbol{d}\right] = \mathrm{E}[\boldsymbol{d}'\boldsymbol{D}\boldsymbol{d}], \end{aligned} \tag{28}$$

where the second equality is obtained by taking conditional expectation given $\boldsymbol{u}$. Next consider

$$\mathrm{E}[\boldsymbol{d}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{B}\boldsymbol{e}] = \mathrm{E}[\boldsymbol{d}'\boldsymbol{R}'\boldsymbol{S}\boldsymbol{e}] = \sum_{i,j=1}^{p} \mathrm{E}\left[\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\|\boldsymbol{d}'_i\boldsymbol{S}_{ij}\boldsymbol{u}_j\right],$$

where the first equality is due to $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{B}$ and the second follows from $\boldsymbol{e}_j = \|\boldsymbol{e}_j\|\boldsymbol{u}_j$. By taking conditional expectation given $\boldsymbol{u}$, we see that the above quantity is equal to

$$\sum_{i,j=1}^{p} \mathrm{E}\left[t_{ij}(\boldsymbol{u})\boldsymbol{d}'_i\boldsymbol{S}_{ij}\boldsymbol{u}_j\right] = \mathrm{E}[\boldsymbol{d}'\boldsymbol{F}\boldsymbol{u}]. \tag{29}$$

Hence combining (28) and (29) yields

$$\begin{aligned}&\mathrm{E}\left[(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})'\boldsymbol{H}(\boldsymbol{B}\boldsymbol{e} + \boldsymbol{R}\boldsymbol{d})\right] \\ &\quad = \mathrm{E}\left[\boldsymbol{d}'\boldsymbol{D}\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{F}\boldsymbol{u} + \boldsymbol{u}'\boldsymbol{F}'\boldsymbol{d} + \boldsymbol{e}'\boldsymbol{B}'\boldsymbol{H}\boldsymbol{B}\boldsymbol{e}\right] \\ &\quad = \mathrm{E}\left[\left(\boldsymbol{d} + \boldsymbol{D}^{-1}\boldsymbol{F}\boldsymbol{u}\right)'\boldsymbol{D}\left(\boldsymbol{d} + \boldsymbol{D}^{-1}\boldsymbol{F}\boldsymbol{u}\right)\right] + \mathrm{E}\left[\boldsymbol{e}'\boldsymbol{B}'\boldsymbol{H}\boldsymbol{B}\boldsymbol{e} - \boldsymbol{u}'\boldsymbol{F}'\boldsymbol{D}^{-1}\boldsymbol{F}\boldsymbol{u}\right],\end{aligned}$$

and this quantity is minimized uniquely when $\boldsymbol{d} = -\boldsymbol{D}^{-1}\boldsymbol{F}\boldsymbol{u}$ as long as $\boldsymbol{D}$ is nonsingular.

The nonsingularity of $\boldsymbol{D}$ follows from a well-known matrix result: Theorem 5.2.1 of Horn and Johnson (1991) states that if $\boldsymbol{X}$ is a positive definite matrix and $\boldsymbol{Y}$ is a positive semidefinite matrix whose diagonal elements are all positive, then the Hadamard product $\boldsymbol{X} \circ \boldsymbol{Y}$ is positive definite. To apply this result, let $\boldsymbol{r} = (\|\boldsymbol{e}_1\|, \ldots, \|\boldsymbol{e}_p\|)' : p \times 1$. The matrix $\boldsymbol{T}(\boldsymbol{u}) = (t_{ij}(\boldsymbol{u}))$ is the second moment of the conditional distribution of $\boldsymbol{r}$ given $\boldsymbol{u}$:

$$\boldsymbol{T}(\boldsymbol{u}) = \mathrm{E}[\boldsymbol{r}\boldsymbol{r}'|\boldsymbol{u}],$$

which is positive semidefinite. Since the vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p$ are not null on the sample space $\mathcal{Y}$, the quantities $\|\boldsymbol{e}_i\|$s are positive on $\mathcal{Y}$, which implies that $t_{ii}(\boldsymbol{u}) = \mathrm{E}[\|\boldsymbol{e}_i\|^2|\boldsymbol{u}] > 0$ $(i = 1, \ldots, p)$. Thus viewing $\boldsymbol{X} = \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{Y} = \boldsymbol{T}$, we see that

$\boldsymbol{\Lambda}^{-1} \circ \boldsymbol{T}$ is positive definite. This implies that $\boldsymbol{D}$ in (24) is positive definite. This completes the proof. □

Corollary 4 *When $\boldsymbol{\Lambda} = \boldsymbol{I}_p$, that is, the cross-correlations are all zero, the BEE reduces to the OLSE $\boldsymbol{b}$.*

Proof In this case, the matrix $\boldsymbol{S}$ in (22) is zero:

$$\boldsymbol{S} = \boldsymbol{H}\left\{\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right\} = \boldsymbol{0} : k \times n,$$

which implies $\boldsymbol{F} = \boldsymbol{0}$ and hence $\boldsymbol{d}_*(\boldsymbol{u}) = \boldsymbol{0}$ in (23). □

In applications, one needs to evaluate $\boldsymbol{T} = (t_{ij})$ in $\hat{\boldsymbol{\beta}}_*(\boldsymbol{y})$, which depends on the specification of h in (4). The result below gives an explicit form of $\hat{\boldsymbol{\beta}}_*(\boldsymbol{y})$ when the model consists of two equations ($p = 2$) and the distribution of the error term $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2')'$ is the normal distribution $N_n(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_m)$.

