Projektbereich B
Discussion Paper No. B–300

On Inequality Constrained Generalized Least Squares Selections in the General Possibly Singular Gauß-Markov Model: A Projector Theoretical Approach

Hans Joachim WERNER*, Bonn, Germany
and
Cemil YAPAR, Trabzon, Turkey

1994

This paper appears in: *Linear Algebra Appl. (5th Special Issue devoted to Statistics, in press for publication in 1995).*

* Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

On Inequality Constrained Generalized Least Squares Selections in the General Possibly Singular Gauß-Markov Model: A Projector Theoretical Approach

H. J. Werner [1]
Institute for Econometrics and Operations Research
Econometrics Unit, University of Bonn
D-53113 Bonn, Germany

C. Yapar
Karadeniz Technical University
Faculty of Arts and Sciences, Dept. of Stat. & Comp. Sci.
61080 Trabzon, Turkey

Abstract

This paper deals with the general possibly singular linear model. It is assumed that in addition to the sample information we have some nonstochastic prior information concerning the unknown regression coefficients that can be expressed in form of linear independent inequality constraints. Since these constraints are *part and parcel* of the model the *inequality constrained generalized least squares* (ICGLS) problem arises that contains some unknown aspects up to now. Based on a projector theoretical approach we show in this paper how the set of *ICGLS selections* under the constrained model is related to the set of *GLS selections* under the associated unconstrained model. As a by-product we obtain an interesting method for determining an ICGLS selection from a GLS selection. The insights gained from our considerations might also be useful in a future study of the statistical properties of ICGLS estimators. Certain special model cases are also considered. Some of the results discussed in [29] and [7] are reobtained.

Keywords: Gauß-Markov model, singular model, perfect multicollinearity, linear inequality constraints, inequality constrained generalized least squares problem, oblique projectors, generalized inverses.

JEL-Classification: C20.

---

[1] Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

On Inequality Constrained Generalized Least Squares Selections in the General Possibly Singular Gauß-Markov Model: A Projector Theoretical Approach

H. J. Werner [1]
Institute for Econometrics and Operations Research
Econometrics Unit, University of Bonn
D-53113 Bonn, Germany

C. Yapar
Karadeniz Technical University
Faculty of Arts and Sciences, Dept. of Stat. & Comp. Sci.
61080 Trabzon, Turkey

Abstract

This paper deals with the general possibly singular linear model. It is assumed that in addition to the sample information we have some nonstochastic prior information concerning the unknown regression coefficients that can be expressed in form of linear independent inequality constraints. Since these constraints are *part and parcel* of the model the *inequality constrained generalized least squares* (ICGLS) problem arises that contains some unknown aspects up to now. Based on a projector theoretical approach we show in this paper how the set of *ICGLS selections* under the constrained model is related to the set of *GLS selections* under the associated unconstrained model. As a by-product we obtain an interesting method for determining an ICGLS selection from a GLS selection. The insights gained from our considerations might also be useful in a future study of the statistical properties of ICGLS estimators. Certain special model cases are also considered. Some of the results discussed in [29] and [7] are reobtained.

---

1. Introduction

Let $\mathbb{R}^n$, $\mathbb{R}^{n,m}$, and $\mathcal{P}^{n,n}$ denote the set of n-dimensional real column vectors, the set of $n \times m$ real matrices, and the set of real nonnegative definite and symmetric $n \times n$ matrices (nnds), respectively. Given $A \in \mathbb{R}^{n,m}$, the symbols A^t, $\mathcal{R}(A)$, $\mathcal{N}(A)$, and $\operatorname{rank}(A)$ will denote the transpose, the range space, the null space, and the rank, respectively, of A. In addition,

[1] Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

let A^- denote an arbitrary generalized inverse of A satisfying $AA^-A = A$. Further, I and 0, respectively, will stand for the identity matrix and the zero matrix of whatever size is appropriate to the context.

Consider the following restricted Gauß-Markov model

$$\mathcal{L}_r := (y,\ X\beta,\ V \mid R\beta \geq r), \tag{1.1}$$

in which y is an observable random vector with expectation $X\beta$ and dispersion V; the vector β of unknown regression coefficients satisfies the a priori constraints $R\beta \geq r$; $X \in \mathbb{R}^{n,m}$ and $V \in \mathcal{P}^{n,n}$ are known model matrices that need not be of full column rank as in [29] and [7]; $R \in \mathbb{R}^{p,m}$ and $r \in \mathcal{R}(R)$ are also known; and R has full row rank p. Only for the sake of simplicity, it is further assumed throughout this paper that

$$\mathcal{R}(X) \subseteq \mathcal{R}(V). \tag{1.2}$$

Restrictions in the form of inequalities frequently arise on the unknown parameters as work with disequilibrium models [1, 10], simultaneous Tobit and Probit models [21], and other related models has shown. Although models like $\mathcal{L}_r$ are thus of growing interest, they contain some unknown aspects up to now. Unfortunately, inequality constraints pose statistical problems, and so there has been a subsequent lag in the determination of the statistical properties of estimators in such models. Much of the existing literature has avoided these problems by assuming that X, V and R are all of a very particular structure; see Section 1 in [29] for more details in this respect. If the matrices X and V in model $\mathcal{L}_r$ are both of full column rank, then some closed form expressions for the (highly nonlinear) ICGLS estimator of β can be found in [29]. The main purpose of this paper is to drop these restrictive rank assumptions.

In the sequel it is convenient to denote by

$$\mathcal{L}_u := (y,\ X\beta,\ V) \tag{1.3}$$

the model that is obtained from $\mathcal{L}_r$ by ignoring the inequality restrictions $R\beta \geq r$. Note that, under the assumption (1.2),

$$y \in \mathcal{R}(V) \quad \text{(a.s.)}, \tag{1.4}$$

irrespective of under which of the two models y is observed. By Theorem 2.3, due to invariance, an arbitrary generalized inverse V^- of V can be used to define a norm

$$\|x\|_{V^-} := (x^t V^- x)^{\frac{1}{2}}$$

on $\mathcal{R}(V)$. The mathematical programming problem

$$\text{minimize} \quad \|y - Xb\|^2_{V^-} \quad \text{subject to} \quad Rb \geq r \tag{1.5}$$

is hence well defined for each $y \in \mathcal{R}(V)$. Any optimal solution to this convex-quadratic optimization problem, that is any vector from

$$\text{argmin}_{Rb \geq r}\ \|y - Xb\|^2_{V^-} \tag{1.6}$$

is called an *ICGLS solution* (for β) and is henceforth denoted by $\tilde{\beta}(y)$. Although (1.5) possesses an optimal solution for each $y \in \mathcal{R}(V)$, (1.6) need not be a singleton. In which case there do exist many different functions f with $f(y)$ representing an ICGLS solution for each $y \in \mathcal{R}(V)$. It seems reasonable to call any such function an *ICGLS selection* for β and to reserve the term *ICGLS estimator* for exactly those situations where there does exist only one ICGLS selection on $\mathcal{R}(V)$. The set of all ICGLS selections will be denoted by $\{\tilde{\beta}(\cdot)\}$.

The paper is organized as follows. Section 2 consists of some miscellaneous results on generalized inverses and (generally oblique) projectors, which are important in this text. Section 3 and Section 4 deal with the ICGLS problem. In particular, it is shown there how the ICGLS selections for β under model $\mathcal{L}_r$ are related to the GLS selections for β under the associated unconstrained model $\mathcal{L}_u$. A nice method for determining an ICGLS selection from a GLS selection is obtained as a bonus. Two spcecial (extreme) model cases where either

$$\mathcal{R}(R^t) \subseteq \mathcal{R}(X^t) \tag{1.7}$$

or

$$\mathcal{R}(R^t) \cap \mathcal{R}(X^t) = \{0\}. \tag{1.8}$$

is assumed in addition to (1.2) are also investigated in detail. Many of the results obtained in this paper appear to be new. Some of the results in [29] and [7] are reobtained.

It is interesting to mention here that (1.7) is a necessary and sufficient condition for $R\beta$ to be *linearly unbiasedly estimable* under model $\mathcal{L}_u$; cf. [19] or [23]. If the matrices X and R satisfy (1.8) they are called *weakly complementary* to one another; see Section 2. It is well-known that the set of GLS solutions for β under model $\mathcal{L}_u$ is not a singleton whenever X is deficient in column rank; see also (3.13). As in fixed effects models with balanced data, constraints in the form of linear equations on the regression coefficents are then frequently *artificially* introduced for the sole purpose of reducing the original set of GLS solutions. Notice that it is essential for such a restrictor matrix to be weakly complementary to the regressor matrix (compare [5], [8], [9], [25]; see also Theorem 3.9). These imposed constraints which are carefully to be distinguished from restrictions considered as an *integral part* of the underlying model have a long history; cf. [30], [14], [20], [5], [15], [8], [9], and [25].

2. G-Inverses and Projectors

Let $\mathcal{M}$ and $\mathcal{N}$ be linear subspaces in the n-dimensional real space $\mathbb{R}^n$. Then $\mathcal{M}^\perp$ will stand for the orthogonal complement of $\mathcal{M}$ in $\mathbb{R}^n$ (with respect to the usual inner product), and if $\mathcal{M} \cap \mathcal{N} = \{0\}$, then $\mathcal{M} \oplus \mathcal{N}$ will denote the direct sum of $\mathcal{M}$ and $\mathcal{N}$. Next, if $\mathcal{N}$ is a direct complement of $\mathcal{M}$ (i.e. $\mathbb{R}^n = \mathcal{M} \oplus \mathcal{N}$), then $P_{\mathcal{M},\mathcal{N}}$ will denote the well-defined (generally *oblique*) projector on $\mathcal{M}$ along $\mathcal{N}$, and if $\mathcal{N} = \mathcal{M}^\perp$, then $P_{\mathcal{M}}$ will denote the corresponding (*orthogonal*) projector. Notice that $P_{\mathcal{M},\mathcal{N}}$ may be defined by $P_{\mathcal{M},\mathcal{N}}u = u$ if $u \in \mathcal{M}$ and $P_{\mathcal{M},\mathcal{N}}u = 0$ if $u \in \mathcal{N}$ (see, e.g., [16, pp. 106–113]). Observe that the relations

$$\mathcal{R}(A^t)^\perp = \mathcal{N}(A) \quad \text{and} \quad \mathcal{N}(A^t)^\perp = \mathcal{R}(A) \tag{2.1}$$

hold for each matrix $A \in \mathbb{R}^{n,m}$. Recall that any projector $P_{\mathcal{M},\mathcal{N}}$ is an idempotent matrix, i.e. $P^2_{\mathcal{M},\mathcal{N}} = P_{\mathcal{M},\mathcal{N}}$, and that conversely every idempotent matrix P is a projector, namely

$P = P_{\mathcal{R}(P),\mathcal{N}(P)}$. If $P^2 = P$, then $(I-P)^2 = I - P$ and $(P^t)^2 = P^t$. Check that

$$I - P_{\mathcal{M},\mathcal{N}} = P_{\mathcal{N},\mathcal{M}}. \tag{2.2}$$

In view of (2.1) it is further clear that

$$(P_{\mathcal{M},\mathcal{N}})^t = P_{\mathcal{N}^\perp,\mathcal{M}^\perp}. \tag{2.3}$$

For given $A \in \mathbb{R}^{n,m}$ and $\mathcal{M},\mathcal{N} \subseteq \mathbb{R}^m$, it is convenient to denote by $\mathcal{M}+\mathcal{N}$, $A\mathcal{M}$, $\mathcal{N}_c(A)$, and $\mathcal{R}_c(A)$, respectively, the Minkowski sum of $\mathcal{M}$ and $\mathcal{N}$, the image of $\mathcal{M}$ under A, the set of all direct complements of $\mathcal{N}(A)$, and the set of all direct complements of $\mathcal{R}(A)$. Recall that $(\mathcal{M}+\mathcal{N})^\perp = \mathcal{M}^\perp \cap \mathcal{N}^\perp$ and $(\mathcal{M}\cap\mathcal{N})^\perp = \mathcal{M}^\perp + \mathcal{N}^\perp$. Also notice that $\mathcal{N}^\perp \subseteq \mathcal{M}^\perp$ whenever $\mathcal{M} \subseteq \mathcal{N}$.

