AMSTER

18/89

FACULTY

~~INSTITUTE~~ OF ACTUARIAL SCIENCE & ECONOMETRICS

REPORT AE 18/89

R^2 in Seemingly Unrelated Regression Equations

H. Neudecker
F.A.G. Windmeijer

University of Amsterdam

Title: R^2 in Seemingly Unrelated Regression Models

Authors: H. Neudecker
F.A.G. Windmeijer

Address: University of Amsterdam
Institute of Actuarial Science and Econometrics
Jodenbreestraat 23
1011 NH Amsterdam
The Netherlands

Date: November 1989

Series and Number: Report AE 18/89

Pages: 18

Price: No charge

JEL Subject Classification: 211

Keywords: SURE-model, goodness of fit, correlation coefficient, asymptotic properties.

Abstract: In this paper we will discuss some properties of McElroy's measure of goodness of fit for Zellner's seemingly unrelated regression equations (McElroy (1977)). Amongst them are asymptotic properties.
The same will be done for another goodness-of-fit measure, the squared sample correlation coefficient of $(\Omega^{-\frac{1}{2}}\otimes I)\underline{y}$ and $(\Omega^{-\frac{1}{2}}\otimes I)\hat{\underline{y}}$ where $\hat{\underline{y}}$ is the theoretical value of the dependent variable $\underline{y}$ and $\Omega\otimes I$ is the variance of $\underline{y}$. A comparison will be made between the two measures and it turns out that McElroy's measure possesses more desirable characteristics in case all equations contain a constant term.

R^2 in Seemingly Unrelated Regression Equations.

H. Neudecker
F.A.G.Windmeijer
Instituut voor Actuariaat en Econometrie
Jodenbreestraat 23
NL-1011 NH Amsterdam
The Netherlands

In this paper we will discuss some properties of McElroy's measure of goodness of fit for Zellner's seemingly unrelated regression equations (McELROY (1977)). Amongst them are asymptotic properties.
The same will be done for another goodness-of-fit measure, the squared sample correlation coefficient of $(\Omega^{-1/2}\otimes I)\underline{y}$ and $(\Omega^{-1/2}\otimes I)\hat{\underline{y}}$ where $\hat{\underline{y}}$ is the theoretical value of the dependent variable $\underline{y}$ and $\Omega\otimes I$ is the variance of $\underline{y}$. A comparison will be made between the two measures and it turns out that McElroy's measure possesses more desirable characteristics in case all equations contain a constant term.

Key Words & Phrases: SURE-model, goodness of fit, correlation coefficient, asymptotic properties.

1. Introduction

Some ten years ago McELROY(1977) presented a measure of goodness of fit for Zellner's seemingly unrelated regression equations.
In this paper we will discuss some properties of this measure, some of which are known, some of which are new e.g. that it is the sample correlation coefficient of $(\Omega^{-1/2}\otimes N)\underline{y}$ and $(\Omega^{-1/2}\otimes N)\hat{\underline{y}}$ where $\hat{\underline{y}}$ is the theoretical value of the dependent variable $\underline{y}$ and $\Omega\otimes I$ is the variance of $\underline{y}$, and asymptotic properties.
Next we will look at the properties of another goodness-of-fit measure, the squared sample correlation coefficient of $(\Omega^{-1/2}\otimes I)\underline{y}$ and $(\Omega^{-1/2}\otimes I)\hat{\underline{y}}$. McElroy wrongly stated that her measure is this correlation coefficient. A comparison will be made between the two measures and it turns out that McElroy's measure possesses more desirable characteristics in case all equations contain a constant term.

In section 2 Zellner's SURE-model is given and the assumptions underlying the model are stated.
Section 3 gives the definition of McElroy's measure (R_z^2) and discusses its properties.
In section 4 the properties of the squared sample correlation coefficient between $(\Omega^{-\frac{1}{2}}\otimes I)\underline{y}$ and $(\Omega^{-\frac{1}{2}}\otimes I)\hat{\underline{y}}$ are given.
In the concluding section 5 a comparison between the two measures is made.