In the theorem below, we need the following notation:

$$\begin{aligned} W &= \boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m - \left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\left\{\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\left(\boldsymbol{\Lambda}^{-1} \otimes \boldsymbol{I}_m\right) \\ &\equiv \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \text{ with } W_{ij} : m \times m, \end{aligned}$$

and

$$v_{ij} = \frac{\boldsymbol{e}_i' W_{ij} \boldsymbol{e}_j}{\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\|} \quad (i, j = 1, 2). \tag{30}$$

Theorem 5 *Let $p = 2$. Suppose that $\boldsymbol{\varepsilon}$ is distributed as the normal distribution $N_n(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_m)$. Then the BEE $\hat{\boldsymbol{\beta}}_*(\boldsymbol{y})$ of $\boldsymbol{\beta}$ is given by (23) with*

$$\boldsymbol{T} = (t_{ij}(\boldsymbol{u})) = \frac{1}{K(0,0)}\begin{pmatrix} K(2,0) & K(1,1) \\ K(1,1) & K(0,2) \end{pmatrix}, \tag{31}$$

where

$$\begin{aligned} K(a,b) &= \frac{2^{(a+b+n-k)/2-2}}{v_{11}^{(a+m-k_1)/2} v_{22}^{(b+m-k_2)/2}} \sum_{\ell=0}^{\infty} \frac{1}{\ell!}\left(-\frac{2v_{12}}{\sqrt{v_{11}v_{22}}}\right)^{\ell} \Gamma\left(\frac{a+m-k_1+\ell}{2}\right) \\ &\quad \times \Gamma\left(\frac{b+m-k_2+\ell}{2}\right). \end{aligned} \tag{32}$$

Proof We can set without loss of generality $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}$. Choose arbitrarily an $m \times (m - k_i)$ matrix $\boldsymbol{Z}_i$ satisfying $\boldsymbol{Z}_i'\boldsymbol{X}_i = \boldsymbol{0}$ and $\boldsymbol{Z}_i'\boldsymbol{Z}_i = \boldsymbol{I}_{m-k_i}$ $(i = 1, 2)$, and define

$$\boldsymbol{Z} = \operatorname{diag}\{\boldsymbol{Z}_1, \boldsymbol{Z}_2\} : n \times (n - k),$$

where $n = 2m$, $k = k_1 + k_2$. Further, let $z_i = \boldsymbol{Z}_i' \boldsymbol{y}_i : (m - k_i) \times 1$ $(i = 1, 2)$ and

$$z = \boldsymbol{Z}' \boldsymbol{y} = (z_1', z_2')' : (n - k) \times 1, \tag{33}$$

which has a pdf and is distributed as $N_{n-k}(\boldsymbol{0}, \boldsymbol{Z}'(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m)\boldsymbol{Z})$. The quantities z and $\boldsymbol{e}$ are in one-to-one correspondence through

$$z_i = \boldsymbol{Z}_i' \boldsymbol{e}_i \;\text{ and }\; \boldsymbol{e}_i = \boldsymbol{Z}_i z_i. \tag{34}$$

Similarly, $\boldsymbol{u}_i = \boldsymbol{e}_i / \|\boldsymbol{e}_i\|$ and $z_i/\|z_i\|$ correspond in one-to-one, since $\|\boldsymbol{e}_i\| = \|z_i\|$. Let

$$\boldsymbol{\Omega} = \left\{\boldsymbol{Z}' \left(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_m\right) \boldsymbol{Z}\right\}^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \text{ with } \boldsymbol{\Omega}_{ij} : (m - k_i) \times (m - k_j).$$

Then it can be shown that v_{ij}s in (30) are expressed as

$$v_{ij} = \frac{z_i' \boldsymbol{\Omega}_{ij} z_j}{\|z_i\| \|z_j\|} \quad (i, j = 1, 2). \tag{35}$$

In fact, by using (34), we have

$$\frac{z_i' \boldsymbol{\Omega}_{ij} z_j}{\|z_i\| \|z_j\|} = \frac{\boldsymbol{e}_i' \boldsymbol{Z}_i \boldsymbol{\Omega}_{ij} \boldsymbol{Z}_j' \boldsymbol{e}_j}{\|\boldsymbol{e}_i\| \|\boldsymbol{e}_j\|}.$$

Here, the matrix $\boldsymbol{Z}_i \boldsymbol{\Omega}_{ij} \boldsymbol{Z}_j' : m \times m$ is the (i, j)th block of $\boldsymbol{Z} \boldsymbol{\Omega} \boldsymbol{Z}' = \boldsymbol{Z} \left\{\boldsymbol{Z}'(\boldsymbol{\Lambda} \otimes \boldsymbol{I})\boldsymbol{Z}\right\}^{-1} \boldsymbol{Z}'$. This matrix is in turn equal to $\boldsymbol{W}$ in (30), since, in general, the following matrix identity holds for any $n \times n$ positive definite matrix Ψ:

$$\Psi^{-1} = \Psi^{-1} \boldsymbol{X} (\boldsymbol{X}' \Psi^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \Psi^{-1} + \boldsymbol{Z} (\boldsymbol{Z}' \Psi \boldsymbol{Z})^{-1} \boldsymbol{Z}'. \tag{36}$$