Now let $A \in \mathbb{R}^{n,m}$, let $\mathcal{M} \in \mathcal{N}_c(A)$, and let $\mathcal{S} \in \mathcal{R}_c(A)$. Consider the matrix equations

$$\begin{array}{llll} (G1) & AXA = A, & (G\mathcal{M}) & XA = P_{\mathcal{M},\mathcal{N}(A)}, \\ (G2) & XAX = X, & (G\mathcal{S}) & AX = P_{\mathcal{R}(A),\mathcal{S}}. \end{array} \tag{2.4}$$

Suppose that $\emptyset \neq \eta \subseteq \{1,2,\mathcal{M},\mathcal{S}\}$. Then let $A\eta$ denote the set of all those matrices X which satisfy equations (Gi) for all $i \in \eta$. Any $X \in A\eta$ is called an *η-inverse* of A, and is also denoted by A^η. $\{1\}$-inverses are usually called *generalized inverses* or *g-inverses* and are also denoted by A^-. For an extensive discussion of the theory of g-inversion, we refer, e.g., to the books by Ben-Israel and Greville [2], Hartung and Werner [9], Pringle and Rayner [15], Rao and Mitra [16]; for a geometric approach, to Werner [24, chapter 1] and Rao and Yanai [17]; and for a projector theoretical one, e.g., to the paper by Langenhop [11]. Only for the sake of clarity and for easier reference, a few basic results are summarized in Theorem 2.1 (cf. [24], see also [27]).

Theorem 2.1.

(i) The $\{2,\mathcal{M},\mathcal{S}\}$-inverse of A exists uniquely. The $\{2,\mathcal{R}(A^t),\mathcal{N}(A^t)\}$-inverse of A coincides with the *Moore-Penrose inverse* of A and is usually denoted by $A^\dagger$. Hence, $(A^\dagger)^\dagger = A$.

(ii) Any $\{\mathcal{M}\}$-inverse of A and likewise any $\{\mathcal{S}\}$-inverse of A is always a $\{1\}$-inverse of A. Conversely, for each $\{1\}$-inverse of A there uniquely exist an $\mathcal{M} \in \mathcal{N}_c(A)$ and an $\mathcal{S} \in \mathcal{R}_c(A)$ such that $X \in A\{\mathcal{M},\mathcal{S}\}$. Moreover, if $X \in A\{\mathcal{M},\mathcal{S}\}$, then $XAX = A^{\{2,\mathcal{M},\mathcal{S}\}}$.

(iii) If $X \in A\{\mathcal{M},\mathcal{S}\}$, then $\mathcal{M} = \mathcal{R}(XA) \subseteq \mathcal{R}(X)$, and $\mathcal{N}(X) \subseteq \mathcal{S} = \mathcal{N}(AX)$. In particular, $X\mathcal{S} \subseteq \mathcal{N}(A)$. Moreover, $X = A^{\{2,\mathcal{M},\mathcal{S}\}}$ iff $\mathcal{R}(X) = \mathcal{M}$ and $\mathcal{N}(X) = \mathcal{S}$. Furthermore, $A\{1\} = A\{1,2\}$ iff A is of full column rank and/or A is of full row rank. In particular, $A^-A = I$ iff A is of full column rank. Likewise, $AA^- = I$ iff A is of full row rank.

(iv) If $X \in A\{\mathcal{M},\mathcal{S}\}$, then $X^t \in A^t\{\mathcal{S}^\perp,\mathcal{M}^\perp\}$. Hence $(A^t)^\dagger = (A^\dagger)^t$.

(v) If A is nonsingular, then its $\{1\}$-inverses all coincide with its *regular inverse*, i.e. $A\{1\} = \{A^{-1}\}$.

Next, let A and B be two real matrices, both of the same column number. From Werner [27] we have the following definitions:

(a) B is said to be *weakly complementary* to A, if $\mathcal{R}(A^t) \cap \mathcal{R}(B^t) = \{0\}$.

(b) B is said to be *weakly bicomplementary* to A, if B and B^t are weakly complementary to A and A^t, respectively.

A pair of weakly bicomplementary matrices is also often said to be a pair of *disjoint* matrices (also written $A+B = A \oplus B$), cf. [13]. The connections between these concepts and the concept of generalized inversion are discussed in detail in [27]. Here we only cite the following important result.

Theorem 2.2. For given $A,\ B \in \mathbb{R}^{n,m}$, the following conditions are equivalent:

(i) $A + B = A \oplus B$;

(ii) $(A + B)\{1\} \subseteq A\{1\}$;

(iii) $(A + B)\{\mathcal{M}, \mathcal{S}\} \subseteq B\{\mathcal{M} \cap \mathcal{N}(A), \mathcal{S} \oplus \mathcal{R}(A)\}$ for each $\mathcal{M} \in \mathcal{N}_c(A + B)$ and $\mathcal{S} \in \mathcal{R}_c(A + B)$.

In which case, in particular,

$$B(A + B)^{-}A = 0, \tag{2.5}$$

irrespective of the choice of $(A + B)^{-}$.

Although the following invariance property is known (cf. [16]), a simple illustrative proof is given.

Theorem 2.3. Let $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$. If $\mathcal{R}(A) \subseteq \mathcal{R}(W)$, then

(i) $A^tW^{-}A$ is invariant for any choice of W^{-};

(ii) $A^tW^{-}A \in \mathcal{P}^{m,m}$;

(iii) $\mathcal{R}(A^tW^{-}A) = \mathcal{R}(A^t)$ and $\mathcal{N}(A^tW^{-}A) = \mathcal{N}(A)$.

Proof: If $\mathcal{R}(A) \subseteq \mathcal{R}(W)$, then clearly $A = WZ$ for some matrix Z. Consequently $A^tW^{-}A = Z^tWW^{-}WZ = Z^tWZ$, and the results are all plain because W is nnds. ∎

We are now in the position to derive some interesting results on (generally oblique) projectors and associated g-inverses that will play a key role in Section 3. To that end, let $W \in \mathcal{P}^{n,n}$ and $A \in \mathbb{R}^{n,m}$ be given matrices. Then we have the following direct-sum decomposition

$$\mathcal{R}(A, W) = \mathcal{R}(A) \oplus W\mathcal{N}(A^t) \tag{2.6}$$

for the range space of the block partitioned matrix (A, W) (cf. Lemma 3.2 in [26]). It is convenient to denote by $\mathcal{P}(A \mid W)$ the set of all those projectors $P \in \mathbb{R}^{n,n}$ which satisfy

$$\mathcal{R}(P) = \mathcal{R}(A) \quad \text{and} \quad W\mathcal{N}(A^t) \subseteq \mathcal{N}(P). \tag{2.7a-b}$$

That such a projector does always exist is evident from (2.6). A general efficient representation for $\mathcal{P}(A \mid W)$ is given by

$$\mathcal{P}(A \mid W) = \{P_{\mathcal{R}(A), W\mathcal{N}(A^t)\oplus\mathcal{T}} \mid \mathcal{T} \in \mathcal{R}_c(A, W)\}, \tag{2.8}$$

that is, there is a one-to-one correspondence between $P \in \mathcal{P}(A \mid W)$ and $\mathcal{T} \in \mathcal{R}_c(A, W)$. (2.8) tells us that $\mathcal{P}(A \mid W)$ is a singleton iff $\mathcal{R}(A, W) = \mathbb{R}^n$. Clearly, on account of (2.7a), for each $P \in \mathcal{P}(A \mid W)$, there exists a matrix, say A^P, such that

$$P = AA^P. \tag{2.9}$$

From the defintion of η-inversion we further know that

$$AA^P = P_{\mathcal{R}(A), W\mathcal{N}(A^t)\oplus \mathcal{T}} \quad \text{iff} \quad A^P \in A\{W\mathcal{N}(A^t)\oplus \mathcal{T}\}.$$

In other words, only particular g-inverses of A can serve as A^P in (2.9). For convenience we denote the set of all these g-inverses by the symbol $\mathcal{G}(A \mid W)$. Evidently,

$$\mathcal{G}(A \mid W) = \bigcup_{\mathcal{T}\in \mathcal{R}_c(A,W)} A\{W\mathcal{N}(A^t)\oplus \mathcal{T}\}.$$

Although Theorem 2.1 (iii) tells us something about the geometry of these inverses, a natural question is: Given W and A, how can we explicitly compute such an inverse in terms of W and A? A satisfactory answer is given in the next three theorems.

Theorem 2.4. For given $W \in \mathcal{P}^{n,n}$ and $A \in \mathbb{R}^{n,m}$, let $\mathcal{P}(A \mid W)$ be defined as before. Then

$$\mathcal{P}(A \mid W) = \{A\left[A^t(W+AA^t)^- A\right]^- A^t(W+AA^t)^- \mid (\cdot)^- \in (\cdot)\{1\}\}.$$

In particular,

$$A\left[A^t(W+AA^t)^- A\right]^- A^t(W+AA^t)^{\{\mathcal{T}\}} = P_{\mathcal{R}(A), W\mathcal{N}(A^t)\oplus \mathcal{T}},$$

for each $\mathcal{T} \in \mathcal{R}_c(A,W)$.

Proof: Since $W + AA^t \in \mathcal{P}^{n,n}$ and $\mathcal{R}(A) \subseteq \mathcal{R}(W+AA^t)$, we know from Theorem 2.3 that $H := A^t(W+AA^t)^- A$ is a matrix being nnds and invariant for any choice of $(W+AA^t)^-$. In addition, also by Theorem 2.3, $\mathcal{R}(H) = \mathcal{R}(A^t)$ and $\mathcal{N}(H) = \mathcal{N}(A)$. Next consider an arbitrary but fixed g-inverse $(W+AA^t)^-$ of $W+AA^t$. In view of Theorem 2.1 (ii), clearly $(W+AA^t)^- = (W+AA^t)^{\{\mathcal{T}\}}$ for some $\mathcal{T} \in \mathcal{R}_c(W+AA^t) = \mathcal{R}_c(A,W)$. Observing $\mathcal{N}(W + AA^t) = \mathcal{N}(W)\cap\mathcal{N}(A^t)$ yields $A^t(W+AA^t)^{\{\mathcal{T}\}}\mathcal{T} = \{0\}$ according to Theorem 2.1 (iii). Since $(W+AA^t)^{\{\mathcal{T}\}}(W+AA^t)$ is a projector along $\mathcal{N}(W+AA^t) = \mathcal{N}(W)\cap\mathcal{N}(A^t)$, we further get $A^t(W+AA^t)^{\{\mathcal{T}\}}(W+AA^t) = A^t$ and hence $A^t(W+AA^t)^{\{\mathcal{T}\}}W\mathcal{N}(A^t) = A^t(W+AA^t)^{\{\mathcal{T}\}}(W+AA^t)\mathcal{N}(A^t) = A^t\mathcal{N}(A^t) = \{0\}$. So we arrive at

$$A^t(W+AA^t)^{\{\mathcal{T}\}}[W\mathcal{N}(A^t)\oplus \mathcal{T}] = \{0\}, \tag{2.10}$$

which implies $AH^-A^t(W+AA^t)^{\{\mathcal{T}\}}[W\mathcal{N}(A^t)\oplus \mathcal{T}] = \{0\}$. Since H^-H is a projector along $\mathcal{N}(H) = \mathcal{N}(A)$, we further have $A = AH^-H = AH^-A^t(W+AA^t)^{\{\mathcal{T}\}}A$. Combining all observations results in $AH^-A^t(W+AA^t)^{\{\mathcal{T}\}} = P_{\mathcal{R}(A), W\mathcal{N}(A^t)\oplus \mathcal{T}}$, and the proof is complete. ∎

It is no surprise that somewhat simpler representations for $\mathcal{P}(A \mid W)$ can be obtained under additional assumptions relating A and W. Below we consider only two extreme situations.

Theorem 2.5. For given $W \in \mathcal{P}^{n,n}$ and $A \in \mathbb{R}^{n,m}$, let $\mathcal{P}(A \mid W)$ be defined as before. If $\mathcal{R}(A) \subseteq \mathcal{R}(W)$, then $\mathcal{R}(A,W) = \mathcal{R}(W)$ and

$$\mathcal{P}(A \mid W) = \{A(A^tW^-A)^-A^tW^- \mid W^- \in W\{1\}\}.$$

In particular,

$$A(A^tW^-A)^-A^tW^{\{\mathcal{T}\}} = P_{\mathcal{R}(A), W\mathcal{N}(A^t)\oplus \mathcal{T}},$$

for each $\mathcal{T} \in \mathcal{R}_c(W)$.

