# Estimation of Constrained Singular Seemingly Unrelated Regression Models

By Harry Haupt and Walter Oberhofer

**Abstract**

The topic of this paper is the problem of a singular disturbance covariance matrix in (seemingly unrelated) systems of linear regression equations. This singularity is considered as being caused by exact linear restrictions on the endogenous variables, adding-up to a predetermined aggregate. It is well known, that the estimation of such systems require the substitution of one equation by the "adding up condition" as proposed by Barten (1969), or, as alternatively proposed by Theil (1971), a modification of the GLS estimator. The consequently occuring question of the invariance of the parameter estimates to the choice of which equation is deleted has been discussed exhaustively by Powell (1969) for GLS and Barten (1969) for ML estimation. Dhrymes and Schwarz (1987a,b) pointed out the parallels between the different procedings and argued that the estimator of Theil (1971) fails to exist in most of the practically relevant constellations. The rank conditions given by Theil (1971) and the corresponding objections of Dhrymes and Schwarz (1987a,b) are substantially simplified and generalized.

**JEL classification:** C30.

**Keywords:** Seemingly unrelated regression, singular equation systems, adding-up restriction, system estimation, GLS, FGLS.

**Address of the authors:**
University of Regensburg
Department of Economics and Econometrics
Universitaetsstr. 31, D-93040 Regensburg
Tel.: (..49) (941) 943-2736, Fax.: (..49) (941) 943-4917
http: \\ www.wiwi.uni-regensburg.de \ oberhofer \

# 1 Introduction

Much of the the attention dedicated to regression models with a singular covariance matrix of disturbances has been stimulated by the work on allocation models, mainly demand systems. Typically this models contain several, say $n$ behavioral equations and an additional identity. This identity implies that $n$ dependendent variables $y_i$ $(1 \leq i \leq n)$ (expenditures for the different commodities) add up to a fixed, known value (total expenditures). Consequently, having $n+1$ equations and $n$ dependent variables the resulting equation system is singular. Considering models with $n$ endogenous variables and $T$ observations on every variable, we denote the observations by $y_{t,i}, (t = 1, \dots, T), (i = 1, \dots, n)$.
Of course there are several possible forms of the adding up condition. For example

$$\text{(a)} \qquad \sum_{i=1}^{n} y_{t,i} = s_t \qquad \forall t,$$

$$\text{(b)} \qquad \sum_{i=1}^{n} y_{t,i} = 1 \qquad \forall t \qquad \text{and}$$

$$\text{(c)} \qquad \sum_{i=1}^{n} a_i y_{t,i} = s_t \qquad \forall t.$$

In demand theory the identity usually is of form (a) or (b). Case (b) is given when the $n$ dependent variables are represented as shares of the predetermined aggregate. A more general identity is given in case (c), with a fundamental set partitioned in $n$ disjoint subsets, where $y_{t,i}$ represents the share value in subset $i$ and $y_t$ represents the fixed share value of the fundamental set. Denoting the non-stochastic shares of subset $i$ of the fundamental set as $a_i$, the weighted $y_{t,i}$ will sum up identically to $s_t$. While (b) is nested in (a), both are special cases of (c).

Whatever identity is considered, the implicit adding-up condition causes the seemingly unrelated variables $y_{t,i}$ to be related, i.e. correlated. In regression analysis the residuals inherit this correlation from the dependent variables, resulting in a singular covariance matrix. Usually, in the context of unconstrained estimation, the resulting rank deficit of the covariance matrix is one. The consequently occuring problems when computing GLS or ML estimates can be solved either by deleting one equation before the estimation procedure or by modification of the estimator (see Bewley, 1986).

In the framework of the classical multivariate single equation regression model with the assumption of a singular covariance matrix Theil (1971) (and Kreijger and Neudecker (1977)) developed a best linear unbiased GLS and constrained GLS estimator using the Moore-Penrose generalized inverse. Deleting one equation is one possible procedure to be able to estimate the unknown parameters of the model. As stated by Bewley (1986) in his monograph on allocation models "The major question that then arises is does the estimator depend on which equation is deleted?" Powell (1969) considering both cases (a) and (c) using GLS estimates and Barten (1969) considering case (a) using a generalisation of the ML system estimator, treated the invariance problem in the system of equations model. Dhrymes and Schwarz (1987a,b) pointed out, that "the literature has generally dealt with a special case – in which all explanatory variables appear in every equation", but the Barten estimator does "not hold for the general case" and the Theil-Kreijger-Neudecker (TKN) estimator "fails to exist when the equations of the system contain one or more variables in common", since the rank conditions postulated by Theil (1971) are not fulfilled.

Similar problems occur in the context of panel data models, where sometimes a set of restrictions of the form

$$\text{(d)} \qquad \sum_{i=1}^{n} y_{t,i,j} = r_{t,j} \qquad \forall j, \forall t \qquad \text{and} \qquad \sum_{j=1}^{m} y_{t,i,j} = c_{t,i} \qquad \forall i, \forall t.$$

has to be considered, whereas in the macroeconometric or simultaneous systems approach usually the problem of a singular system covariance matrix occurs due to identities of endogenous variables also occuring in behavioral equations. This two models represent practically relevant cases where the rank deficit (due to restrictions) of the system covariance matrix could be greater than one.