In the sequel, we work on z in (33). Transform z_i into $(r_i, \boldsymbol{\theta}_i)$ $(i = 1, 2)$ through the polar coordinate decomposition: $r_i = \|z_i\|$ and

$$z_i = r_i \begin{pmatrix} \cos\theta_{i1} \\ \sin\theta_{i1} \cos\theta_{i2} \\ \vdots \\ \sin\theta_{i1} \sin\theta_{i2} \ldots \sin\theta_{i,m-k_i-2} \cos\theta_{i,m-k_i-1} \\ \sin\theta_{i1} \sin\theta_{i2} \ldots \sin\theta_{i,m-k_i-2} \sin\theta_{i,m-k_i-1} \end{pmatrix} = r_i p_i(\boldsymbol{\theta}_i) \text{ (say)},$$

where $r_i > 0$, $0 < \theta_{i1}, \ldots, \theta_{i,m-k_i-2} \leq \pi$, $0 < \theta_{i,m-k_i-1} \leq 2\pi$. Then, as is well-known [see, for example, Theorem 2.1.3 of Muirhead (1982)],

$$dz_1 dz_2 = J_1(\boldsymbol{\theta}_1) J_2(\boldsymbol{\theta}_2) r_1^{m-k_1-1} r_2^{m-k_2-1} dr_1 dr_2 d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2$$

with $J_i(\boldsymbol{\theta}_i) = \sin^{m-k_i-2}\theta_{i1}\sin^{m-k_i-3}\theta_{i2}\ldots\sin\theta_{i,m-k_i-2}$ $(i = 1, 2)$. From this, the joint pdf of $(r_1, r_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is obtained as

$$\begin{aligned}
&\tilde{g}(r_1, r_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= c r_1^{m-k_1-1} r_2^{m-k_2-1} \exp\left(-\frac{1}{2}(r_1 p_1(\boldsymbol{\theta}_1)', r_2 p_2(\boldsymbol{\theta}_2)') \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \begin{pmatrix} r_1 p_1(\boldsymbol{\theta}_1) \\ r_2 p_2(\boldsymbol{\theta}_2) \end{pmatrix}\right) \\
&\quad \text{with } c = (2\pi)^{-(n-k)/2} |\boldsymbol{\Omega}|^{1/2} J_1(\boldsymbol{\theta}_1) J_2(\boldsymbol{\theta}_2).
\end{aligned} \tag{37}$$

The quadratic form in the exponent is expressed as

$$\begin{aligned}
&r_1^2 p_1(\boldsymbol{\theta}_1)' \boldsymbol{\Omega}_{11} p_1(\boldsymbol{\theta}_1) + 2 r_1 r_2 p_1(\boldsymbol{\theta}_1)' \boldsymbol{\Omega}_{12} p_2(\boldsymbol{\theta}_2) + r_2^2 p_2(\boldsymbol{\theta}_2)' \boldsymbol{\Omega}_{22} p_2(\boldsymbol{\theta}_2) \\
&= r_1^2 v_{11} + 2 r_1 r_2 v_{12} + r_2^2 v_{22} \quad (\text{since } p_i(\boldsymbol{\theta}_i) = z_i / \|z_i\|) \\
&= \boldsymbol{r}' V \boldsymbol{r} \quad \text{with } \boldsymbol{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}.
\end{aligned} \tag{38}$$

Expand the factor $\exp(-r_1 r_2 v_{12})$ as

$$\exp(-r_1 r_2 v_{12}) = \sum_{\ell=0}^{\infty} r_1^{\ell} r_2^{\ell} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!},$$

which converges for all values of r_1, r_2 and v_{12}. Hence

$$\begin{aligned}
&\tilde{g}(r_1, r_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= c r_1^{m-k_1-1} \exp(-r_1^2 v_{11}^2/2) r_2^{m-k_2-1} \exp(-r_2^2 v_{22}/2) \sum_{\ell=0}^{\infty} r_1^{\ell} r_2^{\ell} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!} \\
&= c \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!}\, r_1^{m-k_1+\ell-1} \exp(-r_1^2 v_{11}^2/2)\, r_2^{m-k_2+\ell-1} \exp(-r_2^2 v_{22}/2).
\end{aligned}$$

Thus for each fixed $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and for $a, b = 0, 1, 2$, we have

$$\begin{aligned}
&\int_0^{\infty}\int_0^{\infty} r_1^a r_2^b\, \tilde{g}(r_1, r_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) dr_1 dr_2 \\
&= c \int_0^{\infty}\int_0^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!} r_1^{a+m-k_1+\ell-1} r_2^{b+m-k_2+\ell-1} \\
&\quad \times \exp(-r_1^2 v_{11}^2/2) \exp(-r_2^2 v_{22}/2) dr_1 dr_2 \\
&= c \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!} \int_0^{\infty}\int_0^{\infty} r_1^{a+m-k_1+\ell-1} r_2^{b+m-k_2+\ell-1} \\
&\quad \times \exp(-r_1^2 v_{11}^2/2) \exp(-r_2^2 v_{22}/2) dr_1 dr_2 \\
&= c \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell} v_{12}^{\ell}}{\ell!} \frac{2^{(a+m-k_1+\ell)/2-1}}{v_{11}^{(a+m-k_1+\ell)/2}} \frac{2^{(b+m-k_2+\ell)/2-1}}{v_{22}^{(b+m-k_2+\ell)/2}} \Gamma\left(\frac{a+m-k_1+\ell}{2}\right)
\end{aligned} \tag{39}$$

$$\times\Gamma\left(\frac{b+m-k_2+\ell}{2}\right) \tag{40}$$
$$= c\frac{2^{(a+b+n-k)/2-2}}{v_{11}^{(a+m-k_1)/2}v_{22}^{(b+m-k_2)/2}}\sum_{\ell=0}^{\infty}\frac{1}{\ell!}\left(-\frac{2v_{12}}{\sqrt{v_{11}v_{22}}}\right)^{\ell}\Gamma\left(\frac{a+m-k_1+\ell}{2}\right)$$
$$\times\Gamma\left(\frac{b+m-k_2+\ell}{2}\right)$$
$$= cK(a,b).$$

