Effects of non-orthogonality on the efficiency of seemingly unrelated regression (SUR) models

W. B. Yahya*[1], S. B. Adebayo[2], E. T. Jolayemi[2], B. A. Oyejola[2] and O. O. M. Sanni[2]

[1]Department of Medical Statistics and Epidemiology, Technical University of Munich
Klinikunm rechts der Isar, Ismaningerstraße 22, 81675 Munich, Germany
*E-mail: wb_yahya@daad-alumni.de

[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstracts

This work examines the relative gain/loss in efficiency of Seemingly Unrelated Regression (SUR) estimators when one or more pair of the predictors in the system of equations are correlated (non-orthogonal). Literature has revealed that multicollinearity often affects the efficiency of SUR estimators. This paper, however, addresses such challenges by determining the 'Tolerable Non-orthogonal Correlation Points' (TNCP) among the predictors at which the efficiency of SUR estimators will still be preserved. Results from our simulation studies showed that SUR estimators are still efficient up to the range of the TNCP values, especially when the predictors have Gaussian distribution. By this result, the strict condition of purely uncorrelated covariates often advanced in the literature as a criteria for achieving efficient SUR estimators are relaxed within the range of TNCP values proposed in this work. Comparisons of the SUR estimators with that of the equation-by-equation from the ordinary least squares (OLS) are also taken up to assess their respective performances in the presence of non-orthogonal explanatory variables. In all cases considered, SUR estimators are consistently more efficient than the OLS.

Key words: Average Mean Square Error (AMSE), Contemporaneous correlation, Feasible Generalized Least Squares, Multicollinearity, Tolerable Non-orthogonal Correlation Point (TNCP), SUR.

2000 Mathematics Subject Classification: 62J05, 11D04

*[1] *Corresponding Author. E-mail:* wb_yahya@daad-alumni.de *Tel: +49 1520 6011 729 Fax: +49 89 4140 4850*

1. Introduction

A Seemingly Unrelated Regression (SUR) is a system of regression equations which consists of a set of m regression equations, each of which contains different explanatory variables and satisfies the classical assumptions of the standard regression (CASR) model. The SUR estimation procedures which enable an efficient joint estimation of all the regression parameters was first reported by Zellner (1962) which involves the application of Aitken's generalised least-squares (AGLS), (Aitken 1934, Powell 1965) to the whole system of equations. Zellner (1962 & 1963), Zellner & Theil (1962) submitted that the joint estimation procedure of SUR is more efficient than the equation-by-equation estimation procedure of the ordinary least square (OLS) and the gain in efficiency would be magnified if the contemporaneous correlation between each pair of the disturbances in the SUR system of equations is very high and explanatory variables (covariates) in different equations are uncorrelated. In other words, the efficiency in the SUR formulation increases the more the correlation between error vector differs from zero and the closer the explanatory variables for each response are to being uncorrelated.

After the much celebrated Zellner's joint generalized least squares estimator, several other estimators for different SUR systems were developed by many scholars to address different situations being investigated. For instance, Jackson (2002) developed an estimator for SUR system that could be used to model election returns in a multiparty election. Sparks (2004) developed a SUR procedure that is applicable to environmental situations especially when missing and censored data are inevitable. In share equation systems with random coefficients, Mandy & Martins-Filho (1993) proposed a consistent and asymptotically efficient estimators for SUR systems that have additive heteroscedastic contemporaneous correlation. They followed Amemiya (1977) by using Generalized Least Squares (GLS) to estimate the parameters of the covariance matrix. Furthermore, Lang, Adebayo & Fahrmeir (2002), Adebayo (2003), and Lang *et al* (2003) in their works also extended the usual parametric SUR model to Semiparametric SUR (SSUR) and Geoadditive SUR models within a Bayesian context. Also O'Donnell *et al* (1999) and Wilde *et al* (1999) developed

SUR estimators that are applicable in agricultural economics. More recently, Foschi (2004), Foschi *et al* (2003) and Foschi & Kontoghiorghes (2002, 2003) provided some new numerical procedures that could successively and efficiently solve a large scale of SUR model. Several other estimators proposed within the SUR frame- work could be found in Telser (1964), Parks (1967), Kakwani (1967), Kmenta & Gilbert (1968), Revankar (1974 & 1976), Mehta & Swamy (1976), Dwivedi & Srivastava (1978), Maeshiro (1980), Blattberg & George (1991), Kontoghiorghes & Clerke (1995), Kontoghiorghes & Dinenis (1996, 1997), Smith & Kohn (2000) and Kontoghiorghes (2003).

In all the estimation procedures developed for different SUR situations as reported above, Zellner's basic recommendation for high contemporaneous correlation between the error vectors with uncorrelated explanatory variables within each response equations was also maintained. However, in spite of the danger of multicollinearity reported in all these works, no remedy or how best to select covariate in the presence of multicollinearity within the context of SUR was provided.