Proof: The proof follows similar to the previous one. Since the required modifications are obvious, details are left to the reader. ∎

Theorem 2.6. For given $W \in \mathcal{P}^{n,n}$ and $A \in \mathbb{R}^{n,m}$, let $\mathcal{P}(A \mid W)$ be defined as before. If $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$, that is, if A^t is weakly complementary to W, then $\mathcal{R}(A, W) = \mathcal{R}(A) \oplus \mathcal{R}(W)$ and

$$\mathcal{P}(A \mid W) = \{AA^t(W + AA^t)^- \mid (W + AA^t)^- \in (W + AA^t)\{1\}\}.$$

In particular,

$$AA^t(W + AA^t)^{\{\mathcal{T}\}} = P_{\mathcal{R}(A),\mathcal{R}(W)\oplus\mathcal{T}},$$

for each $\mathcal{T} \in \mathcal{R}_c(A, W)$.

Proof: Let $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$. Then $\mathcal{R}(A, W) = \mathcal{R}(A) \oplus \mathcal{R}(W)$, and it is obvious that W is weakly bicomplementary to AA^t. That $AA^t(W + AA^t)^-$ is a projector onto $\mathcal{R}(AA^t)$ thus follows from Theorem 2.2. Clearly $\mathcal{R}(AA^t) = \mathcal{R}(A)$. Therefore $AA^t(W + AA^t)^- A = A$. Moreover, as in the proof of Theorem 2.4 [see (2.10)], $(W + AA^t)^- = (W + AA^t)^{\{\mathcal{T}\}}$ for some $\mathcal{T} \in \mathcal{R}_c(A, W)$, and $A^t(W + AA^t)^{\{\mathcal{T}\}}[W\mathcal{N}(A^t) \oplus \mathcal{T}] = \{0\}$. Since $W\mathcal{N}(A^t) = \mathcal{R}(W)$, the proof is complete. ∎

We conclude this section with proving in addition to Theorem 2.3 some further interesting invariance properties.

Theorem 2.7. For given $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$, let $\mathcal{P}(A \mid W)$ be defined as before. Then we have:

(i) PW and $(I - P)W$ are invariant for any choice of $P \in \mathcal{P}(A \mid W)$. In particular, $\mathcal{R}(PW) = \mathcal{R}(A) \cap \mathcal{R}(W)$ and $\mathcal{R}((I - P)W) = W\mathcal{N}(A^t)$.

(ii) $WP^tW^-(I - P)W = 0$ for each $P \in \mathcal{P}(A \mid W)$, irrespective of the choice of W^-.

(iii) $PW(I - P)^t = 0$ for each $P \in \mathcal{P}(A \mid W)$.

Proof: Notice that (2.6) implies $\mathcal{R}(W) = [\mathcal{R}(W) \cap \mathcal{R}(A)] \oplus W\mathcal{N}(A^t)$. Therefore $W = T + U$ for some unique matrices T, U with $\mathcal{R}(T) = \mathcal{R}(W) \cap \mathcal{R}(A)$ and $\mathcal{R}(U) = W\mathcal{N}(A^t)$. But then, in view of (2.8), $PW = T$ and $(I - P)W = U$, irrespective of the choice of $P \in \mathcal{P}(A \mid W)$, and the proof of (i) is done. In view of (i), $PW = WZ_1$ for some matrix Z_1, and $(I - P)W = WZ_2$ for some matrix Z_2 with $\mathcal{R}(Z_2) \subseteq \mathcal{N}(A^t)$. Because $\mathcal{N}(A^t)^\perp = \mathcal{R}(A)$ [see (2.1b)], and since $\mathcal{R}(PW) \subseteq \mathcal{R}(A)$, we now also get $WP^tW^-(I-P)W = Z_1^tWW^-WZ_2 = Z_1^tWZ_2 = (PW)^tZ_2 = 0$, that is, we arrive at (ii). In order to prove (iii), observe that $\mathcal{R}((I - P)^t) = \mathcal{N}(A^t)$ follows from (2.8) by means of (2.2) and (2.3). But $W\mathcal{N}(A^t) \subseteq \mathcal{N}(P)$, so that $PW(I - P)^t = 0$ is obvious. This completes the proof. ∎

Theorem 2.8. For given $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$, let $\mathcal{P}(A \mid W)$ and $\mathcal{G}(A \mid W)$ be defined as before. Then we have:

(i) If $\mathcal{R}(A) \subseteq \mathcal{R}(W)$, then $GWG^t \in (A^tW^-A)\{1\}$ for each matrix $G \in \mathcal{G}(A \mid W)$.

(ii) If $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$, then $PW = 0$ for each projector $P \in \mathcal{P}(A \mid W)$.

Proof: Suppose first that $\mathcal{R}(A) \subseteq \mathcal{R}(W)$. Recall that A^tW^-A is invariant for any choice of W^-; see Theorem 2.3. Let $G \in \mathcal{G}(A \mid W)$ be arbitrary but fixed. Then $AG \in \mathcal{P}(A \mid W)$. We have to show that $(A^tW^-A)(GWG^t)(A^tW^-A) = (A^tW^-A)$. Put $P := AG$. By Theorem 2.7 (iii), $PWP^t = PW$. Because of $\mathcal{R}(A) \subseteq \mathcal{R}(W)$ we further have $WW^-A = A$. Trivially $PA = A$. Hence $(A^tW^-A)(GWG^t)(A^tW^-A) = A^tW^-PWP^tW^-A = A^tW^-PWW^-A = A^tW^-PA = A^tW^-A$, and the proof of (i) is complete. Claim (ii) follows directly from Theorem 2.7 (i). ∎

This theorem admits the following corollary.

Corollary 2.9. For given $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$, let $\mathcal{P}(A \mid W)$ and $\mathcal{G}(A \mid W)$ be defined as before. In addition, let A be of full column rank. Then we have:

(i) If $\mathcal{R}(A) \subseteq \mathcal{R}(W)$, then $GWG^t = (A^tW^-A)^{-1}$, irrespective of the choice of $G \in \mathcal{G}(A \mid W)$.

(ii) If $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$, then $GW = 0$ and hence $GWG^t = 0$ for each matrix $G \in \mathcal{G}(A \mid W)$.

Proof: We first consider the case where $\mathcal{R}(A) \subseteq \mathcal{R}(W)$. Since A is of full column rank, $\mathcal{N}(A) = \{0\}$. Therefore, in virtue of Theorem 2.3 (iii), A^tW^-A is nonsingular. But then $(A^tW^-A)\{1\} = \{(A^tW^-A)^{-1}\}$, and (i) follows from Theorem 2.8 (i). Next, let $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$. Since $\operatorname{rank}(A) = m$, clearly $AGW = 0$ iff $GW = 0$. As a consequence of Theorem 2.8 (ii) we now get (ii). ∎

Theorem 2.10. For given $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$, let $\mathcal{P}(A \mid W)$ and $\mathcal{G}(A \mid W)$ be defined as before. In addition, let A be of full column rank. Then we have:

(i) GW is invariant for any choice of $G \in \mathcal{G}(A \mid W)$.

(ii) GWG^t is invariant for any choice of $G \in \mathcal{G}(A \mid W)$.

(iii) For each $G \in \mathcal{G}(A \mid W)$: $GWG^t \in \mathcal{P}^{m,m}$, $\mathcal{N}(GWG^t) = A^t\mathcal{N}(W)$, and $\mathcal{R}(GWG^t) = \{x \mid Ax \in \mathcal{R}(A) \cap \mathcal{R}(W)\}$.

(iv) For each $G \in \mathcal{G}(A \mid W)$, GWG^t is nonsingular iff $\mathcal{R}(A) \subseteq \mathcal{R}(W)$.

(v) For each $G \in \mathcal{G}(A \mid W)$, $GWG^t = 0$ iff $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$.

Proof: From Theorem 2.7 (i) we get (i) because A is of full column rank. Writing $GWG^t = G(GW)^t$ shows that (ii) follows from (i). For proving (iii), let $G \in \mathcal{G}(A \mid W)$ be arbitrary but fixed. Put $P := AG$. Since W is nnds, trivially $GWG^t \in \mathcal{P}^{m,m}$ and $\mathcal{N}(GWG^t) = \mathcal{N}(WG^t)$. But $\mathcal{N}(WG^t) = A^t\mathcal{N}(WG^tA^t) = A^t\mathcal{N}(WP^t)$. In view of (2.1a), $\mathcal{N}(WP^t) = \mathcal{R}(PW)^\perp$. By Theorem 2.7 (i), $\mathcal{R}(PW) = \mathcal{R}(A) \cap \mathcal{R}(W)$. Consequently $\mathcal{N}(WP^t) = \mathcal{N}(A^t) + \mathcal{N}(W)$. Combining observations now results in $\mathcal{N}(GWG^t) = A^t\mathcal{N}(W)$. Since A is of full column rank, we further get $\mathcal{R}(GW) = \{x \mid Ax \in \mathcal{R}(A) \cap \mathcal{R}(W)\}$ directly from $\mathcal{R}(PW) = \mathcal{R}(A) \cap \mathcal{R}(W)$. Observing that $\mathcal{R}(GWG^t) = \mathcal{R}(GW)$ completes the proof of (iii). To establish (iv) observe that, in view of (iii), GWG^t is nonsingular iff $A^t\mathcal{N}(W) = \{0\}$, that is, iff $\mathcal{N}(W) \subseteq \mathcal{N}(A^t)$. Since the latter condition is equivalent to $\mathcal{R}(A) \subseteq \mathcal{R}(W)$ [note (2.1)]), the desired result emerges. That $\mathcal{R}(A) \cap \mathcal{R}(W)$ is a sufficient condition for $GWG^t = 0$ to hold is the result of Corollary 2.9 (ii). Necessity is seen as follows. Let $GWG^t = 0$. Then $\mathcal{N}(GWG^t) = \mathbb{R}^m$ and hence, by (iii), $A^t\mathcal{N}(W) = \mathbb{R}^m$ or, equivalently, $\mathcal{N}(W) + \mathcal{N}(A^t) = \mathbb{R}^n$. This condition, however, is equivalent to $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$. ∎

Theorem 2.11. For given $A \in \mathbb{R}^{n,m}$ and $W \in \mathcal{P}^{n,n}$, let $\mathcal{G}(A \mid W)$ be defined as before. Further let A be of full column rank, and put $H := A^t(W + AA^t)^-A$. Then we have:

(i) $GWG^t = H^{-1} - I$, irrespective of the choice of $G \in \mathcal{G}(A \mid W)$.

(ii) $\mathcal{R}(A) \subseteq \mathcal{R}(W)$ iff $H^{-1} - I$ is nonsingular. In which case

$$H^{-1} - (A^t W^- A)^{-1} = I. \tag{2.11}$$

(iii) $\mathcal{R}(A) \cap \mathcal{R}(W) = \{0\}$ iff $H = I$.

Proof: That GWG^t is invariant for any choice of $G \in \mathcal{G}(A \mid W)$ is the statement of Theorem 2.10 (ii). Because $\mathcal{R}(A) \subseteq \mathcal{R}(W + AA^t)$, and since A is of full column rank, it follows from Theorem 2.3 that H is nonsingular, whence we get $H^{-1}A^t(W + AA^t)^\dagger \in \mathcal{G}(A \mid W)$ in view of Theorem 2.4 and Theorem 2.1. Making use of Theorem 2.1 and Theorem 2.3, we now obtain $GWG^t = H^{-1}A^t(W+AA^t)^\dagger W(W+AA^t)^\dagger AH^{-1} = H^{-1}A^t(W+AA^t)^\dagger[(W+AA^t)-AA^t](W+AA^t)^\dagger AH^{-1} = H^{-1}(H - H^2)H^{-1} = H^{-1} - I$, and the proof of (i) is complete. Combining (i) with Corollary 2.9 (i) and Theorem 2.10 (iv) results in (ii). Finally we get (iii) from (i) and Theorem 2.10 (v). ∎

3. ICGLS-Problem: General Model

In this section we consider the restricted model $\mathcal{L}_r$ [see (1.1)] under the assumption that

$$\mathcal{R}(X) \subseteq \mathcal{R}(V). \tag{1.2}$$

Throughout we further assume that the matrix R is of full row rank. As preannounced, our aim here is to exhibit how the ICGLS selections for β under this model are related to the GLS selections for β under the associated unconstrained model $\mathcal{L}_u$ [see (1.3)].

It is pertinent to begin this section with characterizing the feasible solutions of the ICGLS optimization program (1.5).