In the second section we introduce a general model, where we discuss all necessary preliminaries, assumptions and constraints and consider relevant model alternatives. Estimation topics are discussed in section three. In section four we derive less restrictive rank conditions than postulated by Theil (1971) in a more general framework as in the corresponding considerations of Dhrymes and Schwarz (1987a,b). The special case of identical regressors in every equation of the system and homoscedasticity is considererd in section five. Section six deals with the problem of the deletion choice. Strongly related to the work of Powell (1969) and Barten (1969) we show, that under consideration of the constraints, the estimator is invariant to the deletion choice and equivalent to the modified GLS estimator. In section seven we briefly discuss starting conditions for feasible GLS estimation.

# 2 The framework

## 2.1 Alternative specifications

The basic model is the following general system of regression equations model

$$\text{(i)} \qquad y_{t,i} = \sum_{k} x_{t,k,i} b_{k,i} + u_{t,i}.$$

Nested within (i) is the model considered by Powell (1969)

$$\text{(ii)} \qquad y_{t,i} = \sum_{k} x_{t,k} b_{k,i} + \sum_{l} z_{t,l,i} c_{l,i} + u_{t,i},$$

which contains both, independent variables $x_{t,k}$ that are common to every equation and independent variables $z_{t,l,i}$ that are specific for some of the $n$ equations. A special case of (ii) is the model with identical regressors in each of the $n$ equations,

$$\text{(iii)} \qquad y_{t,i} = \sum_{k} x_{t,k} b_{k,i} + u_{t,i}.$$

Powell, using specification (ii), considered adding-up conditions

$$\text{(I)} \qquad \sum_{i} y_{t,i} = x_{t,1} \qquad \forall t, \qquad \text{or}$$

$$\text{(II)} \qquad \sum_{i} a_i y_{t,i} = x_{t,1} \qquad \forall t,$$

with the resulting constraints

$$\text{(I')} \qquad \sum_i b_{1,i} = 1 \quad \text{and} \quad \sum_i b_{k,i} = 0 \quad \text{for} \quad k \geq 2, \qquad \text{or}$$

$$\text{(II')} \qquad \sum_i a_i b_{1,i} = 1 \quad \text{and} \quad \sum_i a_i b_{k,i} = 0 \quad \text{for} \quad k \geq 2,$$

and assumed

$$\text{(I'')} \qquad \sum_i z_{t,l,i} = 0 \qquad \forall t, \forall l, \qquad \text{or}$$

$$\text{(II'')} \qquad \sum_i a_i z_{t,l,i} = 0 \qquad \forall t, \forall l.$$

He proved that, under condition (I) considering (I') and assuming (I''), the Aitken estimator (modified by using the Moore-Penrose inverse) is invariant to the choice of which equation is deleted and that condition (II) represents a linear transformation of the minimisation problem under (I). Thus, given the invariance of the Aitken estimator under linear transformations, the invariance result is equally valid for weighted aggregates.

## 2.2 A general model

Consider the system of regression equations

$$\text{(1)} \qquad y_{t,i} = \sum_{k=1}^{K_i} b_{k,i} x_{t,k,i} + u_{t,i} \qquad 1 \leq i \leq n, 1 \leq t \leq T,$$

with the adding-up restriction

$$\text{(2)} \qquad \sum_{i=1}^{n} a_i y_{t,i} = s_t \qquad \forall t,$$

where in general $t$ is a time index and $i$ is a cross-section index. Equation (1) could be rendered to

$$\text{(3)} \qquad \mathbf{y}_{\bullet,i} = \mathbf{X}_{\bullet,i} \mathbf{b}_{\bullet,i} + \mathbf{u}_{\bullet,i} \qquad 1 \leq i \leq n,$$

where $\mathbf{y}_{\bullet,i}$ is a vector of $T$ observations of the dependent variable, $\mathbf{X}_{\bullet,i}$ is a $T \times K_i$ regressor matrix and $\mathbf{b}_{\bullet,i}$ contains the $K_i$ parameters of the $i$th equation.

Assumption 1: All regressors are nonstochastic.

Assumption 2: $\mathrm{E}\mathbf{u}_{\bullet,i} = \mathbf{0} \qquad \forall i.$

Assumption 3: $\operatorname{cov}(u_{t,i} u_{t,j}) = \begin{cases} 0 & \text{for} \quad t \neq s, \quad 1 \leq t, s \leq T, \\ \sigma^2 \omega_{t,i,j} & \text{for} \quad t = s, \quad 1 \leq i, j \leq n. \end{cases}$

Thus the covariance matrix is denoted by $\mathbf{\Omega}_t$ with known elements $\omega_{t,i,j}$.

Assumption 4: $\sum_i a_i^2 > 0$ and all $a_i$ are known and combined to the column vector $\mathbf{a}$.

ASSUMPTION 5: All $s_t$ are known and combined to the column vector $\mathbf{s}$.

Since the adding-up restriction should be identically fulfilled in the disturbances from (3) we get

$$(4) \qquad \sum_i a_i \mathbf{u}_{\bullet,i} = \mathbf{0}$$

and

$$(5) \qquad \sum_i a_i \mathbf{X}_{\bullet,i} \mathbf{b}_{\bullet,i} = \mathbf{s}.$$

From (4) follows

$$(6) \qquad \mathbf{\Omega_t a} = \mathbf{0}.$$

Without loss of generality $\mathbf{a}'\mathbf{a} = 1$.