2. Model and assumptions

Zellner's model consists of n observations on g seemingly unrelated stochastic equations written as

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \underline{y}_g \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ 0 & 0 & & X_g \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_g \end{bmatrix} + \begin{bmatrix} \underline{\varepsilon}_1 \\ \underline{\varepsilon}_2 \\ \cdot \\ \cdot \\ \underline{\varepsilon}_g \end{bmatrix} \tag{1}$$

where for the j-th equation $\underline{y}_j$ is nx1, X_j is nxk_j of rank k_j and fixed, β_j is k_jx1 and unknown, and $\underline{\varepsilon}_j$ is nx1 and stochastic with mean zero.
Furthermore it is assumed that every equation contains a constant term, so X_j can be partitioned as (s_n, Z_j), with $s_n = (1,1,...,1)'$, for all j.
We shall write (1) in compact form as

$$\underline{y} = X\beta + \underline{\varepsilon}, \tag{2}$$

where $\underline{y}$ and $\underline{\varepsilon}$ are ngx1, X is ngxk, β is kx1 and $k=\sum_{j=1}^{g} k_j$.
For $\underline{\varepsilon}$ we have $E(\underline{\varepsilon})=0$ and

$$D(\underline{\varepsilon}) = \Omega \otimes I_n,$$

Ω being the gxg positive definite contemporaneous variance.
For simplicity we assume that Ω is known. If not, it can be replaced in all relevant formulae by a consistent estimator, e.g. $\hat{\underline{\Omega}} = \frac{1}{n}\underline{E}'\underline{E}$, where $\underline{E} = (\underline{e}_1,...,\underline{e}_g)$ and $\underline{e}_j$ is the LS residual of the j^{th} equation.
Following McElroy we rewrite (2) as

$$\underline{y} = Z\beta_z + W\beta_w + \underline{\varepsilon}, \tag{3}$$

$$Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & & . \\ . & & . & . \\ . & & & . \\ 0 & \dots & & Z_g \end{bmatrix}, \qquad W = I_g \otimes s, \tag{4}$$

The estimated counterparts of (2) and (3) will be written as

$$\underline{y} = X\underline{b} + \underline{e} = Z\underline{b}_z + W\underline{b}_w + \underline{e}, \tag{5}$$

where

$$\underline{b} = (X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)\underline{y}. \tag{6}$$

The theoretical value of $\underline{y}$, $\hat{\underline{y}}$, is given by

$$\hat{\underline{y}} = X\underline{b} = Z\underline{b}_z + W\underline{b}_w. \tag{7}$$

3. Definition and properties of McElroy's R_z^2

McElroy defines as a measure of goodness of fit for the estimated model

$$R_z^2 = \frac{\underline{b}_z'Z'(\Omega^{-1}\otimes N_n)Z\underline{b}_z}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}} = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}}$$

where

$$N_n = I_n - \tfrac{1}{n}s_n s_n', \tag{8}$$

and the second equality holds because $W'(\Omega^{-1}\otimes N_n)=0$ by virtue of (4), (7) and (8).

R_z^2 can be seen as the ratio of the estimated weighted variation and the total weighted variation in $\underline{y}$, because $\Omega^{-1}\otimes N_n = (I_g\otimes N_n)(\Omega^{-1}\otimes I_n)(I_g\otimes N_n)$.

Properties of R_z^2 are:

i) $0 \le R_z^2 \le 1$, $\quad R_z^2 = 1$ if $\hat{\underline{y}} = \underline{y}$,
$R_z^2 = 0$ if $\underline{b}_z = 0$.

ii) R_z^2 has a one to one relation with the F test statistic for testing the hypothesis that all coefficients except the constant terms (β_w) are 0 :

$$F_{k-g,ng-k} = \frac{R_z^2}{1-R_z^2} \cdot \frac{ng-k}{k-g}.$$

iii) R_z^2 is a generalization of Buse's definition of R^2 in the univariate GLS-model (BUSE (1973), (1979)).
Consider a GLS-model

$$\underline{y} = X\beta + \underline{\varepsilon}$$

with $E(\underline{\varepsilon})=0$, $D(\underline{\varepsilon})=\sigma^2 V$ and $X=(s,Z)$.
Buse's definition of R^2 for this model is given by

$$R^2_{Bu} = \frac{(\hat{\underline{y}}-s\underline{b}_0^*)'V^{-1}(\hat{\underline{y}}-s\underline{b}_0^*)}{(\underline{y}-s\underline{b}_0^*)'V^{-1}(\underline{y}-s\underline{b}_0^*)}$$

where

$$\underline{b}_0^* = (s'V^{-1}s)^{-1}s'V^{-1}\underline{y} = (s'V^{-1}s)^{-1}s'V^{-1}\hat{\underline{y}}$$