Let us confirm that the above term-by-term integration is valid. Define a power series

$$P(x)=\sum_{\ell=0}^{\infty}\frac{x^{\ell}}{\ell!}r_1^{a+m-k_1+\ell-1}r_2^{b+m-k_2+\ell-1}\exp(-r_1^2v_{11}^2/2)\exp(-r_2^2v_{22}/2).$$

Then clearly the series $P(-v_{12})$ is the integrand in (39). Since $P(-v_{12})\le P(|v_{12}|)$ holds, we can view $P(|v_{12}|)$ as a dominating function. The function $P(|v_{12}|)$ is a nonnegative series and hence we can integrate it term by term (due to the monotone convergence theorem). The resulting series is $Q(2|v_{12}/(v_{11}v_{22})^{1/2}|)$, where

$$Q(x)=\sum_{\ell=0}^{\infty}\frac{x^{\ell}}{\ell!}\Gamma\left(\frac{a+m-k_1+\ell}{2}\right)\Gamma\left(\frac{b+m-k_2+\ell}{2}\right).$$

We have to show that $Q\left(2\left|v_{12}/(v_{11}v_{22})^{1/2}\right|\right)$ converges. A routine calculation shows that the convergence radius of $Q(x)$ is 2. Hence it suffices to see that $\left|v_{12}/(v_{11}v_{22})^{1/2}\right|<1$. Since V in (38) can be written as

$$V=\boldsymbol{U}'\boldsymbol{\Omega}\boldsymbol{U}\ \ \text{with}\ \ \boldsymbol{U}=\text{diag}\{z_1/\|z_1\|,z_2/\|z_2\|\}:(n-k)\times 2$$

and $\boldsymbol{U}$ is of full rank on $\mathcal{Y}$, the matrix V is positive definite. This guarantees $|v_{12}/(v_{11}v_{22})^{1/2}|<1$ on $\mathcal{Y}$, and thus the series $Q(2|v_{12}/(v_{11}v_{22})^{1/2}|)$ converges. Hence by the dominating convergence theorem, the validity of the term-by-term integration is verified.

Since there exists a one-to-one correspondence between $\boldsymbol{u}=(\boldsymbol{u}_1',\boldsymbol{u}_2')'$ and $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, we have

$$t_{12}(\boldsymbol{u})=\text{E}[\|\boldsymbol{e}_1\|\|\boldsymbol{e}_2\||\boldsymbol{u}]=\text{E}[r_1r_2|\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]$$
$$=\frac{\int_0^{\infty}\int_0^{\infty}r_1r_2\tilde{g}(r_1,r_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)dr_1dr_2}{\int_0^{\infty}\int_0^{\infty}\tilde{g}(r_1,r_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)dr_1dr_2}=K(1,1)/K(0,0),$$

where in the second equality, $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2$ are chosen for given $\boldsymbol{u}$. Similarly, $t_{11}(\boldsymbol{u})=\text{E}[\|\boldsymbol{e}_1\|^2|\boldsymbol{u}]=K(2,0)/K(0,0)$ and $t_{22}(\boldsymbol{u})=\text{E}[\|\boldsymbol{e}_2\|^2|\boldsymbol{u}]=K(0,2)/K(0,0)$. Thus we obtain (31). This completes the proof. □

4 BEE as a generalized least squares estimator

In this section, the BEE $\hat{\boldsymbol{\beta}}_*(\boldsymbol{y})$ is examined through the theory of GLS estimation.

Consider the SUR model (2). As is well known, when the matrix $\boldsymbol{\Sigma}$ is known (up to a multiplicative constant), the Gauss–Markov estimator (GME)

$$b(\boldsymbol{\Sigma}) = \left(\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{y}$$

is the best linear unbiased estimator of $\boldsymbol{\beta}$. That is, for any linear unbiased estimator $\hat{\boldsymbol{\beta}}$, the variance–covariance matrix of the GME is uniformly smaller than or equal to that of $\hat{\boldsymbol{\beta}}$:

$$\left(\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1} = \mathrm{V}[b(\boldsymbol{\Sigma})] \leq \mathrm{V}[\hat{\boldsymbol{\beta}}]$$

where the inequality is in terms of positive semidefiniteness. This result is known as the Gauss–Markov theorem.

Suppose that the distribution of $\boldsymbol{y}$ belongs to the class $\mathcal{P}$ in (4). If the matrix $\boldsymbol{\Sigma}$ is fully unknown, a typical estimator of $\boldsymbol{\beta}$ may be a generalized least squares estimator (GLSE) $b(\hat{\boldsymbol{\Sigma}}) = \left(\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{y}$, which is the GME with unknown $\boldsymbol{\Sigma}$ replaced with an estimator $\hat{\boldsymbol{\Sigma}}$. In most cases, $\hat{\boldsymbol{\Sigma}}$ is a function of the residual vector $\boldsymbol{e}$. Hence, in this section, we limit our consideration to such estimators, say $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\boldsymbol{e})$. Thus any GLSE $b(\hat{\boldsymbol{\Sigma}})$ is rewritten as

$$b(\hat{\boldsymbol{\Sigma}}) = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y}$$

by letting $\boldsymbol{C}(\boldsymbol{e}) = \left(\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)$. The matrix $\boldsymbol{C}(\boldsymbol{e})$ is a (random) left inverse of $\boldsymbol{X}$, where a $k \times n$ matrix $\boldsymbol{A}$ is called a left inverse of $\boldsymbol{X}$ if $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{I}_k$. Let $\mathcal{L}$ be the set of left inverses of $\boldsymbol{X}$, and consider the following class of estimators:

$$\mathcal{C}_{\mathrm{GLSE}} = \left\{\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \mid \boldsymbol{C} : \mathbb{R}^n \to \mathcal{L}\right\}.$$