ASSUMPTION 6: $\mathbf{\Omega}_t$ has a single zero eigenvalue.

This assumption implies that the singularity of $\mathbf{\Omega}_t$ is solely caused by the adding-up restriction and gives way to the diagonalization

$$(7) \qquad \mathbf{\Omega}_t = \begin{pmatrix} \mathbf{a} & \mathbf{F}_t \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{\Lambda}_t \end{pmatrix} \begin{pmatrix} \mathbf{a}' \\ \mathbf{F}_t' \end{pmatrix}, \qquad \forall t,$$

where $\mathbf{\Lambda}_t$ is an $n-1$ nonsingular diagonal matrix and $\begin{pmatrix} \mathbf{a} & \mathbf{F}_t \end{pmatrix}$ is orthogonal. The latter implies

$$\mathbf{a}'\mathbf{F}_t = 0, \quad \mathbf{a}'\mathbf{a} = 1, \quad \mathbf{F}_t'\mathbf{F}_t = \mathbf{I}_{n-1} \quad \text{and} \quad \mathbf{aa}' + \mathbf{F}_t\mathbf{F}_t' = \mathbf{I}_n.$$

From (7) follows

$$(8) \qquad \mathbf{\Omega}_t = \mathbf{F}_t \mathbf{\Lambda}_t \mathbf{F}_t',$$

and consequently for the Moore-Penrose inverse

$$(9) \qquad \mathbf{\Omega}_t^+ = \mathbf{F}_t \mathbf{\Lambda}_t^{-1} \mathbf{F}_t'.$$

Dhrymes and Schwarz (1987b, p.237) stated that "the heart of the problem is that the conditions on the parameters force the singularity of the covariance matrix – and to a certain degree the converse is true, i.e., the singularity of the covariance matrix implies certain restrictions". It is important to note that, as stated by Bewley (1986, p.10, 12), "a necessary and sufficient condition for the OLS estimates to satisfy the adding-up criterion is that some linear combination of the regressors must be identically equal to the sum of regressands if the model ... is to be logically consistent."

Since the constraints in (5) depend on the values of the regressors, we postulate that the constraints are identically valid in the regressors. This induces restrictions on the parameters which are independent from the regressors. Thus let $\mathbf{Z}$ be a $T \times p$ matrix of $T$-vectors $\mathbf{z_1}, \ldots, \mathbf{z_p}$, which constitute a base of the vector space containing the $\sum_i K_i$ regressors of all $n$ equations. The obvious consequence is the existence of $n$ matrices $\mathbf{C}_i$ of order $p \times K_i$ with

$$\mathbf{X}_{\bullet,i} = \mathbf{ZC}_i, \qquad \forall i.$$

Thus, the constraints in (5) could be written as

$$\mathbf{Z}\begin{pmatrix} a_1\mathbf{C}_1 & \dots & a_n\mathbf{C}_n \end{pmatrix}\mathbf{b} = \mathbf{s}, \tag{10}$$

where $\mathbf{b}$ is the $(\sum_i K_i)$-vector of parameters in equation system (3).

Assumption 7: $\mathbf{s}$ is element of the linear space spanned by the $(\sum_i K_i)$ regressors.

This implies

$$\mathbf{s} = \mathbf{Zc}, \tag{11}$$

where $\mathbf{c}$ is a suitable vector. Combining (10) and (11) we get

$$\begin{pmatrix} a_1\mathbf{C}_1 & \dots & a_n\mathbf{C}_n \end{pmatrix}\mathbf{b} = \mathbf{c}, \tag{12}$$

or, more compact

$$\mathbf{Sb} = \mathbf{c}, \tag{13}$$

where $\mathbf{S}$ is an $p \times (\sum_i K_i)$ matrix. If any of the regressors cannot be reproduced as linear combination of others, then (12) implies that the corresponding parameter could be obtained through the constraints. Consider an illustrative example for the special case $a_i = 1,\ \forall i$ given in Bewley (1986, p.22)

$$\begin{aligned} y_{t,1} &= b_{1,1}x_{t,1} + b_{2,1}x_{t,2} + u_{t,1}, \\ y_{t,2} &= b_{2,2}x_{t,2} + b_{3,2}x_{t,3} + u_{t,2}, \\ y_{t,1} &= b_{1,3}x_{t,1} + u_{t,3}, \end{aligned}$$

and adding-up restrictions

$$y_{t,1} + y_{t,2} + y_{t,3} = x_{t,1}.$$

Summing up the behavioral equations gives

$$(y_{t,1} + y_{t,2} + y_{t,3}) = (b_{1,1} + b_{1,3})x_{t,1} + (b_{2,1} + b_{2,2})x_{t,2} + b_{3,2}x_{t,3} + (u_{t,1} + u_{t,2} + u_{t,3}).$$

Thus we have

$$\begin{aligned} u_{t,1} + u_{t,2} + u_{t,3} &= 0, \\ b_{1,1} + b_{1,3} &= 1, \\ b_{2,1} + b_{2,2} &= 0, \\ b_{3,2} &= 0. \end{aligned}$$

In various applications there is the need to impose additional exogenous restrictions of the form

$$\mathbf{Rb} = \mathbf{r}, \tag{14}$$

where $\mathbf{R}$ is an $q \times (\sum_i K_i)$ matrix.
Combining (13) and (14) we get

$$\begin{pmatrix} \mathbf{S} \\ \mathbf{R} \end{pmatrix}\mathbf{b} = \begin{pmatrix} \mathbf{c} \\ \mathbf{r} \end{pmatrix}. \tag{15}$$

We further assume that the constraints (15) are not contradictory.