Theorem 3.1. Let $R \in \mathbb{R}^{p,m}$ be of rank p and let $r \in \mathbb{R}^p$. Then b is a solution of $Rb \geq r$ iff

$$Rb = r + \mu \quad \text{for some} \quad \mu \geq 0 \tag{3.1}$$

or, equivalently,

$$b = R^-(r + \mu) + z \quad \text{for some pair} \quad \mu \geq 0,\ z \in \mathcal{N}(R). \tag{3.2}$$

Since Theorem 3.1 can be established similar to Lemma 3.1 in [29], its proof is omitted. It should be emphasized that the previous characterizations rely heavily on the assumption that R is of full row rank. For recall that this rank condition is necessary and sufficient for $RR^- = I$ to hold; see Theorem 2.1 (iii). If, on the other hand, R is not of full row rank, then condition (3.2), although still necessary, need not be sufficient for b to be a solution of $Rb \geq r$. A set of necessary as well as sufficient conditions in this more general situation consists of (3.2) and

$$RR^-(r + \mu) = r + \mu. \tag{3.3}$$

If R is of full row rank, then condition (3.3) is fortunately automatically satisfied, so that we have the possibilty to represent the general feasible solution of (1.5) in the form (3.2), where μ is *free to vary* in $\mathbb{R}^p$ and where z is free to vary in $\mathcal{N}(R)$. In context with Theorem 3.1, it is also worth mentioning that if R is of full row rank, then b satisfies

$$Rb = r \quad \text{iff} \quad b = R^-r + z \quad \text{for some vector} \quad z \in \mathcal{N}(R). \tag{3.4}$$

Note that characterization (3.4) remains true even when R is not of full row rank provided $r \in \mathcal{R}(R)$; cf. [2].

For convenience, we introduce the matrix

$$\Gamma := X^t V^- X. \tag{3.5}$$

Because of $\mathcal{R}(X) \subseteq \mathcal{R}(V)$ it follows from Theorem 2.3 that Γ is a nonnegative definite and symmetric matrix, is invariant for any choice of $V^- \in V\{1\}$, and satisfies

$$\mathcal{R}(\Gamma) = \mathcal{R}(X^t) \quad \text{as well as} \quad \mathcal{N}(\Gamma) = \mathcal{N}(X). \tag{3.6a-b}$$

Since V and Γ are nnds, we further know from Section 2 that

$$\mathcal{P}(X \mid V) := \{P_{\mathcal{R}(X), V\mathcal{N}(X^t)\oplus\mathcal{T}} \mid \mathcal{T} \in \mathcal{R}_c(X, V)\} \tag{3.7}$$

and

$$\mathcal{P}(R^t \mid \Gamma) := \{P_{\mathcal{R}(R^t), \Gamma\mathcal{N}(R)\oplus\mathcal{U}} \mid \mathcal{U} \in \mathcal{R}_c(R^t, \Gamma)\} \tag{3.8}$$

are well-defined classes of (generally oblique) projectors. In the sequel, let $X^P \in \mathcal{G}(X \mid V)$ and $(R^Q)^t \in \mathcal{G}(R^t \mid \Gamma)$ be arbitrary but fixed matrices. Then we have

$$P := XX^P \in \mathcal{P}(X \mid V) \quad \text{and} \quad Q^t := (R^Q R)^t \in \mathcal{P}(R^t \mid \Gamma). \tag{3.9}$$

Define the matrix

$$\Omega := (R^Q)^t \Gamma R^Q. \tag{3.10}$$

As R^t is of full column rank, it follows from Theorem 2.10 that Ω is not only nnds but in addition even invariant for any choice of $(R^Q)^t \in \mathcal{G}(R^t \mid \Gamma)$. On this occasion we mention already here that

$$\mathcal{N}(\Omega) = R\mathcal{N}(X); \tag{3.11}$$

this identity is a consequence of Theorem 2.10 (iii) by observing (3.6b).

On account of Theorem 2.7 (ii), $VP^tV^-(I-P)V = 0$. Trivially $PX = X$ and hence $(I-P)X = 0$. In view of $\mathcal{R}(X) \subseteq \mathcal{R}(V)$ [see (1.2)], we therefore get for each realization $y \in \mathcal{R}(V)$:

$$\begin{aligned} \|y - Xb\|^2_{V^-} &= \|P(y - Xb) + (I-P)(y - Xb)\|^2_{V^-} \\ &= \|P(y - Xb)\|^2_{V^-} + \|(I-P)(y - Xb)\|^2_{V^-} \\ &= \|Py - Xb\|^2_{V^-} + \|(I-P)y\|^2_{V^-} \\ &= \|X(X^P y - b)\|^2_{V^-} + \|(I-P)y\|^2_{V^-} \\ &= \|X^P y - b\|^2_{\Gamma} + \|(I-P)y\|^2_{V^-}. \end{aligned} \tag{3.12}$$

From (3.12) it is seen that the set of ordinary GLS-solutions for β under the unconstrained model $\mathcal{L}_u$, that is, $\{\hat{\beta}_u(y)\} := \operatorname{argmin}_{b\in\mathbb{R}^m} \|y - Xb\|^2_{V^-}$, is representable in the form

$$\{\hat{\beta}_u(y)\} = \{X^P y + P_{\mathcal{N}(X)} w(y) \mid w(y) \text{ arbitrary}\} \tag{3.13}$$

(compare [22]); notice (3.6b). In what follows, let $\hat{\beta}_u = \hat{\beta}_u(y)$ denote an arbitrary but fixed GLS solution for β under model $\mathcal{L}_u$. Put

$$\hat{\mu} = \hat{\mu}(\hat{\beta}_u) := R\hat{\beta}_u - r. \tag{3.14}$$

In the decomposition given below observe that according to Theorem 2.7 (iii), $Q^t\Gamma(I-Q) = 0$. Also recall that R^Q is a particular g-inverse of R, so that $I - R^Q R$ is a projector onto $\mathcal{N}(R)$. In addition, notice that $RR^Q = I$; see Theorem 2.1 (iii). By considering only feasible solutions to (1.5), we can express b alternatively in the form (3.2) where R^- can be replaced by R^Q. Doing so, (3.12) can be rewritten step by step as follows:

$$\begin{aligned}
||y - Xb||^2_{V^-} &= ||\hat{\beta}_u - b||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&= ||Q(\hat{\beta}_u - b) + (I-Q)(\hat{\beta}_u - b)||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&= ||Q(\hat{\beta}_u - b)||^2_\Gamma + ||(I-Q)(\hat{\beta}_u - b)||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&= ||R^Q(R\hat{\beta}_u - Rb)||^2_\Gamma + ||(I - R^Q R)(\hat{\beta}_u - b)||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&= ||R\hat{\beta}_u - Rb||^2_\Omega + ||(I - R^Q R)(\hat{\beta}_u - b)||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&= ||R\hat{\beta}_u - r - \mu||^2_\Omega + ||(I - R^Q R)\hat{\beta}_u - z||^2_\Gamma + ||(I-P)y||^2_{V^-} \\
&\geq ||\hat{\mu} - \tilde{\mu}||^2_\Omega + ||(I-P)y||^2_{V^-},
\end{aligned} \tag{3.15}$$

where $\tilde{\mu} = \tilde{\mu}(y)$ denotes any vector contained in

$$\tilde{M}(y) := \text{argmin}_{\mu \geq 0}\ ||\hat{\mu} - \mu||^2_\Omega. \tag{3.16}$$

Note that 0 is the optimal value of program

$$\min_{z \in \mathcal{N}(R)}\ ||(I - R^Q R)\hat{\beta}_u - z||^2_\Gamma. \tag{3.17}$$

As Ω is nnds, the objective function of the optimization program

$$\min_{\mu \geq 0}\ ||\hat{\mu} - \mu||^2_\Omega \tag{3.18}$$

is convex-quadratic. The feasible region of program (3.18) is the closed convex nonnegative orthant. (3.18) is thus a very particular convex program. That (3.18) has at least one optimal solution can now easily be seen by means of the Weierstrass theorem; cf. [4]; see also the proof of Theorem 3.6. In other words, $\tilde{M}(y)$ is nonempty.

It is interesting to mention here that $\tilde{M}(y)$ is always invariant for any choice of $\hat{\beta}_u \in \{\hat{\beta}_u(y)\}$. That the objective function of program (3.18) has this property follows by means of (3.11) and (3.13). It is evident that $\tilde{M}(y)$ inherits this property. Using (3.6b), (3.11) and (3.13), it can be seen that the program $\min_{z \in \mathcal{N}(R)}\ ||(I - R^Q R)\hat{\beta}_u - z||^2_\Gamma$ possesses the same property. Evidently

$$\text{argmin}_{z \in \mathcal{N}(R)}\ ||(I - R^Q R)\hat{\beta}_u - z|| = \{(I - R^Q R)\hat{\beta}_u + P_{\mathcal{N}(R)\cap\mathcal{N}(X)} w(y) \mid w(y) \text{ arbitrary}\}, \tag{3.19}$$

and it is clear that this set is also invariant for any choice of $\hat{\beta}_u$.

The following theorem relating the set $\{\hat{\beta}_u(\cdot)\}$ of GLS-selections under model $\mathcal{L}_u$ to the set $\{\tilde{\beta}(\cdot)\}$ of ICGLS-selections under model $\mathcal{L}_r$ is now an easy consequence of all these observations.

Clearly, instead of minimizing (1.5) subject to $\{b \mid Rb \geq r\}$ we can just as well determine (3.16) and (3.19); (3.2) does tell us how the optimal solutions of program (1.5) are related to the optimal solutions of the auxiliary programs (3.18) and (3.17).

Theorem 3.2. Consider model $\mathcal{L}_r$ under (1.2), and let R be of full row rank. Further, let Γ be defined by (3.5), and let $\hat{\beta}_u$ denote an arbitrary but fixed GLS selection for β under model $\mathcal{L}_u$. Finally, let X^P and R^Q, respectively, stand for an arbitrary but fixed g-inverse of X and R with $XX^P \in \mathcal{P}(X \mid V)$ and $(R^Q R)^t \in \mathcal{P}(R^t \mid \Gamma)$. The complete set of ICGLS selections for β under model $\mathcal{L}_r$ can then be represented as

$$\{\tilde{\beta}(\cdot)\} = \{R^Q(r + \tilde{\mu}) + (I - R^Q R)\hat{\beta}_u + P_{\mathcal{N}(R)\cap\mathcal{N}(X)} w(\cdot) \mid \tilde{\mu} \in \tilde{M}(\cdot),\ w(\cdot) \text{ arbitrary}\}, \quad (3.20)$$

where $\tilde{M}(\cdot)$ is pointwise defined by (3.16). The set of GLS selections for β under model $\mathcal{L}_u$ is given by

$$\{\hat{\beta}_u(\cdot)\} = \{X^P y + P_{\mathcal{N}(X)} w(\cdot) \mid w(\cdot) \text{ arbitrary}\}. \quad (3.21)$$

In context with Theorem 3.2 it is pertinent to mention that the ECGLS (equality constrained GLS) selections for β under the related model

$$\mathcal{L}_e := (y,\ X\beta,\ V \mid R\beta = r) \quad (3.22)$$

can be represented similarly.

Theorem 3.3. Consider model $\mathcal{L}_e$ under (1.2), and let R be of full row rank. Further, let Γ, R^Q, and $\hat{\beta}_u$ be as in the preceding theorem. The complete set $\{\hat{\beta}_e(\cdot)\}$ of ECGLS selections for β under model $\mathcal{L}_e$ can then be represented as

$$\{\hat{\beta}_e(\cdot)\} = \{R^Q r + (I - R^Q R)\hat{\beta}_u + P_{\mathcal{N}(X)\cap\mathcal{N}(R)} w(\cdot) \mid w(\cdot) \text{ arbitrary}\}. \quad (3.23)$$

Proof: Using characterization (3.4), this result can be established step by step similar to the preceding theorem; the required modifications are obvious. ∎

We mention that representation (3.23) remains true even when R is not of full row rank provided $r \in \mathcal{R}(R)$. Combining (3.23) with (3.20) results in the following alternative representation for $\{\tilde{\beta}(\cdot)\}$.

Theorem 3.4. Consider model $\mathcal{L}_r$ under (1.2), and let R be of full row rank. Further, let R^Q and $\tilde{M}(\cdot)$ be as in Theorem 3.2. Finally, let $\hat{\beta}_e$ stand for an arbitrary but fixed ECGLS selection for β under the associated model $\mathcal{L}_e$. Then

$$\{\tilde{\beta}(\cdot)\} = \{R^Q \tilde{\mu} + \hat{\beta}_e + P_{\mathcal{N}(X)\cap\mathcal{N}(R)} w(\cdot) \mid \tilde{\mu} \in \tilde{M}(\cdot),\ w(\cdot) \text{ arbitrary}\}. \quad (3.24)$$

When glancing at (3.24) one could be tempted to believe that, contrary to (3.20), in this alternative representation it is not necessary to know a GLS selection $\hat{\beta}_u$. Unfortunately, however, this is not the case because for determining $\tilde{M}(\cdot)$ we still need a selection $\hat{\mu} = R\hat{\beta}_u - r$.