We call the class $\mathcal{C}_{\mathrm{GLSE}}$ the class of GLSEs and call elements of $\mathcal{C}_{\mathrm{GLSE}}$ GLSEs. In fact, the class $\mathcal{C}_{\mathrm{GLSE}}$ contains all the GLSEs defined above. Conversely, for each $\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \in \mathcal{C}_{\mathrm{GLSE}}$, there exists an $n \times n$ (random) positive definite matrix $\Psi = \Psi(\boldsymbol{e})$ depending only on $\boldsymbol{e}$ such that $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\Psi^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Psi^{-1}\boldsymbol{y}$ [see Theorem 3.2 of Kariya and Toyooka (1985) or Proposition 2.3 of Kariya and Kurata (2004)].

For $\mathcal{C}_{\mathrm{GLSE}}$, Kariya (1985) and Kariya and Toyooka (1985) obtained the following fundamental results. They established the results under a quite mild distributional condition, which we do not mention (since it is satisfied by our setting 4).

Proposition 6 (Kariya (1985) and Kariya and Toyooka (1985)) *Let* $\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \in \mathcal{C}_{\mathrm{GLSE}}$.

(i) *If* $\boldsymbol{C}(\cdot)$ *is an even function, i.e.,* $\boldsymbol{C}(\boldsymbol{x}) = \boldsymbol{C}(-\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$*, then* $\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \in \mathcal{C}_{GLSE}$ *is unbiased as long as the expectation is finite.*

(ii) *If* $\boldsymbol{C}(\cdot)$ *is continuous and scale invariant (i.e.,* $\boldsymbol{C}(a\boldsymbol{x}) = \boldsymbol{C}(\boldsymbol{x})$ *for all* $a > 0$ *and* $\boldsymbol{x} \in \mathbb{R}^n$*), then* $\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \in \mathcal{C}_{GLSE}$ *has finite second moments.*

(iii) *The mean squared error matrix of* $\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{e})\boldsymbol{y} \in \mathcal{C}_{GLSE}$ *is bounded from below by the variance–covariance matrix of the GME* $b(\boldsymbol{\Sigma})$*:*

$$\left(\boldsymbol{X}'\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1} = \mathrm{V}[b(\boldsymbol{\Sigma})] \le \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right].$$

The BEE $\hat{\boldsymbol{\beta}}_*$ belongs to the class $\mathcal{C}_{\mathrm{GLSE}}$, that is, the BEE is a GLSE. In fact, by letting

$$\tilde{\boldsymbol{R}}(\boldsymbol{e}) = \mathrm{diag}\left\{\|\boldsymbol{e}_1\|\boldsymbol{I}_m, \ldots, \|\boldsymbol{e}_p\|\boldsymbol{I}_m\right\} : n \times n \text{ and } \boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

and using

$$\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{e}) = \begin{pmatrix} \boldsymbol{e}_1/\|\boldsymbol{e}_1\| \\ \vdots \\ \boldsymbol{e}_p/\|\boldsymbol{e}_p\| \end{pmatrix} = \tilde{\boldsymbol{R}}(\boldsymbol{e})^{-1}\boldsymbol{e} = \tilde{\boldsymbol{R}}(\boldsymbol{e})^{-1}\boldsymbol{M}\boldsymbol{y},$$

we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_*(\boldsymbol{y}) &= \boldsymbol{b} - \boldsymbol{R}(\boldsymbol{e})\boldsymbol{D}(\boldsymbol{u})^{-1}\boldsymbol{F}(\boldsymbol{u})\boldsymbol{u} \\
&= \boldsymbol{b} - \boldsymbol{R}(\boldsymbol{e})\boldsymbol{D}(\boldsymbol{u}(\boldsymbol{e}))^{-1}\boldsymbol{F}(\boldsymbol{u}(\boldsymbol{e}))\boldsymbol{u}(\boldsymbol{e}) \\
&= \left\{\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' - \boldsymbol{R}(\boldsymbol{e})\boldsymbol{D}(\boldsymbol{u}(\boldsymbol{e}))^{-1}\boldsymbol{F}(\boldsymbol{u}(\boldsymbol{e}))\tilde{\boldsymbol{R}}(\boldsymbol{e})^{-1}\boldsymbol{M}\right\}\boldsymbol{y} \\
&= \boldsymbol{C}_*(\boldsymbol{e})\boldsymbol{y} \quad \text{(say)},
\end{aligned}$$

where $\boldsymbol{R}(\boldsymbol{e})$ is defined in (18). The matrix $\boldsymbol{C}_*(\boldsymbol{e})$ clearly satisfies $\boldsymbol{C}_*(\boldsymbol{e})\boldsymbol{X} = \boldsymbol{I}_k$, from which $\hat{\boldsymbol{\beta}}_* \in \mathcal{C}_{GLSE}$ follows.