ASSUMPTION 8: $\mathrm{rk}\begin{pmatrix} \mathbf{S} & \mathbf{c} \\ \mathbf{R} & \mathbf{r} \end{pmatrix} = \mathrm{rk}\begin{pmatrix} \mathbf{S} \\ \mathbf{R} \end{pmatrix}$.

The $n$ seemingly unrelated equations are connected by the adding-up restriction. Since this connection is independent from time, we introduce an alternative representation of the system by aggregating groupewise for every $t$, given by

$$(16) \qquad \mathbf{y}_{t,\bullet} = \mathbf{X}_{t,\bullet}\mathbf{b} + \mathbf{u}_{t,\bullet}, \qquad \forall t,$$

where the $n \times (\sum_i K_i)$ matrix $\mathbf{X}_{t,\bullet}$ is given by

$$\mathbf{X}_{t,\bullet} \equiv \begin{pmatrix} \mathbf{x}'_{t,\bullet,1} & 0 & \dots & 0 \\ 0 & \mathbf{x}'_{t,\bullet,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{x}'_{t,\bullet,n} \end{pmatrix},$$

and $\mathbf{x}'_{t,\bullet,i}$ represents the column vector of the $K_i$ regressors in equation $i$ at time $t$. Aggregating (16) subject to

$$\mathbf{y} \equiv \begin{pmatrix} \mathbf{y}_{1,\bullet} \\ \mathbf{y}_{2,\bullet} \\ \vdots \\ \mathbf{y}_{T,\bullet} \end{pmatrix}, \mathbf{u} \equiv \begin{pmatrix} \mathbf{u}_{1,\bullet} \\ \mathbf{u}_{2,\bullet} \\ \vdots \\ \mathbf{u}_{T,\bullet} \end{pmatrix}, \mathbf{b} \equiv \begin{pmatrix} \mathbf{b}_{\bullet,1} \\ \mathbf{b}_{\bullet,2} \\ \vdots \\ \mathbf{b}_{\bullet,n} \end{pmatrix} \quad \text{and} \quad \mathbf{X} \equiv \begin{pmatrix} \mathbf{X}_{1,\bullet} \\ \mathbf{X}_{2,\bullet} \\ \vdots \\ \mathbf{X}_{T,\bullet} \end{pmatrix},$$

we get

$$(17) \qquad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}.$$

We need, however, a rank condition to guarantee the identification of the parameters. The parameters in equation (17) are not identified in case of the existence of a nonzero vector $\mathbf{\Delta b}$ with $\mathbf{X\Delta b} = \mathbf{0}$, $\mathbf{S\Delta b} = \mathbf{0}$ and $\mathbf{R\Delta b} = \mathbf{0}$.

ASSUMPTION 9: The matrix $\mathbf{H} \equiv \begin{pmatrix} \mathbf{X} \\ \mathbf{S} \\ \mathbf{R} \end{pmatrix}$ has full column rank.

By suitable row conversions the matrix $\mathbf{H}$ could be rendered to

$$\begin{pmatrix} \mathbf{X}_{\bullet,1} & 0 & \dots & 0 \\ 0 & \mathbf{X}_{\bullet,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_{\bullet,n} \\ a_1\mathbf{C}_1 & a_2\mathbf{C}_2 & \dots & a_n\mathbf{C}_n \\ & \mathbf{R} & & \end{pmatrix}.$$

Then, obviously full column rank of $\mathbf{X}_{\bullet,i}$, $\forall i$, is a sufficient condition for Assumption 9 to be fulfilled. The covariance matrix of $\mathbf{u}$ in (17) is given by

$$(18) \qquad \mathrm{E}\mathbf{uu}' \equiv \mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega}_1 & 0 & \dots & 0 \\ 0 & \mathbf{\Omega}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{\Omega}_T \end{pmatrix}.$$

Hence, considering (6)

$$\boldsymbol{\Omega}\bar{\mathbf{a}} = \mathbf{0}, \tag{19}$$

with $\bar{\mathbf{a}} \equiv \mathbf{e}_T \otimes \mathbf{a}$, where $\mathbf{e}_T$ is an T-vector with all elements equal to unity. From (8) follows

$$\boldsymbol{\Omega} = \begin{pmatrix} \mathbf{F}_1 & 0 & \dots & 0 \\ 0 & \mathbf{F}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{F}_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Lambda}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{\Lambda}_T \end{pmatrix} \begin{pmatrix} \mathbf{F}'_1 & 0 & \dots & 0 \\ 0 & \mathbf{F}'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{F}'_T \end{pmatrix}, \tag{20}$$

and from (9) consequently

$$\boldsymbol{\Omega}^+ = \begin{pmatrix} \mathbf{F}_1 & 0 & \dots & 0 \\ 0 & \mathbf{F}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{F}_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1^{-1} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Lambda}_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{\Lambda}_T^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{F}'_1 & 0 & \dots & 0 \\ 0 & \mathbf{F}'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{F}'_T \end{pmatrix}. \tag{21}$$