When two or more exogenous (explanatory) variables are highly correlated or have the same predictive power with respect to the endogenous (response) variable, they are said to be *multicollinear*. In other words, whenever the independent variables are correlated among themselves, inter-correlation or multicollinearity among them is said to exist (Neter & Wasserman, 1974 and Neter *et al,* 1996). Whenever there is a complete absence of linear relationship (correlation) among these explanatory variables then, they are said to be orthogonal,(Silvey ,1969). However, in most practical situations, the explanatory variables across the different equations in SUR systems are often correlated. For instance, suppose we consider a study similar to that of Taylor *et al* (1986 & 1990) where it may be required to estimate, within the SUR framework, the number or proportion of HIV seropositive and seroconverter subjects who eventually developed AIDS from a particular cohort. The seropositive cohort consists of those subjects who already have HIV infections (HIV antibody positive) at enrolment in the study and the seroconverter cohort are those subjects who became infected during the follow-up period (those who changed from antibody negative to antibody positive during the follow-up period). Some of the covariates usually considered for such estimations among others, include the HIV antibody serology tests (both ELISA and Western blot), T-helper cell percentage, Platelet count, haemoglobin, age and weight of the subjects. However, it would constitute a statistical fallacy to ignore the possibility of some level of correlation between the age and weight of the subjects when performing such an estimation. Therefore, the recognition and management of this type of association requires adequate attention in order to ensure efficient estimates.

In another situation, the number or proportion of patients cured through two or more medical interventions with equal potency may be related to their dose levels. Determination and management of the degree of association among the available interventions may assist to control their application on the patients. For instance, if the association between a pair of intervention is very strong, it may be necessary to reduce if not discontinue the application of one of them to prevent overdose or effects of multiple reactions. This kind of dose-response study can be accommodated within the SUR framework whereby the tolerable level of association among the available medical interventions may be desirable.

Also, it may be necessary to jointly regress the demand for two or more complementary products like automobiles and gasoline on peoples' income and expenditures on other products within the SUR framework. While the two demands (responses) would obviously correlate through their error, satisfying the first basic requirement of SUR estimation, people's income and their expenditure on other products should not be expected to be uncorrelated. Thereby, violating the second important condition. Therefore, the existence of this kind of relationship needed to be recognized and accorded proper management within the SUR context such that the efficiency of SUR estimator would not be compromised.

It is now obvious, due to several instances of SUR highlighted above, that the independent variables are often correlated (collinear). Since the severity of these correlations could compromise the efficiency of SUR estimator as submitted in many earlier works, (Zellner, 1962, Liu & Wang,1999, Kurata, 1999, Sparks, 2004 etc.), our main task in this work, is to determine the admissible level of correlation between any pair of covariates within the SUR set-up at which the efficiency of SUR estimator would still be preserved. We shall call such correlation levels the *Tolerable Non-orthogonal Correlation Points* (TNCP). Therefore, the TNCP would be a range of correlation values that can exist between any pair of covariates in each of the separate equations for which the efficiency of SUR estimator would not be lost. We shall equally attempt to examine the performances of both SUR and OLS estimators at varying degree of collinearity and sample sizes.

In Section 2, the structural parametric framework of SUR system is presented while the simulation studies carried out in the work is discussed in Section 3. Detail discussion of our results is presented in Section 4 while Section 5 provides some concluding remarks and further suggestions to the study.

2. The Parametric SUR Framework

Consider a complete system of regression equations with m response variables each containing n observations denoted by the vector $\mathbf{Y}' = (y_1, y_2, \ldots, y_m)$ with associated distinct vector of explanatory variables $X_1, X_2, \ldots, X_m$ respectively. Each of the equations in this system of regression equations is assumed to satisfy the Gauss-Markov properties of homoscedasticity and no serial correlations of the error terms. That is, for each of the response equations the popular distributional assumptions on the error term of

$$\varepsilon_i \sim N\left(0, \sigma_i^2\right) \tag{2.10}$$

for $i = 1, 2, \ldots, m$ and

$$Cov\left(\varepsilon_{n_i}, \varepsilon_{n_i'}\right) = 0 \tag{2.11}$$

are maintained for $n_i, n_i' = 1, 2, \ldots, n$.