The BEE $\hat{\boldsymbol{\beta}}_*$ is unbiased. To see this, it suffices to show that $\boldsymbol{C}_*(\boldsymbol{e})$ is an even function of $\boldsymbol{e}$. The matrices $\boldsymbol{R}(\boldsymbol{e})$ and $\tilde{\boldsymbol{R}}(\boldsymbol{e})$ are clearly even functions of $\boldsymbol{e}$. Since the matrices $\boldsymbol{D}(\boldsymbol{u})$ and $\boldsymbol{F}(\boldsymbol{u})$ depends on $\boldsymbol{e}$ only through $t_{ij}(\boldsymbol{u})$s in (21), we use the following lemma:

Lemma 7 *The quantities* $t_{ij}(\boldsymbol{u}) = \mathrm{E}_{\mathbf{0},\boldsymbol{\Lambda}}\left[\|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\||\boldsymbol{u}\right]$*s, when they are viewed as functions of* $\boldsymbol{e}$ *via* $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{e})$*, are even functions of* $\boldsymbol{e}$ *for all* $i, j = 1, \ldots, p$*. That is,* $t_{ij}(\boldsymbol{u}(\boldsymbol{e})) = t_{ij}(\boldsymbol{u}(-\boldsymbol{e}))$*.*

Proof First we note that the distributions of $-\boldsymbol{e}$ and $-\boldsymbol{u}$ are the same as those of $\boldsymbol{e}$ and $\boldsymbol{u}$, respectively. Fix i, j arbitrarily, and let $f_{ij}(\boldsymbol{e}) = \|\boldsymbol{e}_i\|\|\boldsymbol{e}_j\|$. Then $f_{ij}(\boldsymbol{e})$ is an even function of $\boldsymbol{e}$, and the quantity $t_{ij}(\boldsymbol{u})$ is expressed as

$$t_{ij}(\boldsymbol{u}) = \mathrm{E}_{\mathbf{0},\boldsymbol{\Lambda}}[f_{ij}(\boldsymbol{e})|\boldsymbol{u}].$$

Let $\mathbb{P}$ be the distribution of $\boldsymbol{e}$ on the Borel space $(\mathbb{R}^n, \mathbb{B})$, and let $(\mathcal{U}, \mathbb{C}, \mathbb{P}_U)$ be the probability space induced by $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{e})$ from $(\mathbb{R}^n, \mathbb{B}, \mathbb{P})$, where $\mathbb{P}_U(C) = \mathbb{P}(\boldsymbol{u}(\boldsymbol{e}) \in C)$ for $C \in \mathbb{C}$.

Let $\mathbb{Q}$ be a measure defined on $\mathcal{U}$ as

$$\mathbb{Q}(C) = \int_{\{\boldsymbol{u}(\boldsymbol{e}) \in C\}} f_{ij}(\boldsymbol{e})\mathbb{P}(d\boldsymbol{e}), \quad C \in \mathbb{C}.$$

Since the distribution of $-\boldsymbol{e}$ is the same as that of $\boldsymbol{e}$, and since $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{e})$ is an odd function of $\boldsymbol{e}$ (i.e., $\boldsymbol{u}(-\boldsymbol{e}) = -\boldsymbol{u}(\boldsymbol{e})$), we can see that the measure $\mathbb{Q}$ satisfies $\mathbb{Q}(C) = \mathbb{Q}(-C)$ for each C, where $-C = \{-\boldsymbol{u} | \boldsymbol{u} \in C\}$. In fact,

$$\begin{aligned}\mathbb{Q}(-C) &= \int_{\{\boldsymbol{u}(\boldsymbol{e}) \in -C\}} f_{ij}(\boldsymbol{e})\mathbb{P}(d\boldsymbol{e}) = \int_{\{\boldsymbol{u}(-\boldsymbol{e}) \in C\}} f_{ij}(\boldsymbol{e})\mathbb{P}(d\boldsymbol{e}) \\ &= \int_{\{\boldsymbol{u}(\boldsymbol{x}) \in C\}} f_{ij}(-\boldsymbol{x})\mathbb{P}(d\boldsymbol{x}) = \int_{\{\boldsymbol{u}(\boldsymbol{x}) \in C\}} f_{ij}(\boldsymbol{x})\mathbb{P}(d\boldsymbol{x}) \\ &= \mathbb{Q}(C),\end{aligned}$$

where the second equality follows since $\boldsymbol{u}(\boldsymbol{e})$ is odd in $\boldsymbol{e}$, the third is obtained by transforming $\boldsymbol{e} = -\boldsymbol{x}$ and using the invariance of the distribution $\mathbb{P}$, and the fourth is due to $f_{ij}(-\boldsymbol{x}) = f_{ij}(\boldsymbol{x})$. The conditional expectation $t_{ij}(\boldsymbol{u})$ is, by definition, measurable with respect to $\mathbb{C}$ and satisfies

$$\mathbb{Q}(C) = \int_C t_{ij}(\boldsymbol{u})\mathbb{P}_U(d\boldsymbol{u}) \text{ for any } C \in \mathbb{C}.$$

Hence we have for each $C \in \mathbb{C}$,

$$\int_C t_{ij}(-\boldsymbol{u})\mathbb{P}_U(d\boldsymbol{u}) = \int_{-C} t_{ij}(\boldsymbol{x})\mathbb{P}_U(d\boldsymbol{x}) = \mathbb{Q}(-C) = \mathbb{Q}(C) = \int_C t_{ij}(\boldsymbol{u})\mathbb{P}_U(d\boldsymbol{u}),$$

where the first equality follows by transforming $\boldsymbol{x} = -\boldsymbol{u}$ and noting that the distribution $\mathbb{P}_U$ of $\boldsymbol{u}$ is invariant under $\boldsymbol{u} \to -\boldsymbol{u}$. This shows that

$$t_{ij}(-\boldsymbol{u}) = t_{ij}(\boldsymbol{u}),$$

which implies that $t_{ij}(\boldsymbol{u}(-\boldsymbol{e})) = t_{ij}(-\boldsymbol{u}(\boldsymbol{e})) = t_{ij}(\boldsymbol{u}(\boldsymbol{e}))$, and hence t_{ij} is an even function of $\boldsymbol{e}$. □