# 3 Estimation

The estimation problem is given by calculating $\mathbf{b}$ from (17) subject to the restrictions (15). Premultiplying (16) by $\mathbf{M}_t \equiv \begin{pmatrix} \mathbf{a}' \\ \boldsymbol{\Lambda}_t^{-1}\mathbf{F}'_t \end{pmatrix}$, produces the restrictions (4) and (5) and the reduced system

$$\mathbf{y}^*_{t,\bullet} = \mathbf{X}^*_{t,\bullet}\mathbf{b} + \mathbf{u}^*_{t,\bullet}, \qquad \forall t, \tag{22}$$

where $\mathbf{y}^*_{t,\bullet} \equiv \boldsymbol{\Lambda}_t^{-1}\mathbf{F}'_t\mathbf{y}_{t,\bullet}$, $\mathbf{X}^*_{t,\bullet} \equiv \boldsymbol{\Lambda}_t^{-1}\mathbf{F}'_t\mathbf{X}_{t,\bullet}$ and $\mathbf{u}^*_{t,\bullet} \equiv \boldsymbol{\Lambda}_t^{-1}\mathbf{F}'_t\mathbf{u}_{t,\bullet}$. Note that $\text{rk}(\mathbf{M}_t) = n$. Combining the equations of (22) leads to

$$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{u}^*, \tag{23}$$

with

$$\text{var}(\mathbf{u}^*) = \sigma^2\mathbf{I}_{T(n-1)} \tag{24}$$

and

$$\mathbf{X}^{*\prime}\mathbf{X}^* = \mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X} \qquad \text{and} \qquad \mathbf{X}^{*\prime}\mathbf{y}^* = \mathbf{X}'\boldsymbol{\Omega}^+\mathbf{y}. \tag{25}$$

OLS estimation of (23) subject to (15) yields the normal equations

$$\begin{pmatrix} \mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \nu \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\boldsymbol{\Omega}^+\mathbf{y} \\ \mathbf{h} \end{pmatrix}, \tag{26}$$

where $\mathbf{H} \equiv \begin{pmatrix} \mathbf{S} \\ \mathbf{R} \end{pmatrix}$, $\mathbf{h} \equiv \begin{pmatrix} \mathbf{c} \\ \mathbf{r} \end{pmatrix}$ and $\nu$ is the vector of Lagrange multipliers.

Oberhofer and Haupt (1999) have shown, that under assumptions one to nine, system (26) has a best linear unbiased solution

$$\hat{\mathbf{b}} = \mathbf{C}\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{y} + (\mathbf{I}_K - \mathbf{C}\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X})\mathbf{b}^*,$$

where $\mathbf{b}^*$ is an arbitrary vector fulfilling $\mathbf{H}\mathbf{b}^* = \mathbf{h}$ and $\mathbf{C} \equiv \mathbf{N}(\mathbf{N}'\mathbf{X}'\boldsymbol{\Omega}^+\mathbf{X}\mathbf{N})^{-1}\mathbf{N}'$, where $\mathbf{N}$ is a base of the null space on $\mathbf{H}$. The resulting estimator is denoted as modified GLS estimator.

# 4 A note on rank conditions

Theorem 6.6 of Theil (1971, p.285) provides a BLU representation of the constrained GLS estimator. The essential rank conditions for existence of this estimator used by Theil guarantee the invertibility of $\mathbf{X}'\mathbf{\Omega}^+\mathbf{X}$ in (26). Dhrymes and Schwarz (1987a), both using too restrictive conditions and no additional exogenous restrictions, show that the rank conditions postulated by Theil are not fulfilled for several practically relevant cases and provide existence and non-existence conditions for the constrained estimator. As a generalization of the considerations of Dhrymes and Schwarz (1987a) we derive necessary and sufficient conditions for Theils rank conditions to be met in the context of the introduced SUR model. Due to the representation of $\mathbf{\Omega}^+$ in (21) the nonsingularity of $\mathbf{X}'\mathbf{\Omega}^+\mathbf{X}$ is equivalent to a full column rank of the matrix

$$(27)\qquad \begin{pmatrix} \mathbf{F}_1'\mathbf{X}_{1,\bullet} \\ \mathbf{F}_2'\mathbf{X}_{2,\bullet} \\ \vdots \\ \mathbf{F}_T'\mathbf{X}_{T,\bullet} \end{pmatrix}.$$

The existence of a nonzero vector $\mathbf{d}$ fulfilling

$$(28)\qquad \mathbf{F}_t'\mathbf{X}_{t,\bullet}\mathbf{d} = \mathbf{0}, \qquad \forall t,$$

is equivalent to a column rank deficit of the matrix in (27). Due to the orthogonality of $\mathbf{a}$ and $\mathbf{F}_t$, (28) implies the existence of $T$ scalars $g_t$ with

$$(29)\qquad \mathbf{X}_{t,\bullet}\mathbf{d} = \mathbf{a}g_t, \qquad \forall t.$$

By conducting row manipulations (29) could be rendered to

$$(30)\qquad \mathbf{X}_{\bullet,i}\mathbf{d}_i = a_i\mathbf{g}, \qquad \forall i,$$

where $\mathbf{d}_i$ contains a suitable selection of the elements in $\mathbf{d}$.