is the estimator of the constant term under the restriction that all other coefficients are 0.
A generalization of this measure for the SURE-model is

$$R^2_{Bu} = \frac{(\hat{\underline{y}}-W\underline{b}_w^*)'(\Omega^{-1}\otimes I_n)(\hat{\underline{y}}-W\underline{b}_w^*)}{(\underline{y}-W\underline{b}_w^*)'(\Omega^{-1}\otimes I_n)(\underline{y}-W\underline{b}_w^*)}$$

where

$$\underline{b}_w^* = (W'(\Omega^{-1}\otimes I_n)W)^{-1}W'(\Omega^{-1}\otimes I_n)\underline{y} = \tfrac{1}{n}(I_g\otimes s_n')\underline{y},$$

is the estimator of β_w under the restriction β_z=0.
That R_{Bu}^2 for the SURE-model equals R_z^2 can be seen as follows. We have

$$W\underline{b}_w^* = \tfrac{1}{n}(I_g\otimes s_n s_n')\underline{y} = \tfrac{1}{n}(I_g\otimes s_n s_n')\hat{\underline{y}},$$

because of (A.3). (For (A.) see the appendix.)
Hence,

$$\hat{\underline{y}} - W\underline{b}_w^* = (I_g\otimes N_n)\hat{\underline{y}} \quad ; \quad \underline{y} - W\underline{b}_w^* = (I_g\otimes N_n)\underline{y},$$

and therefore

$$R_{Bu}^2 = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}} = R_z^2.$$

iv) R_z^2 is the squared sample correlation coefficient of $(\Omega^{-\frac{1}{2}}\otimes N_n)\underline{y}$ and $(\Omega^{-\frac{1}{2}}\otimes N_n)\hat{\underline{y}}$. Denoting this correlation by r_z^2 we have

$$r_z^2 = \frac{(\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes N_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes N_n)\underline{y})^2}{\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes N_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes N_n)\hat{\underline{y}} \cdot \underline{y}'(\Omega^{-\frac{1}{2}}\otimes N_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes N_n)\underline{y}}. \tag{9}$$

Because $(\Omega^{-\frac{1}{2}}\otimes N_n)\underline{y} = (I_g\otimes N_n)(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{y}$ consists of g n-vectors, all measured as deviations from their means we have

$$\underline{y}'(\Omega^{-\frac{1}{2}}\otimes N_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes N_n)\underline{y} = \underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}.$$

Clearly the same holds for the other two quadratic forms.
Further

$$\begin{aligned}\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\underline{y} &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)(\hat{\underline{y}}+\underline{e}) \\ &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} + \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\underline{e} \\ &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} + \hat{\underline{y}}'(\Omega^{-1}\otimes I_n)(I_g\otimes N_n)\underline{e} \\ &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}\end{aligned}$$

because of (A.4) and (A.1).
Therefore

$$r_z^2 = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}} = R_z^2.$$

v) R_z^2 is invariant under changes of location and changes of scale of the dependent variable.

Consider a change of location of $\underline{y}$ given by

$$\underline{y}^* = \underline{y} + (I_g\otimes s_n)\mu \quad , \quad \mu = (\mu_1,\ldots,\mu_g)'.$$

The theoretical value of $\underline{y}^*$ is then (in the following $\hat{\underline{y}}$ and $\underline{b}$ are the theoretical value of $\underline{y}$ and the estimator of β respectively in the untransformed model)

$$\begin{aligned}\hat{\underline{y}}^* &= X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)(\underline{y}+(I_g\otimes s_n)\mu)\\ &= \hat{\underline{y}} + (I_g\otimes s_n)\mu.\end{aligned}$$

Because $(I_g\otimes N_n)(I_g\otimes s_n) = 0$ it follows that

$$R_z^{2*} = \frac{\hat{\underline{y}}^{*\prime}(\Omega^{-1}\otimes N_n)\hat{\underline{y}}^*}{\underline{y}^{*\prime}(\Omega^{-1}\otimes N_n)\underline{y}^*} = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}} = R_z^2.$$