Thus the matrices $\boldsymbol{D}(\boldsymbol{u}(\boldsymbol{e}))$ and $\boldsymbol{F}(\boldsymbol{u}(\boldsymbol{e}))$ are even functions of $\boldsymbol{e}$, which in turn shows that $\boldsymbol{C}_*(\boldsymbol{e})$ is also even in $\boldsymbol{e}$. Hence, by using (i) of Proposition 6, we see that the BEE $\hat{\boldsymbol{\beta}}_*$ is unbiased:

$$\mathrm{E}[\hat{\boldsymbol{\beta}}_*] = \boldsymbol{\beta}. \tag{41}$$

Next we show

$$\boldsymbol{C}_*(a\boldsymbol{e}) = \boldsymbol{C}_*(\boldsymbol{e}) \text{ for any } a > 0. \tag{42}$$

It is easy to see that the matrices $\boldsymbol{R}(\boldsymbol{e})$ and $\tilde{\boldsymbol{R}}(\boldsymbol{e})$ satisfy

$$\boldsymbol{R}(a\boldsymbol{e}) = a\boldsymbol{R}(\boldsymbol{e}) \text{ and } \tilde{\boldsymbol{R}}(a\boldsymbol{e}) = a\tilde{\boldsymbol{R}}(\boldsymbol{e}) \text{ for any } a > 0.$$

On the other hand, since $\boldsymbol{u}$ is scale invariant as a function of $\boldsymbol{e}$ (i.e., $\boldsymbol{u}(a\boldsymbol{e}) = \boldsymbol{u}(\boldsymbol{e})$), the matrices $\boldsymbol{D}(\boldsymbol{u})$ and $\boldsymbol{F}(\boldsymbol{u})$ satisfy

$$\boldsymbol{D}(\boldsymbol{u}(a\boldsymbol{e})) = \boldsymbol{D}(\boldsymbol{u}(\boldsymbol{e})) \text{ and } \boldsymbol{F}(\boldsymbol{u}(a\boldsymbol{e})) = \boldsymbol{F}(\boldsymbol{u}(\boldsymbol{e})).$$

Thus we have (42). Hence by (ii) of Proposition 6, the BEE $\hat{\boldsymbol{\beta}}_*$ has finite second moments. The efficiency of the BEE can be measured in terms of its variance–covariance matrix $\mathrm{V}[\hat{\boldsymbol{\beta}}_*]$. Applying (iii) of Proposition 6, we have a lower bound for $\mathrm{V}[\hat{\boldsymbol{\beta}}_*]$:

$$\left(\boldsymbol{X}'\left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1} = \mathrm{V}[b(\boldsymbol{\Sigma})] \leq \mathrm{V}\left[\hat{\boldsymbol{\beta}}_*\right] = \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}\right)'\right]. \tag{43}$$

Summarizing the above conclusions (41), (42) and (43) yields

Theorem 8 *The BEE* $\hat{\boldsymbol{\beta}}_*$ *is an unbiased GLSE with finite second moments, whose variance–covariance matrix is bounded from below by that of the GME* $b(\boldsymbol{\Sigma})$.

Consider the following estimator of $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\boldsymbol{e}) = \left(\hat{\sigma}_{ii}^{1/2}\hat{\sigma}_{jj}^{1/2}\rho_{ij}\right) \text{ with } \hat{\sigma}_{ii} = \hat{\sigma}_{ii}(\boldsymbol{e}) = \boldsymbol{e}_i'\boldsymbol{e}_i/m. \tag{44}$$

Then $b(\hat{\boldsymbol{\Sigma}}) = \left(\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_m\right)\boldsymbol{y}$ with $\hat{\boldsymbol{\Sigma}}$ in (44) is a natural estimator that corresponds to the structure of $\boldsymbol{\Sigma}$ in (6). It belongs to $\mathcal{C}_{GLSE}$, is unbiased and has finite second moments. As can be easily seen, this estimator is also an equivariant estimator. Hence we have the inequality

$$R\left(\hat{\boldsymbol{\beta}}_*, (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right) \leq R\left(b(\hat{\boldsymbol{\Sigma}}), (\boldsymbol{\beta}, \boldsymbol{\Sigma})\right).$$

It is also of interest to compare the two estimators in terms of their variance–covariance matrices, which is left open.