Two cases must be distinguished:

(i) Let us assume the existence of an equation $j, (1 \leq j \leq n)$ with $\mathrm{rk}(\mathbf{X}_{\bullet,j}) < K_j$. Then there exists a nonzero vector $\mathbf{f}$ with $\mathbf{X}_{\bullet,j}\mathbf{f} = \mathbf{0}$. Now let $\mathbf{d}_i = \mathbf{0}$ for $i \neq j$, $\mathbf{d}_j = \mathbf{f}$ and $\mathbf{g} = \mathbf{0}$. Hence we have found a nonzero vector $\mathbf{d}$ fulfilling (28) and the rank criterium of Theil is violated.

(ii) If $\mathrm{rk}(\mathbf{X}_{\bullet,i}) = K_i$, $\forall i$, then (30) implies the existence of a nonzero vector $\mathbf{g}$, lying in the linear space spanned by the regressors of equation $i$, for every $i$ with $a_i \neq 0$. Then obviously it is possible to find a nonzero vector $\mathbf{h}_i$ for every $i$ with $a_i \neq 0$ such that $\mathbf{X}_{\bullet,i}\mathbf{h}_i = \mathbf{g}$. Finally, by setting $\mathbf{d}_i = a_i\mathbf{h}_i$ for every $i$ with $a_i \neq 0$ and $\mathbf{d}_i = \mathbf{0}$ for $a_i = 0$, again a nonzero $\mathbf{d}$ satisfying (28) is found.

Obvious from the preceding discussion is the proof of

LEMMA 1: The rank condition of Theil is violated, if there are collinearities in any of the $n$ equations, or, if there exists a nonzero linear combination $\mathbf{g}$ of the regressors of equation $i$ for all equations $i$ with nonzero weight.

Remark on Lemma 1: Obviously the conditions given in Lemma 1 are necessary and sufficient. Dhrymes and Schwarz (1987a, Theorem 2) prove this lemma by using the stronger condition that all $n$ equations have no common regressor. This is a consequence of their construction, assuming that the regressors of each equation are a selection of the $p$ basis vectors. However, in our considerations the regressors lie in a subspace of the linear space spanned by the $p$ basis vectors.

# 5 Identical regressors

In the existing literature often the case of identical regressors (see section 2.1 (iii)) in each equation is considered. Thus equation (1) is replaced by

$$y_{t,i} = \sum_{k=1}^{K} b_{k,i} x_{t,k} + u_{t,i} \qquad 1 \le i \le n, 1 \le t \le T. \tag{31}$$

Let $\mathbf{Y}$ be the $T \times n$ matrix $(y_{t,i})$, $\mathbf{X}$ is the $T \times K$ regressor matrix $(x_{t,k})$, $\mathbf{B}$ be the $K \times n$ matrix $(b_{k,i})$ and $\mathbf{U}$ is the $T \times n$ matrix of disturbances $(u_{t,i})$. Then, combining all observations, system (31) can be written in the form

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \tag{32}$$

and the adding-up restriction may be written as

$$\mathbf{Ya} = \mathbf{s}. \tag{33}$$

From (32) and (33) follows

$$\mathbf{Ya} = \mathbf{XBa} = \mathbf{s} \tag{34}$$

and $\mathbf{Ua} = \mathbf{0}$. Equation (34) implies the existence of a vector $\mathbf{c}$ with $\mathbf{Xc} = \mathbf{s}$. Since (34) should be identically fulfilled in $\mathbf{X}$, we get the parameter restrictions

$$\mathbf{Ba} = \mathbf{c}. \tag{35}$$

In the case of homoscedasticity and if $\mathbf{X}'\mathbf{X}$ has full rank, the modified GLS estimator is equivalent to the OLS estimator and the adding-up conditions are automatically fulfilled (see Worswick and Champernowne (1954-55)), considering equations (32) to (35)

$$\begin{aligned} \hat{\mathbf{B}}\mathbf{a} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Ya} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{XBa} \\ &= \mathbf{Ba} \\ &= \mathbf{c}. \end{aligned} \tag{36}$$

Note that the rank conditions of Theil (1971) are never fulfilled in this case, even if there are no collinearities among the $K$ regressors. In this case the $K$ non-redundant constraints (13) could be written as

$$\mathbf{Sb} = (\mathbf{a}' \otimes \mathbf{I}_K)\mathbf{b} = \mathbf{c}. \tag{37}$$

An immediate consequence is that (5) contains redundant constraints for $T > K$.

# 6 The deletion choice

For the special case that all elements of $\mathbf{a}$ are equal to one, Barten (1969), assuming normal distribution of errors, has shown that the $n-1$ linear independent equations of (1) contain the entire statistical information. Consequently one arbitrary equation could be dropped and then the remaining $n-1$ equations could be estimated. The parameters of the omitted equation could be estimated indirectly by rearranging restrictions (2). As proposed by Barten (1969), the remaining equations are estimated by ML.
What we are going to prove in the framework of the preceding discussion, is the invariance of the estimator to the deletion choice and the equivalence of the resulting estimator to the modified GLS estimator. This gives way to a representation of the estimator which needs no calculation of the Moore-Penrose inverse of $\mathbf{\Omega}$.