Next, consider a change of scale of $\underline{y}$ given by

$$\underline{y}^* = (\Lambda\otimes I_n)\underline{y},$$

with Λ a gxg diagonal matrix.
If we define the matrix

$$A = \begin{bmatrix} \lambda_1 I_{k_1} & 0 & \cdots & 0 \\ 0 & \lambda_2 I_{k_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \lambda_g I_{k_g} \end{bmatrix},$$

then because of the block-diagonality of X it is easily seen that we have

$$(\Lambda\otimes I_n)X = XA.$$

Model (2) now becomes

$$\begin{aligned}\underline{y}^* &= (\Lambda\otimes I_n)X\beta + (\Lambda\otimes I_n)\underline{\varepsilon},\\ &= XA\beta + \underline{\varepsilon}^*\\ &= X\beta^* + \underline{\varepsilon}^*\end{aligned}$$

with $\beta^*=A\beta$, X as before, $E(\underline{\varepsilon}^*)=0$ and

$$D(\underline{\varepsilon}^*) = (\Lambda\otimes I_n)(\Omega\otimes I_n)(\Lambda\otimes I_n) = \Lambda\Omega\Lambda\otimes I_n.$$

Clearly, the estimator $\underline{b}^*$ of β^* is $A\underline{b}$ and consequently

$$\hat{\underline{y}}^* = XA\underline{b} = (\Lambda\otimes I_n)\hat{\underline{y}}.$$

Finally we obtain

$$R_z^{2*} = \frac{\hat{\underline{y}}^{*\prime}(\Lambda^{-1}\Omega^{-1}\Lambda^{-1}\otimes N_n)\hat{\underline{y}}^*}{\underline{y}^{*\prime}(\Lambda^{-1}\Omega^{-1}\Lambda^{-1}\otimes N_n)\underline{y}^*} = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y}} = R_z^2.$$

vi) Asymptotic properties of R_z^2.

We shall investigate the asymptotic properties of R_z^2. It can be considered to be the estimator of a sort of population correlation coefficient. The procedure will be inspired by the approach of HEIJMANS and NEUDECKER(1987). This itself relies strongly on certain properties of characteristic roots. A useful tool for asymptotic results is the following lemma:

Lemma 1; Lukacs' Lemma : If the sequence $\{\underline{z}_n\}$ of random variables is bounded, then $\text{plim}\underline{z}=z$ implies

$$E\,|\,\underline{z}_n - z\,|^r \to 0 \text{ as } n\to\infty \text{ for all } r>0.$$

Proof : see LUKACS(1975, p.38).

We shall make the following assumptions relating to the SURE-model :

i) There exist m,M with $0<m<M<\infty$ such that $m \leq \lambda_{1n} \leq \ldots \leq \lambda_{kn} \leq M$, where $(\lambda_{1n},\ldots,\lambda_{kn})$ are the characteristic roots of $\frac{1}{n}X'X$.
This means that $\frac{1}{n}X'X$ remains a finite matrix of full rank.

ii) $\text{plim}\ \frac{1}{n}\underline{\varepsilon}'(\Omega^{-1}\otimes N_n)\underline{\varepsilon} = g$.

This assumption is inspired by the following facts :

$$E((n-1)^{-1}\underline{\varepsilon}'(\Omega^{-1}\otimes N_n)\underline{\varepsilon}) = g$$

$$D((n-1)^{-1}\underline{\varepsilon}'(\Omega^{-1}\otimes N_n)\underline{\varepsilon}) = 2g/(n-1),$$

in case $\underline{\varepsilon}$ is normally distributed.

It is now possible to prove the following result :

__Lemma 2 :__ If assumptions i) and ii) hold then

(1) $\quad \text{plim}\ \frac{1}{n}(\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} - \beta'X'(\Omega^{-1}\otimes N_n)X\beta) = 0$;

(2) $\quad \text{plim}\ \frac{1}{n}(\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y} - \beta'X'(\Omega^{-1}\otimes N_n)X\beta - ng) = 0.$

If additionally we also assume

$$\lim_{n\to\infty} \frac{1}{n}Z'(\Omega^{-1}\otimes N_n)Z = H,$$

then

$$\text{plim}\ R_z^2 = \frac{\beta_z'H\beta_z}{\beta_z'H\beta_z+g},$$

and so

$$\text{plim } R_z^2 = 0 \quad \text{if} \quad \beta_z = 0$$
$$\text{plim } R_z^2 \to 1 \quad \text{if} \quad \Omega \to 0.$$

$$\text{Further } \lim_{n\to\infty} E(R_z^2) = \frac{\beta_z' H \beta_z}{\beta_z' H \beta_z + g} \quad \text{and} \quad D(R_z^2) \to 0 \text{ as } n \to \infty.$$