From Theorem 3.2 we single out the following corollary. Observing $RR^Q = I$ the claims are straightforward consequences of (3.20) and (3.24).

Corollary 3.5. Consider the models $\mathcal{L}_r$, $\mathcal{L}_u$ and $\mathcal{L}_e$ under (1.2), and let R be of full row rank. Denote the selections for β under the different models as before. Further, let R^Q be such that $(R^Q R)^t \in \mathcal{P}(R^t \mid \Gamma)$, and let $\tilde{M}(\cdot)$ be pointwise defined according to (3.16). Then we have

$$R\tilde{\beta} - r \in \tilde{M}(\cdot) \tag{3.25}$$

for each ICGLS selection $\tilde{\beta}$. In addition we can find for each mixed pair of selections, $(\tilde{\beta}, \hat{\beta}_u)$, a selection $w(\cdot)$ such that

$$(I - R^Q R)(\tilde{\beta} - \hat{\beta}_u) = P_{\mathcal{N}(X)\cap\mathcal{N}(R)}w(\cdot). \tag{3.26}$$

Therefore $(I - R^Q R)\tilde{\beta} = (I - R^Q R)\hat{\beta}_u$, irrespective of the choice of $(\tilde{\beta}, \hat{\beta}_u)$, iff

$$\mathcal{N}(X) \cap \mathcal{N}(R) = \{0\} \quad \text{or, equivalently,} \quad \mathcal{R}(X^t) + \mathcal{R}(R^t) = \mathbb{R}^m. \tag{3.27}$$

In which case, $(I - R^Q R)\tilde{\beta}$ is invariant for any choice of $\tilde{\beta}$. Analogous results are obtained by replacing the GLS selections $\hat{\beta}_u$ by ECGLS selections $\hat{\beta}_e$.

Note that $C\beta$ is linearly unbiasedly estimable in model $\mathcal{L}_e$ iff $\mathcal{R}(C^t) \subseteq \mathcal{R}(X^t) + \mathcal{R}(R^t)$; cf. [19], [22], [23]. Condition (3.27) therefore means that all parametric functions of the form $C\beta$ are linearly unbiasedly estimable in model $\mathcal{L}_e$. It is interesting to mention here that $(I - R^Q R)\hat{\beta}_u = (I - R^Q R)\hat{\beta}_e$ also holds, irrespective of the choice of $(\hat{\beta}_u, \hat{\beta}_e)$, iff $\mathcal{N}(X) \cap \mathcal{N}(R) = \{0\}$. In general, however, we only have $(I - R^Q R)\{\tilde{\beta}(\cdot)\} = (I - R^Q R)\{\hat{\beta}_e(\cdot)\} = (I - R^Q R)\{\hat{\beta}_u(\cdot)\}$; compare (3.20), (3.24) and (3.22).

Next, let us ask for a necessary and/or sufficient condition under which $\tilde{M}(\cdot)$ is a singleton. The answer is given in our next theorem.

Theorem 3.6. Consider model $\mathcal{L}_r$ under (1.2), and let R be of full row rank. Further, let $\tilde{M}(\cdot)$ be pointwise defined according to (3.16), where Ω is given by (3.10). The following conditions are then equivalent:

(i) $\mathcal{R}(R^t) \subseteq \mathcal{R}(X^t)$, this is inclusion (1.7);

(ii) Ω is nonsingular;

(iii) $\tilde{M}(\cdot)$ is a singleton.

In which case

$$R\tilde{\beta}(\cdot) - r = \tilde{\mu}(\cdot) \tag{3.28}$$

holds for each ICGLS selection $\tilde{\beta}(\cdot)$; $\tilde{\mu}(\cdot)$ stands for the unique element of $\tilde{M}(\cdot)$.

Proof: Suppose first that $\mathcal{R}(R^t) \subseteq \mathcal{R}(X^t)$. Since $\mathcal{N}(\Omega) = R\mathcal{N}(X)$ [see (3.11)], Ω is nonsingular iff $\mathcal{N}(X) \subseteq \mathcal{N}(R)$. But this condition is equivalent to $\mathcal{R}(R^t) \subseteq \mathcal{R}(R^t)$, and so we have (i)$\Leftrightarrow$(ii). Next, let Ω be nonsingular, and consider program (3.18) for an arbitrary but fixed realization $y \in \mathcal{R}(V)$. Theorem 2.10 then tells us that Ω is a positive definite matrix. The objective function of program (3.18) is hence not only convex-quadratic but even *strictly convex*; cf. [4]. Also observe that the feasible region of this program is a closed convex set. Programs with a convex objective function and a convex feasible region are called convex programs. Such programs have amazing properties. One property is that each local optimal solution is a global optimal solution. A further property is that there is at most one optimal solution whenever the objective function is strictly convex. Since from the lines directly following (3.18) we already know that (3.18) has at least one optimal solution, it is now clear that $\tilde{M}(y)$ is a singleton for each given $y \in \mathcal{R}(V)$. But then $\tilde{M}(\cdot)$ is a singleton, and the proof of (ii)⇒(iii) is complete. For proving the converse, let us suppose that $\tilde{M}(\cdot)$ is a singleton but Ω is singular. Of course, $\tilde{M}(\cdot)$ can be a singleton only if $\tilde{M}(y)$ is a singleton for each choice of $y \in \mathcal{R}(V)$. Let $v \neq 0$ be such that $\Omega v = 0$; the existence of such a vector is guaranteed by the singularity of Ω. Further, let X^P be as in (3.21). Since X^P is a g-inverse of X, $\mathcal{R}(X^P X) \in \mathcal{N}_c(X)$, that is, $\mathcal{R}(X^P X) \oplus \mathcal{N}(X) = \mathbb{R}^m$. In view of (1.2), therefore $\mathbb{R}^m = \mathcal{R}(X^P V) + \mathcal{N}(X)$, whence we get

$$\bigcup_{y \in \mathcal{R}(V)} \{\hat{\beta}_u(y)\} = \mathbb{R}^m$$

by means of (3.21). Since R is of full row rank, it is now evident that we can find a vector in $\mathcal{R}(V)$, say y_0, such that $\hat{\mu} := R\hat{\beta}_u - r$ is componentwise positive for some GLS solution $\hat{\beta}_u \in \{\hat{\beta}_u(y_0)\}$. Then trivially $\hat{\mu} \in \tilde{M}(y_0)$. But in addition we also have $\hat{\mu} + \lambda v \in \tilde{M}(y_0)$ for some $\lambda \neq 0$, a contradiction. This completes the proof. ∎

Remark 3.7. Consider model $\mathcal{L}_r$ under (1.2), and let R of full row rank be such that inclusion (1.7) or, equivalently, $\mathcal{N}(X) \subseteq \mathcal{N}(R)$ is satisfied. Then we know that the parametric function $R\beta$ is linearly unbiasedly estimable under $\mathcal{L}_u$. From the famous (generalized) Gauß-Markov Theorem (cf. [22]) it then follows that $R\hat{\beta}_u$, being unique for all GLS selections $\hat{\beta}_u$, is the BLUE of $R\beta$. So its expectation is $R\beta$ [see also (3.13)], and it is clear that $\hat{\mu} = R\hat{\beta}_u - r$ is unique and has expectation $\mu := R\beta - r$. It further follows from Theorem 2.8 (i) that the dispersion of $X^P y$ is a particular fixed g-inverse of Γ, say Γ^-, where Γ is defined by (3.5). In view of (3.21) the dispersion of $R\hat{\beta}_u - r$ therefore coincides with $R\Gamma^- R^t$. It is interesting to remark that this expression, in view of (1.7) and (3.6a), is even invariant for any choice of $\Gamma^- \in \Gamma\{1\}$; this follows by means of Theorem 2.3 (i). Moreover, since R^t is of full column rank, $R\Gamma^- R^t$ is nonsingular, in virtue of Theorem 2.3 (iii). Finally observe that β satisfies the inequality constraints $R\beta \geq r$ iff $\mu \geq 0$. Therefore it is natural to compare model $\mathcal{L}_r$ with the transformed model

$$\mathcal{L}_t := (\hat{\mu},\ \mu,\ R\Gamma^- R^t \mid \mu \geq 0). \tag{3.29}$$

From Corollary 2.9 (i) we know that

$$\Omega := (R^Q)^t \Gamma R^Q = (R\Gamma^- R^t)^{-1} \tag{3.30}$$

holds, irrespective of the choice of R^Q. Because of (3.10) it is now clear that the set of ICGLS selections for μ under the transformed model $\mathcal{L}_t$ coincides with the set $\tilde{M}(\cdot)$ introduced earlier. In other words, when considering model $\mathcal{L}_r$ under (1.7), then it turns out that the set $\tilde{M}(\cdot)$ is nothing else but the complete set of ICGLS selections for μ in the transformed model $\mathcal{L}_t$. At this point it should be emphasized once more that (3.25) tells us something about how this set is related to the set of ICGLS selections for β in the original model $\mathcal{L}_r$. We mention that (3.28) has already been obtained recently by Firoozi [7]. He used a different approach and considered only the full rank case, that is, the case where V and X are both of full column rank. Whereas Firoozi, in the framework of his particular model, was primarily interested in finding relation

(3.28), our concern here goes beyond that aim. Note that our projector theoretical approach detects in addition to (3.25) the more informative relation (3.20) [or (3.32)] that shows how an ICGLS selection for β in (our more general) model $\mathcal{L}_r$ can be constructed from $\tilde{\mu} \in \tilde{M}(\cdot)$ and $\hat{\beta}_u \in \{\hat{\beta}_u(\cdot)\}$.

The following result is a consequence of (3.20) and Theorem 3.6. It gives a necessary and sufficient condition for the uniqueness of ICGLS selections.

Theorem 3.8. Consider model $\mathcal{L}_r$ under (1.2), and let R be of full row rank. Then $\{\tilde{\beta}(\cdot)\}$ is a singleton iff X is of full column rank. In which case the ICGLS estimator of β exists.

Theorem 3.6 dealt with model $\mathcal{L}_r$ under the additional assumption that $\mathcal{R}(R^t) \subseteq \mathcal{R}(X^t)$ is satisfied. Our next theorem is devoted to the other extreme model case where the range space of R^t has only the origin in common with the range space of X^t.

Theorem 3.9. Consider model $\mathcal{L}_r$ under (1.2), and let R be of full row rank. Further, let $\tilde{M}(\cdot)$ be pointwise defined according to (3.16), where Ω is given by (3.10). The following conditions are then equivalent:

(i) $\mathcal{R}(R^t) \cap \mathcal{R}(X^t) = \{0\}$, this is (1.8);

(ii) $\Omega = 0$;

(iii) $\tilde{M}(y) = \{\mu \mid \mu \geq 0\}$ for each $y \in \mathcal{R}(V)$.

In which case we have

$$\{\hat{\beta}_e(\cdot)\} \subseteq \{\tilde{\beta}(\cdot)\} \subseteq \{\hat{\beta}_u(\cdot)\}, \tag{3.31}$$

where $\{\hat{\beta}_u(\cdot)\}$, $\{\hat{\beta}_e(\cdot)\}$, and $\{\tilde{\beta}(\cdot)\}$ are defined as before. All inclusions in (3.31) are proper.