LEMMA 2: We assume $a_1 \neq 0$. Let $\mathbf{u}_{t,*}$ be the vector resulting from deletion of the first element of $\mathbf{u}_{t,\bullet}$ and let $\mathbf{\Omega}_{t,*}$ be the matrix obtained by deletion of the first column and row of $\mathbf{\Omega}_t$. Then

$$\mathbf{u}'_{t,*}\mathbf{\Omega}_{t,*}^{-1}\mathbf{u}_{t,*} = \mathbf{u}_{t,\bullet}\mathbf{\Omega}_t^{+}\mathbf{u}_{t,\bullet} = \mathbf{u}_{t,\bullet}\left[\mathbf{\Omega}_t + \mathbf{a}\mathbf{a}'\right]^{-1}\mathbf{u}_{t,\bullet}.$$

PROOF: Due to the definition of $\mathbf{u}_{t,*}$ we get $u_{t,1} = -\mathbf{a}'_*\mathbf{u}_{t,*}$ and consequently

$$\mathbf{u}_{t,\bullet} = \begin{pmatrix} -1 & -\mathbf{a}'_* \\ -\mathbf{a}_* & \mathbf{I}_{n-1} \end{pmatrix}\begin{pmatrix} 0 \\ \mathbf{u}_{t,*} \end{pmatrix} \equiv \mathbf{V}\begin{pmatrix} 0 \\ \mathbf{u}_{t,*} \end{pmatrix} \tag{38}$$

where $\mathbf{a}'_* \equiv \begin{pmatrix} a_2 & a_3 & \dots & a_n \end{pmatrix}/a_1$ and

$$\mathbf{\Omega}_t \equiv \mathrm{E}\mathbf{u}_{t,\bullet}\mathbf{u}'_{t,\bullet} = \mathbf{V}\begin{pmatrix} 0 & 0 \\ 0 & \mathbf{\Omega}_{t,*} \end{pmatrix}\mathbf{V}. \tag{39}$$

Obviously $\mathbf{V}$ is invertible, since $|\mathbf{V}| = -1 - (\sum_{i=2}^{n} a_i^2)/a_1^2 \neq 0$. Hence

$$\begin{aligned}
\mathbf{u}'_{t,*}\mathbf{\Omega}_{t,*}^{-1}\mathbf{u}_{t,*} &= \begin{pmatrix} 0 & \mathbf{u}_{t,*} \end{pmatrix}\begin{pmatrix} a_1^2 & 0 \\ 0 & \mathbf{\Omega}_{t,*} \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ \mathbf{u}_{t,*} \end{pmatrix} \\
&= \mathbf{u}_{t,\bullet}\left[\mathbf{V}\begin{pmatrix} a_1^2 & 0 \\ 0 & \mathbf{\Omega}_{t,*} \end{pmatrix}\mathbf{V}\right]^{-1}\mathbf{u}_{t,\bullet} \\
&= \mathbf{u}_{t,\bullet}\left[\mathbf{a}\mathbf{a}' + \mathbf{\Omega}_t\right]^{-1}\mathbf{u}_{t,\bullet}.
\end{aligned} \tag{40}$$

By definition of $\mathbf{u}_{t,\bullet}$ and $\mathbf{\Omega}_t^{+}$

$$\begin{aligned}
\mathbf{u}_{t,\bullet}\mathbf{\Omega}_t^{+}\mathbf{u}_{t,\bullet} &= \mathbf{u}_{t,\bullet}\left[\mathbf{a}\mathbf{a}' + \mathbf{\Omega}_t^{+}\right]\mathbf{u}_{t,\bullet} \\
&= \mathbf{u}_{t,\bullet}\left[\begin{pmatrix} \mathbf{a} & \mathbf{F}_t \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_t^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{a}' \\ \mathbf{F}'_t \end{pmatrix}\right]\mathbf{u}_{t,\bullet} \\
&= \mathbf{u}_{t,\bullet}\left[\begin{pmatrix} \mathbf{a} & \mathbf{F}_t \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_t \end{pmatrix}\begin{pmatrix} \mathbf{a}' \\ \mathbf{F}'_t \end{pmatrix}\right]^{-1}\mathbf{u}_{t,\bullet} \\
&= \mathbf{u}_{t,\bullet}\left[\mathbf{a}\mathbf{a}' + \mathbf{\Omega}_t\right]^{-1}\mathbf{u}_{t,\bullet}.
\end{aligned} \tag{41}$$

*q.e.d.*

Remark on Lemma 2: Note the computational advantage that $[\mathbf{\Omega}_t + \mathbf{aa}']^{-1}$ can be calculated instead of the Moore-Penrose inverse. Also note that $\mathbf{\Omega}_t + \mathbf{aa}'$ is always invertible if $\mathbf{aa}' \neq \mathbf{0}$.

Corollary: (i) Lemma 2 is valid for every deletion choice $i$ with $a_i \neq 0$.

(ii)The substitution of $\mathbf{X}'_{t,\bullet}\mathbf{\Omega}_t^+\mathbf{X}_{t,\bullet}$ by $\mathbf{X}'_{t,\bullet}[\mathbf{\Omega}_t + \mathbf{aa}']^{-1}\mathbf{X}_{t,\bullet}$ in the system of normal equations (26) is equivalent to the deletion of equation $i$ with $a_i \neq 0$. Then the constraints (13) solely serve to recover the deleted parameters.