Proof : Consider

$$\tfrac{1}{n}(\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} - \beta' X'(\Omega^{-1}\otimes N_n)X\beta)$$
$$= \tfrac{2}{n}\beta' X'(\Omega^{-1}\otimes N_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)\underline{\varepsilon}$$
$$+ \tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-1}\otimes I_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes N_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)\underline{\varepsilon}.$$

The variance of the first right-hand side term equals

$$4/n^2\ \beta' X'(\Omega^{-1}\otimes N_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1} \cdot X'(\Omega^{-1}\otimes N_n)X\beta$$
$$= 4/n^2\ \beta' X'(\Omega^{-1}\otimes N_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes N_n)X\beta$$
$$= 4/n^2\ \beta' X'(\Omega^{-\frac{1}{2}}\otimes N_n)(\Omega^{-\frac{1}{2}}\otimes I_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-\frac{1}{2}}\otimes I_n)(\Omega^{-\frac{1}{2}}\otimes N_n)X\beta$$
$$\leq 4/n^2\ \beta' X'(\Omega^{-\frac{1}{2}}\otimes N_n)(\Omega^{-\frac{1}{2}}\otimes N_n)X\beta = 4/n^2\ \beta' X'(\Omega^{-1}\otimes N_n)X\beta,$$

for the greatest characteristic root of $(\Omega^{-\frac{1}{2}}\otimes I_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-\frac{1}{2}}\otimes I_n)$ is 1.
If then w_g is the greatest characteristic root of Ω^{-1} (and hence of $\Omega^{-1}\otimes N_n$), we get

$$4/n^2\ \beta' X'(\Omega^{-1}\otimes N_n)X\beta \leq 4/n^2 \cdot w_g \cdot \beta' X' X\beta \leq 4/n \cdot w_g \cdot \lambda_{kn}\beta'\beta \to 0 \quad \text{as } n\to\infty$$

by assumption i).
So the variance of the first right-hand side term approaches zero as $n\to\infty$.

The second right-hand side term is

$$\tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-1}\otimes I_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes N_n)X(X'(\Omega^{-1}\otimes I_n)X)^{-1}X'(\Omega^{-1}\otimes I_n)\underline{\varepsilon}$$
$$= (\tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-1}\otimes I_n)X)(\tfrac{1}{n}X'(\Omega^{-1}\otimes I_n)X)^{-1}(\tfrac{1}{n}X'(\Omega^{-1}\otimes N_n)X)(\tfrac{1}{n}X'(\Omega^{-1}\otimes I_n)X)^{-1} \cdot (\tfrac{1}{n}X'(\Omega^{-1}\otimes I_n)\underline{\varepsilon}).$$

Now

$$D(\tfrac{1}{n}X'(\Omega^{-1}\otimes I_n)\underline{\varepsilon}) = 1/n^2\ X'(\Omega^{-1}\otimes I_n)X \rightarrow 0 \quad \text{as} \quad n\rightarrow\infty.$$

This can be proved as follows.
Consider the quadratic form $1/n^2\ a'X'(\Omega^{-1}\otimes I_n)Xa$ with arbitrary a. Then

$$1/n^2\ a'X'(\Omega^{-1}\otimes I_n)Xa \le 1/n^2 \cdot w_g \cdot a'X'Xa \le 1/n \cdot w_g \cdot \lambda_{kn} \cdot a'a \rightarrow 0 \quad \text{as} \quad n\rightarrow\infty$$

by assumption i).
Hence $1/n^2\ X'(\Omega^{-1}\otimes I_n)X \rightarrow 0$ as $n\rightarrow\infty$.

Further it follows that $\frac{1}{n}X'(\Omega^{-1}\otimes N_n)X$ is finite, therefore the probability limit of the second right-hand side term is zero.
This establishes the proof of (1).
The proof of (2) goes in the same manner, as

$$\begin{aligned}&\tfrac{1}{n}(\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y} - \beta'X'(\Omega^{-1}\otimes N_n)X\beta - ng)\\ &= \tfrac{2}{n}\beta'X'(\Omega^{-1}\otimes N_n)\underline{\varepsilon} + \tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-1}\otimes N_n)\underline{\varepsilon} - g.\end{aligned}$$

Clearly

$$D(2/n\ \beta'X'(\Omega^{-1}\otimes N_n)\underline{\varepsilon}) = 4/n^2\ \beta'X'(\Omega^{-1}\otimes N_n)X\beta \rightarrow 0 \quad \text{as} \quad n\rightarrow\infty.$$

Using assumption ii) we finish the proof of (2).
The additional results follow immediately, partly from Lukacs' Lemma.