We conclude this paper with a simple numerical example, in which the risks of the BEE, the OLSE and two GLSEs are compared. One of the two GLSEs is a GLSE $b(\hat{\boldsymbol{\Sigma}})$ with $\hat{\boldsymbol{\Sigma}} = (\boldsymbol{e}_i'\boldsymbol{e}_j/m)$, which does not use the information on the known correlation coefficients. The other is a GLSE in (44). We call these two estimators GLSE1 and GLSE2, respectively. The models treated are two-equation models ($p = 2$) with $m = 10$, say Models 1 and 2. In Model 1, we set $k_1 = 2$ and $k_2 = 3$, that is,

$\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}) : 10 \times 2$ and $\boldsymbol{X}_2 = (\boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \boldsymbol{x}_{23}) : 10 \times 3$, where $\boldsymbol{x}_{11}$ and $\boldsymbol{x}_{21}$ are vectors of all ones, and the other vectors $\boldsymbol{x}_{12}, \boldsymbol{x}_{22}, \boldsymbol{x}_{23}$ are chosen according to the uniform distribution on the interval $[-1, 1]$. Model 2 is obtained by adding several column vectors to Model 1: The model has $\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \boldsymbol{x}_{13}) : 10 \times 3$ and $\boldsymbol{X}_2 = (\boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \ldots, \boldsymbol{x}_{26}) : 10 \times 6$, where $\boldsymbol{x}_{13}, \boldsymbol{x}_{25}$ and $\boldsymbol{x}_{26}$ are the squares of $\boldsymbol{x}_{12}, \boldsymbol{x}_{22}$ and $\boldsymbol{x}_{23}$, respectively, and the elements of $\boldsymbol{x}_{24}$ are the products of those of $\boldsymbol{x}_{22}$ and $\boldsymbol{x}_{23}$. The use of uniform distribution for $\boldsymbol{X}$ is due to Ando (2011) and Zellner and Ando (2010).

The tables below illustrate the estimates of the risks (15) of the above four estimators based on 1,000,000 replications. When $\rho = 0$, the OLSE, the BEE and GLSE2 coincide. For each estimator, its risk monotonically increases as ρ increases. On the other hand, whatever ρ is, the BEE has the smallest risk among the four estimators. The efficiency of the BEE relative to GLSE2 is higher in Model 2 than in Model 1.

Model 1	OLSE	GLSE1	GLSE2	BEE
$\rho = 0.0$	4.995831	5.232456	4.995831	4.995831
$\rho = 0.2$	5.116554	5.255603	5.015679	5.01504
$\rho = 0.4$	5.528819	5.324513	5.060269	5.057798
$\rho = 0.6$	6.551833	5.49757	5.150907	5.141146
$\rho = 0.8$	9.948265	6.109094	5.401158	5.326356

Model 2	OLSE	GLSE1	GLSE2	BEE
$\rho = 0$	9.003876	9.329488	9.003876	9.003876
$\rho = 0.2$	9.190073	9.402881	9.045734	9.036417
$\rho = 0.4$	9.854493	9.666779	9.19308	9.14603
$\rho = 0.6$	11.51532	10.33883	9.558842	9.386615
$\rho = 0.8$	16.94042	12.58923	10.79009	9.982628

Acknowledgments The authors are grateful to two anonymous reviewers for their constructive comments. Kurata's portion of this work was supported by JSPS KAKENHI Grant Number 26330035.

References

Ando, T. (2011). Bayesian variable selection for the seemingly unrelated regression models with a large number of predictors. *Journal of the Japan Statistical Society*, *41*(2), 187–203.

Bilodeau, M. (1990). On the choice of the loss function in covariance estimation. *Statistics and Decisions*, *8*(2), 131–139.

Eaton, M. L. (1989). Group invariance applications in statistics, NSF-CBMS Regional Conference Series in Probability and Statistics, 1.

Fang, K.T., Lam, P.C.B., Wu, Q.G. (1997). Estimation for seemingly unrelated regression equations. *Statistics and Decisions*, *15*(2), 183–189.

Giri, N. C. (1996). *Group invariance in statistical inference*. Singapore: World Scientific Publishing.

Horn, R. A., Johnson, C. R. (1991). *Topics in matrix analysis*. Cambridge: Cambridge University Press.

Kariya, T. (1981). Bounds for the covariance matrices of Zellner's estimator in the SUR model and the 2SAE in a heteroscedastic model. *Journal of the American Statistical Association*, *76*, 975–979.

Kariya, T. (1985). A nonlinear version of the Gauss-Markov theorem. *Journal of the American Statististical Association*, *80*, 476–477.
Kariya, T. (1989). Equivariant estimation in a model with an ancillary statistic. *Annals of Statistcs*, *17*(2), 920–928.
Kariya, T., Kurata, H. (2004). *Generalized least squares*. Chichester: Wiley.
Kariya, T., Toyooka, Y. (1985). Nonlinear versions of the Gauss–Markov theorem and GLSE, Multivariate Analysis VI (Pittsburgh, Pa., 1983), 345–354, North-Holland, Amsterdam.
Kurata, H. (1999). On the efficiencies of several generalized least squares estimators in a seemingly unrelated regression model and a heteroscedastic model. *Journal of Multivariate Analysis*, *70*(1), 86–94.
Kurata, H., Kariya, T. (1996). Least upper bound for the covariance matrix of a generalized least squares estimator in regression with applications to a seemingly unrelated regression model and a heteroscedastic model. *Annals of Statistics*, *24*(4), 1547–1559.
Lehmann, E. L., Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.
Liu, A. (2002). Efficient estimation of two seemingly unrelated regression equations. *Journal Multivariate Analysis*, *82*(2), 445–456.
Ma, T., Ye, R. (2010). Efficient improved estimation of the parameters in two seemingly unrelated regression models. *Journal of Statistical Planning and Inference*, *140*(9), 2749–2754.
Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: Wiley.
Srivastava, V. K., Giles, D. E. A. (1987). *Seemingly unrelated regression equation models*. New York: Marcel Dekker.
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–369.
Zellner, A. (1963). Estimators for seemingly unrelated regression equations: some exact sample results. *Journal of the American Statistical Association*, *58*, 977–992.
Zellner, A., Ando, T. (2010). A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model. *Journal of Econometrics*, *159*(1), 33–45.