Proof: Because R is of full row rank, and since by (3.11) $\mathcal{N}(\Omega) = R\mathcal{N}(X)$, we get $\Omega = 0 \;\Leftrightarrow\; \mathcal{N}(\Omega) = \mathbb{R}^p \;\Leftrightarrow\; R\mathcal{N}(X) = \mathbb{R}^p \;\Leftrightarrow\; \mathcal{N}(X) + \mathcal{N}(R) = \mathbb{R}^m \;\Leftrightarrow\; \mathcal{R}(X^t) \cap \mathcal{R}(R^t) = \{0\}$. So we have (i)$\Leftrightarrow$(ii) [see also Theorem 2.11]. That (ii) implies (iii) is obvious. If (iii) holds, then necessarily $\Omega\mu = 0$ for each $\mu \geq 0$ [compare also (3.37)]. But his happens only if $\Omega = 0$, and so we have (ii). Inclusion $\{\hat{\beta}_e(\cdot)\} \subseteq \{\tilde{\beta}(\cdot)\}$ follows directly from (3.20) and (3.21) because $0 \in \tilde{M}(\cdot)$ [see (iii)]. To establish $\{\tilde{\beta}(\cdot)\} \subseteq \{\hat{\beta}_u(\cdot)\}$, rewrite the general ICGLS selection for β given in (3.20) as follows:

$$\tilde{\beta}(\cdot) = \hat{\beta}_u - R^Q(\hat{\mu} - \tilde{\mu}) + P_{\mathcal{N}(R)\cap\mathcal{N}(X)} w(\cdot), \;\; \tilde{\mu} \in \tilde{M}(\cdot), \;\; w(\cdot) \text{ arbitrary}, \tag{3.32}$$

where $\hat{\beta}_u$ is an arbitrary but fixed GLS selection for β under $\mathcal{L}_u$ and $\hat{\mu} := R\hat{\beta}_u - r$. Notice that (1.8) is equivalent to $\Gamma\mathcal{N}(R) = \mathcal{R}(X^t)$ because of (3.6). Since $R^t(R^Q)^t \in \mathcal{P}(R^t \mid \Gamma)$ [see (3.9) and (3.8)], clearly $\Gamma\mathcal{N}(R) = \mathcal{R}(X^t)$ iff $R^t(R^Q)^t X^t = 0$. But the latter condition is equivalent to $(R^Q)^t X^t = 0$ as R^t is of full column rank. Therefore $XR^Q = 0$ or, equivalently, $\mathcal{R}(R^Q) \subseteq \mathcal{N}(X)$. With this in mind, the desired inclusion follows by comparing (3.32) with (3.21). ∎

In the previous proof we have actually deduced a little bit more. Namely, that (1.8) is not only a sufficient but also a necessary condition for $\{\tilde{\beta}(\cdot)\} \subseteq \{\hat{\beta}_u(\cdot)\}$ or $\{\hat{\beta}_e(\cdot)\} \subseteq \{\hat{\beta}_u(\cdot)\}$ to hold. This means that whenever linear inequality constraints (or linear equality constraints) are *artificially* introduced to $\mathcal{L}_u$ for the sole purpose of *reducing* the set of GLS selections $\{\hat{\beta}_u(\cdot)\}$, then it is indeed *essential* that the restrictor matrix R is weakly complementary to the regressor matrix X; compare Section 1.

Next we consider model $\mathcal{L}_r$ under (1.2), and we assume that R is of full row rank. Procedure 3.10 for computing an ICGLS solution for β under model $\mathcal{L}_r$ from a GLS solution for β under model $\mathcal{L}_u$ [given a realization $y \in \mathcal{R}(V)$] follows directly from representation (3.32). In this context recall that $X^P y$ is a particular GLS solution for β under model $\mathcal{L}_u$; see (3.21). It is also pertinent to mention here that we know from Theorem 2.5 and Theorem 2.4, respectively, that X^P and R^Q can be chosen as

$$X^P = \Gamma^- X^t V^- \quad \text{and} \quad R^Q = (\Gamma + R^t R)^- R^t \left[R(\Gamma + R^t R)^- R^t\right]^{-1},$$

where $\Gamma = X^t V^- X$. Needless to say that Moore-Penrose inversions can be used in these expressions. Finally notice that in our general model framework it follows from Theorem 2.11 that Ω being defined in (3.10) can be written as

$$\Omega = \left[R(\Gamma + R^t R)^- R^t\right]^{-1} - I, \tag{3.33}$$

that is, in terms of the model matrices X, V and R. With all these observations in mind, Procedure 3.10 will be easy to understand.

Procedure 3.10 *(Method for computing an ICGLS solution for β in model $\mathcal{L}_r$ under (1.2)).*

STEP 1. Compute an ordinary GLS solution $\hat\beta_u := \hat\beta_u(y)$ in the associated unconstrained model $\mathcal{L}_u$. Compute $\hat\mu := R\hat\beta_u - r$. If $\hat\mu \geq 0$, then

$$\hat\beta_u(y) \in \{\tilde\beta(y)\}.$$

Else, GOTO STEP 2.

STEP 2. Compute $\Gamma = X^t V^- X$ and $\Omega = [R(\Gamma + R^t R)^- R^t]^{-1} - I$. Determine any vector $\tilde\mu$ such that

$$\tilde\mu \in \tilde M(y) := \operatorname{argmin}_{\mu \geq 0} \|\hat\mu - \mu\|^2_\Omega.$$

Compute $R^Q = (\Gamma + R^t R)^- R^t [R(\Gamma + R^t R)^- R^t]^{-1}$. Then

$$\hat\beta_u(y) - R^Q(\hat\mu - \tilde\mu) \in \{\tilde\beta(y)\}.$$

Evidently the crucial point when using this procedure is STEP 2 where we have to determine an optimal solution of the convex-quadratic auxiliary program (3.18). Of course, when Ω is nonsingular, which according to Theorem 3.6 is the case iff $\mathcal{R}(R^t) \subseteq \mathcal{R}(X^t)$, then we can apply Procedure 4.1 in [29] for determining the unique optimal solution of (3.18). In order to generalize that procedure to situations where Ω is singular, it is pertinent to proceed to characterizing the optimal solutions of program (3.18). To this end we need some further notation as well as a theorem representing the complete set of g-inverses for a particular block partitioned matrix.

For given $\Omega \in \mathcal{P}^{p,p}$ and $\Delta \subseteq \{1, 2, \cdots, p\}$, we define the matix Ω_Δ according to

$$(\Omega_\Delta)_i := \begin{cases} \Omega_i & \text{if } i \notin \Delta, \\ -e_i & \text{if } i \in \Delta, \end{cases}$$

where $(\Omega_\Delta)_i$, Ω_i, and e_i denote, respectively, the ith column vector of Ω_Δ, the ith column vector of Ω, and the ith standard unit vector of $\mathbb{R}^p$.

Theorem 3.11. Consider the following block-partitioned matrix

$$A = \begin{pmatrix} -I & B \\ 0 & C \end{pmatrix},$$

where C and B are such that $\mathcal{R}(B^t) \subseteq \mathcal{R}(C^t)$ or, equivalently, $\mathcal{N}(C) \subseteq \mathcal{N}(B)$ is satisfied. Then $G \in A\{1\}$ iff

$$G = \begin{pmatrix} -I & BC^- + U(I - CC^-) \\ (I - C^-C)Z & C^- \end{pmatrix}$$

for some $C^- \in C\{1\}$ and for some conformable matrices U and Z.

Proof: Let

$$G = \begin{pmatrix} E & F \\ M & N \end{pmatrix}$$

be such that AGA and EIE do exist. Using $\mathcal{N}(C) \subseteq \mathcal{N}(B)$ or, equivalently, $B(I - C^-C) = 0$, it is not difficult to see that $AGA = A$ does hold iff the subblocks in G are as claimed. ∎

Theorem 3.12. Consider the convex-quadratic program $\min_{\mu \geq 0} \|\hat{\mu} - \mu\|^2_{\Omega}$, where $\Omega \in \mathcal{P}^{p,p}$ and $\hat{\mu} \in \mathbb{R}^p$ are given.

(i) If $\Omega\hat{\mu} \leq 0$, then $\tilde{\mu}$ is an optimal solution iff $\tilde{\mu} \geq 0$ and $\tilde{\mu} \in \mathcal{N}(\Omega)$. Note that 0 is a particular optimal solution in this case.

(ii) If $\Omega\hat{\mu} \not\leq 0$, then $\tilde{\mu}$ is an optimal solution iff we can find a subset Δ of $\{1, 2, \cdots, p\}$ and a particular g-inverse of Ω_{Δ}, say G, such that

$$G\Omega\hat{\mu} \geq 0 \tag{3.34}$$

and

$$\tilde{\mu}_i = \begin{cases} (G\Omega\hat{\mu})_i & \text{if } i \notin \Delta, \\ 0 & \text{if } i \in \Delta. \end{cases} \tag{3.35}$$

Proof: From the literature we know (see Lemma 2.2 in [3] or Theorem 1.7.1 in [18]; compare also [22]) that $\tilde{\mu}$ is an optimal solution of our convex-quadratic program iff

$$\tilde{\mu} \geq 0, \quad \forall \mu \geq 0: \ (\hat{\mu} - \tilde{\mu})^t\Omega(\tilde{\mu} - \mu) \geq 0. \tag{3.36a-b}$$

By substituting 0 as well as $2\tilde{\mu}$ for μ in (3.36b), it is seen that the conditions (3.36) are equivalent to the following set of conditions:

$$\tilde{\mu} \geq 0, \quad \Omega(\hat{\mu} - \tilde{\mu}) \leq 0, \quad \tilde{\mu}^t\Omega(\hat{\mu} - \tilde{\mu}) = 0. \tag{3.37a-c}$$

It therefore suffices to show that the claimed conditions are equivalent to these conditions.

First, let $\Omega\hat{\mu} \leq 0$. Condition (3.37c) can be written as $\tilde{\mu}^t\Omega\hat{\mu} = \tilde{\mu}^t\Omega\tilde{\mu}$. On the one hand, clearly $\tilde{\mu}^t\Omega\tilde{\mu} \geq 0$ as Ω is nnds. On the other hand, $\tilde{\mu}^t\Omega\hat{\mu} \leq 0$ for each $\tilde{\mu} \geq 0$ as $\Omega\hat{\mu} \leq 0$. Consequently $\tilde{\mu}^t\Omega\tilde{\mu} = 0$ or, equivalently, $\Omega\tilde{\mu} = 0$, and claim (i) is obvious.

Next, let us consider the alternative case. That is, let $\theta := \Omega\hat{\mu} \not\leq 0$. We prove the 'only if' part first. For that purpose, let $\tilde{\mu}$ be an optimal solution. We then know that the conditions

(3.37) are all satisfied for $\tilde{\mu}$. In order to keep notation simple, let us further assume that $\tilde{\mu}$ can be partitioned as

$$\tilde{\mu} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix},$$

where $\tilde{\mu}_1 = 0$ and $\tilde{\mu}_2$ is componentwise positive. This can be assumed without loss of generality because by means of a suitable permutation it is always possible to arrive at such a partitioning. Notice that we interpret $\tilde{\mu}_2$ as absent if $\tilde{\mu} = 0$. Let

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad \hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \quad \text{and} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

be all partitioned according to $\tilde{\mu}$. From (3.37) we obtain

$$\theta_1 - \Omega_{12}\tilde{\mu}_2 \leq 0 \quad \text{and} \quad \theta_2 - \Omega_{22}\tilde{\mu}_2 = 0. \tag{3.38a-b}$$

We next show that $\theta_2 \neq 0$. For proving this, suppose $\theta_2 = 0$. Then $\Omega_{22}\tilde{\mu}_2 = 0$, in view of (3.38b). Since Ω is nnds, $\mathcal{N}(\Omega_{22}) \subseteq \mathcal{N}(\Omega_{12})$. Therefore $\Omega_{12}\tilde{\mu}_2 = 0$. Consequently $\tilde{\mu} \in \mathcal{N}(\Omega)$. But then in view of (3.37b) $\theta \leq 0$, a contradiction. So we have $\theta_2 \neq 0$. From Section 2 [observe Theorem 2.1 and (2.4)] it now follows that there exists a g-inverse of Ω_{22}, say G, such that

$$\tilde{\mu}_2 = G\theta_2. \tag{3.39}$$

As usual, let $[\Omega/\Omega_{22}] := \Omega_{11} - \Omega_{12}\Omega_{22}^{-}\Omega_{21}$ denote the (generalized) Schur complement of the block Ω_{22} in Ω (cf. [6]); notice that this Schur complement is invariant for any choice of Ω_{22}^{-}. Inserting expression (3.39) for $\tilde{\mu}_2$ in (3.38a) readily results in

$$[\Omega/\Omega_{22}]\hat{\mu}_1 \leq 0.$$

Letting Δ be the set of all subscripts of components of $\tilde{\mu}_1$, clearly

$$\Omega_{\Delta} = \begin{pmatrix} -I & \Omega_{12} \\ 0 & \Omega_{22} \end{pmatrix}.$$

By Theorem 3.11 we know that the matrix

$$\begin{pmatrix} -I & \Omega_{12}\Omega_{22}^{-} + U(I - \Omega_{22}\Omega_{22}^{-}) \\ (I - \Omega_{22}^{-}\Omega_{22})Z & \Omega_{22}^{-} \end{pmatrix}, \tag{3.40}$$

where U, Z, and Ω_{22}^{-} can be chosen arbitrarily, represents the general g-inverse of Ω_{Δ}. By varying $(\Omega_{\Delta})^{-}$ in $\Omega_{\Delta}\{1\}$, $(\Omega_{\Delta})^{-}\Omega\hat{\mu}$ hence attains any vector of the form

$$\begin{pmatrix} -[\Omega/\Omega_{22}]\hat{\mu}_1 \\ (I - \Omega_{22}^{-}\Omega_{22})Z\theta_1 + \Omega_{22}^{-}\theta_2 \end{pmatrix}, \ \Omega_{22}^{-} \text{ and } Z \text{ both arbitrary.} \tag{3.41}$$