∎

4. The squared sample correlation coefficient of $(\Omega^{-1/2}\otimes I)\underline{y}$ and $(\Omega^{-1/2}\otimes I)\hat{\underline{y}}$

McElroy states that R_z^2 is the squared sample correlation coefficient of $(\Omega^{-\frac{1}{2}}\otimes I)\underline{y}$ and $(\Omega^{-\frac{1}{2}}\otimes I)\hat{\underline{y}}$ (McElroy(1977, p.384)). We will show that this is a wrong statement. But could this correlation coefficient be an alternative goodness-of-fit measure? Some properties of this measure will be discussed in this section and a comparison will be made between the two measures in the next section. It turns out that R_z^2 possesses more desirable properties in case all equations contain a constant term.

Let us denote the correlation of $(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{y}$ and $(\Omega^{-\frac{1}{2}}\otimes I_n)\hat{\underline{y}}$ by R^2. Then

$$R^2 = \frac{(\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{y})^2}{\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\hat{\underline{y}} \cdot \underline{y}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{y}}, \qquad (10)$$

with

$$N_{ng} = I_g\otimes N_n + N_g\otimes(I_n-N_n). \qquad (11)$$

We can rewrite the numerator of (10) as the square of

$$\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\hat{\underline{y}} + \hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{e} = \hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\hat{\underline{y}},$$

because

$$\begin{aligned}(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\underline{e} &= (\Omega^{-1}\otimes I_n)(I_g\otimes N_n)\underline{e} + (\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes I_n)(I_g\otimes(I_n-N_n))\underline{e} \\ &= (\Omega^{-1}\otimes I_n)\underline{e}\end{aligned}$$

and

$$\hat{\underline{y}}'(\Omega^{-1}\otimes I_n)\underline{e} = 0,$$

by virtue of (7), (11), (A.1) and (A.4).
We can further derive

$$\begin{aligned}\hat{\underline{y}}'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)\hat{\underline{y}} &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} + \hat{\underline{y}}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\hat{\underline{y}} \\ &= \hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} + \underline{y}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{y},\end{aligned}$$

by employing (5), (7), (11) and (A.4).

So eventually we find

$$R^2 = \frac{\hat{\underline{y}}'(\Omega^{-1}\otimes N_n)\hat{\underline{y}} + \underline{y}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{y}}{\underline{y}'(\Omega^{-1}\otimes N_n)\underline{y} + \underline{y}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{y}}.$$

Let us introduce the definitions

$$\underline{Y} := (\underline{y}_1,...,\underline{y}_g) \text{ and } \hat{\underline{Y}} := (\hat{\underline{y}}_1,...\hat{\underline{y}}_g),$$

so that $\underline{y}=\text{vec}\underline{Y}$ and $\hat{\underline{y}}=\text{vec}\hat{\underline{Y}}$.

We can then establish the alternative expressions for R_z^2 and R^2

$$R_z^2 = \frac{\text{tr}\Omega^{-1}\hat{\underline{Y}}'N_n\hat{\underline{Y}}}{\text{tr}\Omega^{-1}\underline{Y}'N_n\underline{Y}}$$

and

$$R^2 = \frac{\text{tr}\Omega^{-1}\hat{\underline{Y}}'N_n\hat{\underline{Y}} + \frac{1}{n}s_n'\underline{Y}\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\underline{Y}'s_n}{\text{tr}\Omega^{-1}\underline{Y}'N_n\underline{Y} + \frac{1}{n}s_n'\underline{Y}\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\underline{Y}'s_n}.$$