Proof: (i) Obvious from the preceding discussion.

(ii) The system of normal equations (26) results from minimising

$$\sum_{t=1}^{T}(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b})'\mathbf{\Omega}_t^+(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b}), \tag{42}$$

subject to the constraints $\mathbf{Hb} = \mathbf{h}$.

Due to the constraints, $\mathbf{u}_{t,\bullet} = \mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b}$ is normal to $\mathbf{a}$. Thus, by Lemma 2,

$$\begin{aligned}\sum_{t=1}^{T}(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b})'\mathbf{\Omega}_t^+(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b}) = \sum_{t=1}^{T}\mathbf{u}'_{t,*}\mathbf{\Omega}_{t,*}^{-1}\mathbf{u}_{t,*} = \\ \sum_{t=1}^{T}(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b})'[\mathbf{\Omega}_t + \mathbf{aa}']^{-1}(\mathbf{y}_{t,\bullet} - \mathbf{X}_{t,\bullet}\mathbf{b})\end{aligned} \tag{43}$$

is identically fulfilled in $\mathbf{b}$.

*q.e.d.*

Remark on the weights $a_i$:

- In the case of $a_1 \neq 0$, the covariance matrix of the reduced system is nonsingular. Thus we see that, when postmultiplying with $\mathbf{M}_t$, one equation is substituted by restrictions on the parameters. Since $a_1 \neq 0$ implies the nonsingularity of $\mathbf{M}_t$, there is whether a loss of given information nor a gain of new information. Hence, the resulting nonsingular system could be estimated and the remaining parameters (of the deleted equation) can be recovered by the parameter restrictions. This, of course, also holds in the case of further restrictions, e.g. homogeneity or symmetry restrictions. For details see Bewley (1986, chapter 3).
- In the case of $a_1 = 0$, the endogenous variable of the first equation is excluded from the adding-up condition. Obviously the reduced system is singular in this case.
- Usually in the literature all elements of $\mathbf{a}$ are restricted to unity, thus $\mathbf{a} \equiv \mathbf{e}_n$ and we get

  $$(\mathbf{I}_T \otimes \mathbf{e}_n)\mathbf{u} = \mathbf{0}. \tag{44}$$

  Exceptions are given by Powell (1969) and Dhrymes and Schwarz (1987a), who considered weighted aggregates. However, using nontrivial weights should not be considered an extra-generality. An imaginable example could be seen (all readers inclined in finance may forgive us) in the (ex post) regression of asset risk premia on market risk premium (which is equal to a weighted sum of the endogenous variable).

# 7 A glimpse at feasible GLS

We now turn to the practically relevant case of an unknown covariance matrix $\mathbf{\Omega}_t$. Usually the problem of unknown covariance is treated by iteratively applying feasible GLS. Since

$$\mathbf{\Omega}_t = \mathbf{F}_t \mathbf{\Lambda}_t \mathbf{F}_t'$$

it is near at hand to assume for the first iteration step

$$\mathbf{\Lambda}_t^{(1)} = \mathbf{I}_{n-1}$$

to get

$$\mathbf{\Omega}_t^{(1)} = \mathbf{F}_t \mathbf{I}_{n-1} \mathbf{F}_t' = \mathbf{I}_n - \mathbf{aa}',$$

which preserves all correlations caused by the adding-up restrictions. However, if the constraints are considered, due to Lemma 2 the first step estimator is given by

$$[\mathbf{I}_n - \mathbf{aa}' + \mathbf{aa}']^{-1} = \mathbf{I}_n,$$

which brings us to the nice result, that we obtain in the first step the OLS estimation subject to the constraints. Note that the restrictions are automatically fulfilled for homoscedasticity and identical regressors (remember section 5).

# References


[1] Barten A.P. (1969). "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Economic Review* 1: 7-73.

[2] Bewley R. (1986). *Allocation Models: Specification, Estimation and Applications.* Cambridge, Mass.: Ballinger.

[3] Dhrymes P.J. and S. Schwarz (1987a). "On the Existence of Generalized Inverse Estimators in a Singular System of Equations." *Journal of Forecasting* 6: 181-192.

[4] Dhrymes P.J. and S. Schwarz (1987b). "On the Invariance of Estimators for Singular Systems of Equations." *Greek Economic Review* 9 (1): 88-108.

[5] Kreijger R.G. and H. Neudecker (1977). "Exact Linear Restrictions on Parameters in the General Linear Model with a Singular Covariance Matrix." *Journal of the American Statistical Association* 72: 430-432.

[6] Oberhofer W. and H. Haupt (2000). "Best Linear Unbiased Estimates in the Constrained Singular Linear Model", (submitted to *Econometric Theory*).

[7] Powell A.A. (1969). "Aitken Estimators as a Tool in Allocating Predetermined Aggregates." *Journal of the American Statistical Association* 64: 913-922.

[8] Theil H. (1971). *Principles of Econometrics.* New York: J. Wiley & Sons.

[9] Worswick G.D.N. and D.G. Champernowne (1954-55). "A Note on the Adding-up Criterion." *Review of Economic Studies* 22: 57-60.