Comparing (3.41) with (3.39) shows that there is a g-inverse $(\Omega_{\Delta})^{-}$ (choose $\Omega_{22}^{-} = G$ and $Z = 0$) such that

$$(\Omega_{\Delta})^{-}\Omega\hat{\mu} = \begin{pmatrix} -[\Omega/\Omega_{22}]\hat{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}.$$

Combining observations now completes the proof of the 'only if' part. To prove the converse, let $\tilde{\mu}$ be such that (3.34) and (3.35) are satisfied for some $\Delta \subseteq \{1, 2, \cdots, p\}$ and some associated particular g-inverse of Ω_Δ, say G. Without loss of generality, let us again assume that Δ consists of leading and adjacent elements from $\{1, 2, \cdots, p\}$. Moreover, as above let us partition $\tilde{\mu}$, $\hat{\mu}$ and Ω, here and now according to Δ. Finally, let $K \in \Omega_{22}\{1\}$ and Z denote those matrices which, according to (3.40), belong to that particular g-inverse G. By hypothesis, clearly $\tilde{\mu}_1 = 0$ and $\tilde{\mu}_2 \geq 0$. Observing (3.41), we further get $[\Omega/\Omega_{22}]\hat{\mu}_1 \leq 0$ and $0 \leq \tilde{\mu}_2 = (I - K\Omega_{22})Z\theta_1 + K\theta_2$, where θ_1 and θ_2 are defined as in the 'only if' part. By arguments already used, it is now not difficult to see that

$$\Omega(\hat{\mu} - \tilde{\mu}) = \begin{pmatrix} [\Omega/\Omega_{22}]\hat{\mu}_1 \\ 0 \end{pmatrix} \leq 0 \tag{3.42}$$

and $\tilde{\mu}^t\Omega(\hat{\mu} - \tilde{\mu}) = 0$. Since the conditions (3.37) are hence all satisfied, $\tilde{\mu}$ indeed turns out to be an optimal solution of our convex-quadratic program. This completes the proof. ∎

At this point it should be stressed that we did not make use of $\Omega\hat{\mu} \not\leq 0$ when proving the 'if part' of Theorem 3.12 (ii). Clearly, $\Omega\hat{\mu} \leq 0$ iff $-I\Omega\hat{\mu} \geq 0$. Choosing $\Delta = \{1, 2, \cdots, p\}$ in such a case trivially leads to $(\Omega_\Delta)^{-1}\Omega\hat{\mu} = -I\Omega\hat{\mu} \geq 0$. This in turn shows that the criterion under statement (ii) [that is, (3.34) along with (3.35)] is even then able to find a particular optimal solution of program $\min_{\mu \geq 0} \|\hat{\mu} - \mu\|_\Omega^2$, namely 0. For later use it is pertinent to mention here that 0 is an optimal solution iff $\Omega\hat{\mu} \leq 0$; this is an easy consequence of the optimality conditions (3.37). Theorem 3.12 indicates the following procedure for determining an optimal solution of this program.

Procedure 3.13.

STEP 1. Compute $\Omega\hat{\mu}$. If $\Omega\hat{\mu} \leq 0$, then $\tilde{\mu} = 0$ is an optimal solution. Else, choose a subset $\Delta \subseteq \{1, 2, \cdots, p\}$ and GOTO STEP 2.

STEP 2. Check the nonsingularity of Ω_Δ. If Ω_Δ is nonsingular, then GOTO STEP 3. Else, GOTO STEP 4.

STEP 3. Compute $(\Omega_\Delta)^{-1}\Omega\hat{\mu}$. If $(\Omega_\Delta)^{-1}\Omega\hat{\mu} \not\geq 0$, then GOTO STEP 4. Else, compute $\tilde{\mu}$ according to

$$\tilde{\mu}_i = \begin{cases} ((\Omega_\Delta)^{-1}\Omega\hat{\mu})_i & \text{if } i \notin \Delta, \\ 0 & \text{if } i \in \Delta; \end{cases}$$

$\tilde{\mu}$ is an optimal solution of program $\min_{\mu \geq 0} \|\hat{\mu} - \mu\|_\Omega^2$.

STEP 4. Modify Δ and GOTO STEP 2.

Proof of Procedure 3.13: We prove that this procedure does find an optimal solution after a finite number of different steps. As we know from the lines introducing this method, $\tilde{\mu} = 0$ is an optimal solution whenever $\Omega\hat{\mu} \leq 0$. In which case we even know that this particular optimal solution can be detected by the criterion in STEP 3. Next, consider the alternative case, that is, let $\Omega\hat{\mu} \not\leq 0$. In view of the preceding theorem, it suffices to show that we can find an optimal solution which fulfills the criterion in STEP 3. To this end, let $\tilde{\mu}$ be an arbitrary but fixed optimal solution. Note that we can assume that $0 \leq \tilde{\mu} \neq 0$, for otherwise we would be back in the case where $\Omega\hat{\mu} \leq 0$. Without loss of generality we can further adopt the same conventions and assumptions as in the proof of the 'only if' part of Theorem 3.12 (ii). Then we have: $\tilde{\mu}_1 = 0$, $\tilde{\mu}_2$ is componentwise positive, Δ consists of the subscripts in vector $\tilde{\mu}_1$, $G\Omega\hat{\mu} \geq 0$ for some g-inverse G of Ω_Δ, and $\tilde{\mu}$ is related to $G\Omega\hat{\mu}$ according to (3.35). Note that Ω_Δ is nonsingular iff Ω_{22} is nonsingular; in which case the proof is complete. If Ω_{22} is singular, then we construct a different optimal solution, say $\bar{\mu}$, which possesses fewer positive entries than $\tilde{\mu}$. For that purpose, choose any vector $0 \neq v_2 \in \mathcal{N}(\Omega_{22})$. Since $\tilde{\mu}_2$ is componentwise positive, it is possible to find a scalar $\lambda \neq 0$ such that $\tilde{\mu}_2 + \lambda v_2$ is componentwise nonnegative but has at least one zero entry. Bordering v_2 by a zero vector of appropriate size results in $v = (0^t, v_2^t)^t \in \mathbb{R}^p$. Put $\bar{\mu} := \tilde{\mu} + \lambda v$. That $\bar{\mu}$ is a feasible solution is obvious. But it is also an optimal solution; for it follows from $\mathcal{N}(\Omega_{22}) \subseteq \mathcal{N}(\Omega_{12})$ that the objective function attains the same (optimal) value at $\tilde{\mu}$ and at $\bar{\mu}$. Clearly, the proof is now complete when either $\bar{\mu} = 0$ or $\bar{\Omega}_{22}$ (by which we denote the principal submatrix of Ω that corresponds to the positive entries of $\bar{\mu}$) is nonsingular. Otherwise, we proceed with $\bar{\mu}$ as above with $\tilde{\mu}$. Doing this repeatedly results in a sequence of different optimal solutions. It is clear that after a finite number of iterations we must arrive at an optimal solution which can be detected by our procedure. This happens because by construction the number of positive entries in that sequence always goes strictly down. ∎

Combining the observations gained in the preceding proof with Theorem 3.12 even results in the following representation for the set of optimal solutions of program $\min_{\mu \geq 0} \|\hat{\mu} - \mu\|_\Omega^2$.

Theorem 3.14. For given $\Omega \in \mathbb{R}^{p,p}$ and $\hat{\mu} \in \mathbb{R}^p$, consider program $\min_{\mu \geq 0} \|\hat{\mu} - \mu\|_\Omega^2$. Let $\{\tilde{\mu}^1, \tilde{\mu}^2, \cdots, \tilde{\mu}^k\}$ consist of all those feasible solution vectors $\tilde{\mu}$ for which (3.34) and (3.35) hold true for some nonsingular G. Then

$$\operatorname{argmin}_{\mu \geq 0} \|\hat{\mu} - \mu\|_\Omega^2 = \mathcal{P} + \mathcal{K},$$

where $\mathcal{K} := \{\mu \geq 0 \mid \Omega\mu = 0\}$ and $\mathcal{P}$ denotes the convex hull of $\{\tilde{\mu}^1, \tilde{\mu}^2, \cdots, \tilde{\mu}^k\}$.

From the proof of Procedure 3.13 we single out the following interesting remark concerning Procedure 3.10 and our original ICGLS problem.

Remark 3.15. Consider Procedure 3.10. Observe that $\Omega\hat{\mu}$, in contrast to $\hat{\mu} := R\hat{\beta}_u - r$, is always invariant for any choice of $\hat{\beta}_u \in \{\hat{\beta}_u(y)\}$. This follows from representation (3.21) in view of $\mathcal{N}(\Omega) = R\mathcal{N}(X)$ [see (3.11)]. Recall that we have a similar result for $\hat{\mu}$ iff inclusion (1.7) holds true. As we have already seen, $0 \in \tilde{M}(y)$ iff $\Omega\hat{\mu} \leq 0$. In which case, in accordance with Theorem 3.14, $\tilde{M}(y) = \{\mu \geq 0 \mid \Omega\mu = 0\} =: \mathcal{K}$, showing that $\tilde{M}(y)$ is a singleton iff $\mathcal{K} = \{0\}$. Nevertheless, if (1.7) is not satisfied, then there do exist, even when $\mathcal{K} = \{0\}$, realizations $y \in \mathcal{R}(V)$ for which $\tilde{M}(y)$ is not a singleton. For recognize carefully that, according to Theorem 3.6, $\tilde{M}(\cdot)$ is a singleton iff inclusion (1.7) holds true.

4. ICGLS-Problem: Special Model

In this section, we consider model $\mathcal{L}_r$ under (1.2) and (1.7). As in Section 3, it is further assumed that R is of full row rank. Let Γ be defined by (3.5), that is, let $\Gamma = X^t V^- X$. From Theorem 3.6 we then know that $\Omega := (R^Q)^t \Gamma R^Q$ is nonsingular and $\tilde{M}(\cdot)$ is a singleton. We denote the unique element of $\tilde{M}(\cdot)$ by $\tilde{\mu}(\cdot)$. Representation (3.20) for the set of ICGLS selections for β obviously reduces to

$$\{\tilde{\beta}(\cdot)\} = \{R^Q(r + \tilde{\mu}) + (I - R^Q R)\hat{\beta}_u + P_{\mathcal{N}(X)} w(\cdot) \mid w(\cdot) \text{ arbitrary}\}$$

or, equivalently,

$$\{\tilde{\beta}(\cdot)\} = \{\hat{\beta}_u - R^Q(\hat{\mu} - \tilde{\mu}) + P_{\mathcal{N}(X)}w(\cdot) \mid w(\cdot) \text{ arbitrary}\},$$

where $\hat{\beta}_u$ stands for an arbitrary but fixed GLS selection for β under model $\mathcal{L}_u$ and where $\hat{\mu} := R\hat{\beta}_u - r$. Recall that $X^P y$ with $X^P = \Gamma^- X^t V^-$ can be chosen as $\hat{\beta}_u$. By means of Theorem 2.11 and (3.33) [compare also (3.30) in Remark 3.7] it is seen that Ω can be written as

$$\Omega = (R\Gamma^- R^t)^{-1}.$$

Because of (3.6) it follows from Theorem 2.5 that R^Q can be chosen as

$$R^Q = \Gamma^- R^t \Omega. \tag{4.1}$$

By combining Procedure 3.9 with Procedure 3.13, it is now seen that for given $y \in \mathcal{R}(V)$ the *complete* set of ICGLS solutions for β can be determined by the following procedure.

Procedure 4.1.

STEP 1. Compute $\hat{\beta}_u = \Gamma^- X^t V^- y$, $\hat{\mu} = R\hat{\beta}_u - r$, $\Omega = (R\Gamma^- R^t)^{-1}$ and $R^Q = \Gamma^- R^t \Omega$. Choose a subset Δ of the set $\{1, 2, \cdots, p\}$.

STEP 2. Compute $(\Omega_\Delta)^{-1}\Omega\hat{\mu}$.

STEP 3. If $(\Omega_\Delta)^{-1}\Omega\hat{\mu} \geq 0$, GOTO STEP 4. Else, modify Δ and GOTO STEP 2.