The direct relationship between R_z^2 and R^2 is then given by

$$R^2 = \frac{R_z^2+\lambda}{1+\lambda} \tag{12}$$

where

$$\lambda = \frac{1}{n}\,\frac{s_n'\underline{Y}\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\underline{Y}'s_n}{\text{tr}\Omega^{-1}\underline{Y}'N_n\underline{Y}},$$

and it is easily seen that

$$0 \leq R_z^2 \leq R^2 \leq 1.$$

Properties of R^2 are:

i) $0 \leq R^2 \leq 1$, $\quad R^2 = 1$ if $\hat{\underline{y}} = \underline{y}$,
$R^2 = 0$ if $\underline{b}_z = 0$ **and** $N_g\Omega^{-\frac{1}{2}}\underline{Y}'s_n = 0$.

ii) The relation of R^2 and the F test statistic for testing the hypothesis $\beta_w=0$ is given by

$$F_{k-g,ng-k} = \frac{1}{1-R^2}\left(R^2 - \frac{\lambda}{1+\lambda}\right) \cdot \frac{ng-k}{k-g},$$

by virtue of (12) and property ii) of R_z^2.

iii) R^2 is not invariant under changes of location and changes of scale of the dependent variable.

Consider a change of location of $\underline{y}$ given by

$$\underline{y}^* = \underline{y} + (I_g \otimes s_n)\mu \quad , \quad \mu = (\mu_1, \dots, \mu_g)'.$$

We know from property v) of R_z^2 that

$$\hat{\underline{y}}^* = \hat{\underline{y}} + (I_g \otimes s_n)\mu.$$

In this case, because

$$(I_g \otimes s_n')(\Omega^{-\frac{1}{2}} N_g \Omega^{-\frac{1}{2}} \otimes (I_n - N_n)) = ((\Omega^{-\frac{1}{2}} N_g \Omega^{-\frac{1}{2}} \otimes s_n') \neq 0,$$

it follows that $R^{2*} \neq R^2$.

Next consider a change of scale of $\underline{y}$ given by

$$\underline{y}^* = (\Lambda \otimes I_n)\underline{y},$$

with Λ a gxg diagonal matrix.
As we know from property v) of R_z^2 we have

$$D(\underline{y}^*) = \Lambda\Omega\Lambda \otimes I_n \,;$$
$$\hat{\underline{y}}^* = (\Lambda \otimes I_n)\hat{\underline{y}}$$

and so

$$R^{2*} = \frac{\hat{\underline{y}}'(\Omega^{-1} \otimes N_n)\hat{\underline{y}} + \underline{y}'(\Lambda^{\frac{1}{2}}\Omega^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}N_g\Lambda^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}\Lambda^{\frac{1}{2}} \otimes (I_n - N_n))\underline{y}}{\underline{y}'(\Omega^{-1} \otimes N_n)\underline{y} + \underline{y}'(\Lambda^{\frac{1}{2}}\Omega^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}N_g\Lambda^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}\Lambda^{\frac{1}{2}} \otimes (I_n - N_n))\underline{y}} \neq R^2.$$

iv) Asymptotic properties of R^2.

<u>Lemma 3 :</u> If assumption i) holds then

$$\text{plim } \tfrac{1}{n}(\underline{y}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{y} - \beta'X'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))X\beta) = 0.$$

If further assumption ii) holds and the additional assumption

$$\lim_{n\to\infty} \tfrac{1}{n}X'(\Omega^{-\frac{1}{2}}\otimes I_n)N_{ng}(\Omega^{-\frac{1}{2}}\otimes I_n)X = G,$$

then

$$\text{plim } R^2 = \frac{\beta'G\beta}{\beta'G\beta+g},$$

and so

$$\text{plim } R^2 = 0 \quad \text{if} \quad \beta = 0,$$
$$\text{plim } R^2 \to 1 \quad \text{if} \quad \Omega \to 0.$$

Further $\lim_{n\to\infty}E(R^2) = \dfrac{\beta'G\beta}{\beta'G\beta+g}$ and $D(R^2)\to 0$ as $n\to\infty$.

Proof : Consider

$$\tfrac{1}{n}(\underline{y}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{y} - \beta'X'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))X\beta)$$
$$= \tfrac{2}{n}\beta'X'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\varepsilon + \tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))\underline{\varepsilon}.$$

It is easy to see that the variance of the first right-hand term equals

$$4/n^2\ \beta'X'(\Omega^{-\frac{1}{2}}N_g\Omega^{-\frac{1}{2}}\otimes(I_n-N_n))X\beta$$
$$= 4/n^2\ \beta'X'(\Omega^{-\frac{1}{2}}\otimes I_n)(N_g\otimes(I_n-N_n))(\Omega^{-\frac{1}{2}}\otimes I_n))X\beta$$
$$\leq 4/n^2\ \beta'X'(\Omega^{-1}\otimes I_n)X\beta$$
$$\leq 4/n\ w_g\lambda_{kn}\beta'\beta \to 0 \quad \text{as } n\to\infty$$

by assumption i) and using the fact that all characteristic roots of $N_g \otimes (I_n - N_n)$ are less than or equal to one.