STEP 4. Set up $\tilde{\mu}$ according to

$$(\tilde{\mu})_i = \begin{cases} \left[(\Omega_\Delta)^{-1}\Omega\hat{\mu}\right]_i & \text{if } i \notin \Delta, \\ 0 & \text{if } i \in \Delta. \end{cases} \tag{4.2}$$

The set of ICGLS solutions for β is then given by

$$\{\tilde{\beta}(y)\} = \{\hat{\beta}_u - R^Q(\hat{\mu} - \tilde{\mu}) + P_{\mathcal{N}(X)}w \mid w \in \mathbb{R}^m\}. \tag{4.3}$$

After a finite number of different steps this method will find the *complete* set of ICGLS solutions. It should be mentioned that $\tilde{\mu} = 0$ iff $\Omega\hat{\mu} \leq 0$ [compare Remark 3.15]. Moreover, by comparing (4.3) with (3.21), it is seen that $\{\tilde{\beta}(y)\} = \{\hat{\beta}_u(y)\}$ iff $\hat{\mu} \geq 0$, where according to our conventions the complete set of GLS solutions for β under $\mathcal{L}_u$ (given y) is again denoted by $\{\hat{\beta}_u(y)\}$.

Next let us assume, without loss of generality, that Δ in STEP 4 consists of leading and adjacent elements from $\{1, 2, \cdots, p\}$. The vector $\tilde{\mu}$ in STEP 4 can then be partitioned as follows

$$\tilde{\mu} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix},$$

where the set of subscripts in $\tilde{\mu}_1$ coincides with Δ. Of course, $\tilde{\mu}_1 = 0$. Let

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}, \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, \quad r = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

be all partitioned according to $\tilde{\mu}$. We then know from (3.42) that

$$\Omega(\hat{\mu} - \tilde{\mu}) = \begin{pmatrix} [\Omega/\Omega_{22}]\hat{\mu}_1 \\ 0 \end{pmatrix},$$

where $[\Omega/\Omega_{22}]$ stands for the Schur complement of the block Ω_{22} in Ω. From the well-known Schur idendity for partitioned positive definite matrices (cf. [15, p. 45]; see also Theorem 1.23 in [6]) it follows that

$$[\Omega/\Omega_{22}] = \left(R_1 \Gamma^- R_1^t\right)^{-1}.$$

For $R^Q = \Gamma^- R^t \Omega$ [see (4.1)], we hence get

$$R^Q(\hat{\mu} - \tilde{\mu}) = R_1^Q \hat{\mu}_1,$$

where $R_1^Q := \Gamma^- R_1^t \left(R_1 \Gamma^- R_1^t\right)^{-1}$. Evidently $(R_1^Q)^t \in \mathcal{G}(R_1^t \mid \Gamma)$, that is, $(R_1^Q R_1)^t \in \mathcal{P}(R_1^t \mid \Gamma)$. But now it is also clear that $\{\tilde{\beta}(y)\}$ coincides for that realization y with the complete set of ECGLS solutions for β under model $(y,\ X\beta,\ V \mid R_1\beta = r_1)$. This observation proves Theorem 4.2 given below. In the theorem some further notation is used. For given $A \in \mathbb{R}^{p,q}$ and $\Delta \subseteq \{1, 2, \cdots, p\}$, it is convenient to write $A_{r\Delta}$ for the submatrix of A that contains the ith row vector of A if $i \in \Delta$. For given $y \in \mathcal{R}(V)$, the set of all ECGLS solutions for β under model

$$\mathcal{L}_\Delta := (y,\ X\beta, V \mid R_{r\Delta}\beta = r_{r\Delta})$$

is denoted by $\{\hat{\beta}_\Delta(y)\}$. If $\Delta = \emptyset$, we interpret $A_{r\Delta}$ as absent.

Theorem 4.2. Consider model $\mathcal{L}_r$ under (1.2) and (1.7), and let R be of full row rank. Moreover, for given $y \in \mathcal{R}(V)$, let $\tilde{\mu}$ be determined by Procedure 4.1, and let Δ be a subset of $\{1, 2, \cdots, p\}$ being associated with $\tilde{\mu}$ according to (4.2) in STEP 4. Then

$$\{\tilde{\beta}(y)\} = \{\hat{\beta}_\Delta(y)\},$$

and the general ICGLS solution can be represented as follows

$$\tilde{\beta}(y) = \hat{\beta}_u - \Gamma^-(R_{r\Delta})^t \left[R_{r\Delta}\Gamma^-(R_{r\Delta})^t\right]^{-1} (R_{r\Delta}\hat{\beta}_u - r_{r\Delta}) + P_{\mathcal{N}(X)} w, \tag{4.4}$$

where w is arbitrary, Γ is defined as before, and $\hat{\beta}_u$, standing for an arbitrary but fixed GLS solution under $\mathcal{L}_u$, can especially be chosen as $\hat{\beta}_u = \Gamma^- X^t V^- y$.

Representation (4.4) was established in [29, p.385] for the 'full rank' model $\mathcal{L}_r$ (that is, model $\mathcal{L}_r$ with X and V both of full column rank). Notice that the ICGLS estimator of β does exist in such a model; see Theorem 3.8. We mention that an alternative method for determining this estimator is discussed in [29]. Comparing both procedures it turns out that Procedure 4.1 is easier to follow and to compute. One reason is that an explicitly known basis for the null space of R is not required in the above procedure. A further reason is that the matrices we have to g-invert are generally smaller in size than in Procedure 3.3 in [29].

We now conclude this paper with stating two simple versions of Procedure 4.1 in case we have only one restriction (that is, $p = 1$) or the restrictions are such that Ω is a positive diagonal matrix. If $p = 1$, then the ICGLS solutions for β can be obtained according to the following procedure.

Procedure 4.3 *(Two-step procedure).*

STEP 1. Compute a GLS solution for β under $\mathcal{L}_u$, for instance $\hat{\beta}_u = \Gamma^- X^t V^- y$, where $\Gamma = X^t V^- X$. If $R\hat{\beta}_u \geq r$, then $\{\tilde{\beta}(y)\} = \{\hat{\beta}_u + P_{\mathcal{N}(X)} w \mid w \in \mathbb{R}^m\}$. Else, GOTO STEP 2.

STEP 2. Compute $R^Q = \Gamma^- R^t (R\Gamma^- R^t)^{-1}$. The set of ICGLS solutions coincides with the set of ECGLS solutions under $\mathcal{L}_e$, that is,

$$\{\tilde{\beta}(y)\} = \{R^Q r + (I - R^Q R)\hat{\beta}_u + P_{\mathcal{N}(X)} w \mid w \in \mathbb{R}^m\}.$$

If X is of full column rank, then the estimator obtained by this method is often called the two-step estimator (see, e.g, [29] and [12]). Procedure 4.3 is an extension of Procedure 5.1 in [29]. This procedure is readily obtained from Procedure 4.1 by observing Theorem 4.2 as well as the representations (3.21) and (3.23) in Section 3.

The following extended two-step procedure is a direct consequence of Procedure 4.1 and holds true for each $y \in \mathcal{R}(V)$ whenever $\Omega = (R\Gamma^- R^t)^{-1}$ is a (positive) diagonal matrix. We mention that Procedure 5.2 in [29] can be reobtained as special case.

Procedure 4.4 *(Extended two-step procedure).*

STEP 1. Compute a GLS solution for β under $\mathcal{L}_u$, for instance $\hat{\beta}_u = \Gamma^- X^t V^- y$, where $\Gamma = X^t V^- X$. If $R\hat{\beta}_u \geq r$, then $\{\tilde{\beta}(y)\} = \{\hat{\beta}_u + P_{\mathcal{N}(X)} w \mid w \in \mathbb{R}^m\}$. Else, GOTO STEP 2.

STEP 2. Let $\Delta := \{i \mid (R\hat{\beta}_u - r)_i < 0\}$. Compute $(R_{r\Delta})^Q := \Gamma^- (R_{r\Delta})^t \left[R_{r\Delta}\Gamma^- (R_{r\Delta})^t\right]^{-1}$. Then

$$\{\tilde{\beta}(y)\} = \{(R_{r\Delta})^Q r_{r\Delta} + \left[I - (R_{r\Delta})^Q R_{r\Delta}\right] \hat{\beta}_u + P_{\mathcal{N}(X)} w \mid w \in \mathbb{R}^m\}.$$

The authors would like to thank a referee for his comments on this paper.

References

[1] T. Amemiya. *The solvability of a two-market disequilibrium model.* Working paper No. 82, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1977.

[2] A. Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications.* Wiley, New York, 1974 (Krieger, New York, 1980.

[3] A. Berman. *Cones, Matrices and Mathematical Programming.* Lecture Notes in Economics and Mathematical Systems, No. 79, Springer-Verlag, Berlin, 1973.

[4] E. Blum and W. Oettli. *Mathematische Optimierung.* Springer, Berlin, 1975.

[5] J. S. Chipman. *On least squares with insufficient observations.* J. Amer. Statist. Assoc. 59: 1078–1111 (1964).

[6] M. Fiedler. *Special Matrices and their Applications in Numerical Mathematics.* Nijhoff, Dordrecht, 1986.

[7] F. K. Firoozi. *A transformation of the inequality-constrained linear model.* Linear Algebra Appl. 133: 153–163 (1990).

[8] J. Hartung and H. J. Werner. *Zum Auswerten unbalancierter Daten.* EDV in Medizin und Biologie 14: 13–18 (1983).

[9] J. Hartung and H. J. Werner. *Hypothesenprüfung im restringierten linearen Modell: Theorie und Anwendungen.* Vandenhoeck & Ruprecht, Göttingen, 1984.

[10] T. Ito. *Methods of estimation for multi-market disequilibrium models.* Econometrica 48: 97–125 (1980).

[11] C. E. Langenhop. *On generalized inverses of matrices.* SIAM J. Appl. Math. 15: 1239–1246 (1967).

[12] M. C. Lovell and E. Prescott. *Multiple regression with inequality constraints: Pretesting bias, hypothesis testing and efficiency.* J. Amer. Statist. Assoc. 65: 913–925 (1970).

[13] S. K. Mitra. *Fixed rank solutions for linear matrix equations.* Sankhyā Ser. A 34: 387–392 (1972).

[14] R. L. Plackett. *Some theorems in least squares.* Biometrika 37: 149–157 (1950).

[15] R. M. Pringle and A. A. Rayner. *Generalized Inverse Matrices with Applications to Statistics.* Griffin, London, 1971.

[16] C. R. Rao and S. K. Mitra. *Generalized Inverse of Matrices and its Applications.* Wiley, New York, 1971.

[17] C. R. Rao and H. Yanai. *Generalized inverse of linear transformations: A geometric approach.* Linear Algebra Appl. 66: 87–98 (1985).

[18] C. R. Rao and J. Kleffe. *Estimation of Variance Components and Applications.* North-Holland, Amsterdam, 1988.

[19] C. R. Rao. *Linear Statistical Inference and its Applications.* Wiley, New York, 1973 (Wiley Eastern, New Delhi, 1991).

[20] H. Scheffé. *The Analysis of Variance.* Wiley, New York, 1959.

[21] P. Schmidt. *Constraints on the parameters in simultaneous tobit and probit models,* **in:** *Structural Analysis of Discrete Data with Econometric Applications (C. F. Manski et al., eds.).* MIT Press, Cambridge, 1981.

[22] P. Schönfeld and H. J. Werner. *Beiträge zur Theorie und Anwendung linearer Modelle,* **in:** *Ökonomische Prognose- Entscheidungs- und Gleichgewichtsmodelle (W. Krelle, ed.).* VCH Verlagsgesellschaft, Weinheim, 1986.

[23] P. Schönfeld and H. J. Werner. *A note on C. R. Rao's wider definition BLUE in the general Gauß-Markov model.* Sankhyā Ser. B 49: 1–8 (1987).

[24] H. J. Werner. *G-Inverse und monotone Matrizen.* Dissertation, University of Bonn,

1977.

[25] H. J. Werner. *Weak complementarity and missing observations.* Meth. Oper. Res. 50: 421–439 (1985).

[26] H. J. Werner. *More on BLU estimation in regression models with possibly singular covariances.* Linear Algebra Appl. 67: 207–214 (1985).

[27] H. J. Werner. *Generalized inversion and weak bi-complementarity.* Lin. Multilin. Alg. 19: 357–372 (1986).

[28] H. J. Werner. *C. R. Rao's IPM method: a geometric approach,* **in:** *New Perspectives in Theoretical and Applied Statistics (M. L. Puri et al., eds.).* Wiley, New York, 1987.

[29] H. J. Werner. *On inequality constrained generalized least squares estimation.* Linear Algebra Appl. 127: 379–392 (1990).

[30] F. Yates and R. W. Hale. *The analysis of latin squares when two or more rows, columns, or treatments are missing.* J. Royal Satist. Soc. Supplement 6: 67–79 (1939).