The second right-hand side term is

$$\tfrac{1}{n}\underline{\varepsilon}'(\Omega^{-\frac{1}{2}} \otimes I_n)(N_g \otimes (I_n - N_n))(\Omega^{-\frac{1}{2}} \otimes I_n)\underline{\varepsilon}.$$

If we define $(\underline{u}_1', \ldots, \underline{u}_g') = \underline{u}' = \underline{\varepsilon}'(\Omega^{-\frac{1}{2}} \otimes I_n)$, we can write this term as

$$1/n^2 \sum_{i=1}^{g} \sum_{j=1}^{g} (\delta_{ij} - 1/g)\underline{u}_i' s_n s_n' \underline{u}_j,$$

where δ_{ij} equals 1 if i=j, zero otherwise.
Now

$$E(\tfrac{1}{n} s_n' \underline{u}_i) = 0 \qquad i=1,\ldots,g$$
$$D(\tfrac{1}{n} s_n' \underline{u}_i) = \tfrac{1}{n}, \qquad i=1,\ldots,g$$

so $\text{plim}(\frac{1}{n} s_n' \underline{u}_i) = 0$ and therefore the probability limit of the second right-hand side term is zero.
The additional results follow from the results of Lemma 2 and Lemma 1. ◻

5. Conclusions

If we finally compare the properties of the two measures we see that R_z^2 possesses a vital property viz. that it is zero if $\underline{b}_z$ is zero and furthermore plim $R_z^2=0$ if $\beta_z=0$.
For R^2, if $\underline{b}_z=0$, we have by (12)

$$R^2 = \frac{\lambda}{1+\lambda}$$

which is larger than zero in general.

If we partition G like X as

$$G = \begin{bmatrix} G_{ww} & G_{wz} \\ G_{zw} & G_{zz} \end{bmatrix}$$

we have for the probability limit of R^2 in case $\beta_z=0$

$$\text{plim } R^2 = \frac{\beta_w' G_{ww} \beta_w}{\beta_w' G_{ww} \beta_w + g}$$

and again this is not equal to zero in general.
Furthermore, R_z^2 is invariant with respect to changes of location or scale in any $\underline{y}_i$, whereas R^2 is not.
We can conclude that it is preferable to use R_z^2 as goodness-of-fit measure in the SURE-model if all equations contain a constant term. However, if there is at least one equation without a constant term McElroy's definition is of no use and one may consider as a goodness-of-fit measure r_z^2 as given in (9) or R^2 as given in (10).

APPENDIX: BASIC ALGEBRAIC PROPERTIES OF THE MODEL

It follows from (3) and (4) that

$$X'(\Omega^{-1}\otimes I_n)\underline{e} = 0. \tag{A.1}$$

Using the partition of X we find

$$(I_g\otimes s_n')(\Omega^{-1}\otimes I_n)\underline{e} = 0, \tag{A.2}$$

which yields

$$(I_g\otimes s_n')\underline{e} = 0, \tag{A.3}$$

and consequently

$$(I_g\otimes N_n)\underline{e} = \underline{e}. \tag{A.4}$$

REFERENCES

BUSE, A. (1973), Goodness of Fit in Generalized Least Squares Estimation, *The American Statistician 27*, 106-108.

-(1979), Goodness-of-Fit for the Seemingly Unrelated Regressions Model: A Generalization, *Journal of Econometrics 10*, 109-113.

HEIJMANS, R.D.H. and H. NEUDECKER (1987), The Coefficient of Determination revisited, in: R.D.H. Heijmans and H. Neudecker (eds.)(1987), *The Practice of Econometrics*, Kluwer Academic Publishers, Dordrecht, 191-204.

LUKACS, E. (1975), *Stochastic Convergence*,(2nd ed.), Academic Press, New York.

MCELROY, M.B. (1977), Goodness of Fit for Seemingly Unrelated Regressions, *Journal of Econometrics 6*, 381-387.