Therefore, the system can be represented by

$$\left.\begin{array}{c} y_1 = X_1\beta_1 + \varepsilon_1 \\ \vdots \\ y_m = X_m\beta_m + \varepsilon_m \end{array}\right\} \tag{2.12}$$

where; $i = 1, 2, \ldots, m$, y_i is an $n \times 1$ vector of observations on the i^{th} response variable, X_i is an $n \times p_i$ matrix of explanatory variables, β_i is a $p_i \times 1$ vector of regression parameters and ε_i is the corresponding $n \times 1$ vector of disturbances. Thus, each set of the y_i regression equations has p_i parameters. This system of equations in (2.12) can further be presented in a more compact form as

$$\overbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}}^{Y}_{mn \times 1} = \overbrace{\begin{pmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_m \end{pmatrix}}^{X}_{mn \times \sum_i p_i} \overbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}}^{\beta}_{\sum_i p_i \times 1} + \overbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix}}^{\varepsilon}_{mn \times 1} \tag{2.13}$$

and when stacked together the whole system becomes

$$Y = X\beta + \varepsilon \tag{2.14}$$

All the regression equations in (2.12) or (2.13) appear independent (seemingly unrelated) with one another because they do not have common variables or parameters. However, as explained in Section 1, Zellner (1962) was of the opinion that each pair of the system of regression equations above are actually (contemporaneously) correlated through their error terms ε_i, i = 1, 2, …, m, as may occur when regressing the demands or supplies for two related products on some covariates. Hence, the name *Seemingly Unrelated Regression* (SUR) given to such a system of regression equations as represented by (2.12) or (2.13). Estimating each of the equation separately by OLS may still yield consistent but inefficient estimates of the regression parameters. Therefore, in SUR estimation techniques the correlations among the errors in different equations are used to improve the regression estimates.

Looking at the estimation of the stacked equation of (2.14) from OLS point of view, one could have simply provided the estimator of the parameter β as

$$\hat{\beta} = \left(X^{/} X\right)^{-1} X^{/} Y \tag{2.15}$$

The estimator given by (2.15) can only provide a set of consistent but less efficient estimates of the regression in (2.14).

However, the assumption placed on the variance-covariance matrix of the disturbance in equation (2.14) is that

$$E\left(\varepsilon\varepsilon^{/}\right) = \Sigma \otimes I_n \tag{2.16}$$

where Σ is an $m \times m$ matrix of the form

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1m}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^2 & \sigma_{m2}^2 & \cdots & \sigma_{mm}^2 \end{pmatrix} \qquad (2.17)$$

I_n is an $n \times n$ identity matrix and $\otimes$ is the Kronecker product which multiplies each element in Σ by I_n. In addition, the standard assumption that

$$E(\varepsilon) = 0 \qquad (2.18)$$

is also maintained. The representations in (2.16) and (2.17) are obvious from

$$E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1\varepsilon_1') & E(\varepsilon_1\varepsilon_2') & \cdots & E(\varepsilon_1\varepsilon_m') \\ E(\varepsilon_2\varepsilon_1') & E(\varepsilon_2\varepsilon_2') & \cdots & E(\varepsilon_2\varepsilon_m') \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_m\varepsilon_1') & E(\varepsilon_m\varepsilon_2') & \cdots & E(\varepsilon_m\varepsilon_m') \end{pmatrix} \qquad (2.19)$$

$$= \begin{pmatrix} \sigma_{11}^2 I_n & \sigma_{12}^2 I_n & \cdots & \sigma_{1m}^2 I_n \\ \sigma_{21}^2 I_n & \sigma_{22}^2 I_n & \cdots & \sigma_{2m}^2 I_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^2 I_n & \sigma_{m2}^2 I_n & \cdots & \sigma_{mm}^2 I_n \end{pmatrix} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1m}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^2 & \sigma_{m2}^2 & \cdots & \sigma_{mm}^2 \end{pmatrix} \otimes I_n \qquad (2.20)$$

If the $m \times m$ positive definite variance-covariance matrix $\Sigma \otimes I_n$ of (2.16) is denoted by Ω and all values of elements of Σ are known, then the SUR formulation of the regression models in (2.12) through (2.14) produces more efficient regression parameter estimates using generalised least squares (GLS) estimator given by

$$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \qquad (2.21)$$

and the matrix of the standard error of the parameter estimates is given by

$$Var(\hat{\beta}_{GLS}) = \left(X'\Omega^{-1}X\right)^{-1} \qquad (2.22)$$

where

$$\Omega^{-1} = \Sigma^{-1} \otimes I_n \qquad (2.23)$$

The idea behind the development of (2.21) is that since the variance-covariance Ω is a positive definite square matrix of order m, then there exit a matrix P, such that

$$P\Omega P' = I_m \Leftrightarrow \Omega = P^{-1}P^{-1'} \qquad (2.24)$$

and if we pre-multiply the stacked SUR model in (2.14) by matrix P we obtain $PY = PX\beta + P\varepsilon$, which now gives a new transformed model

$$Y^* = X^* \beta + \varepsilon^* \tag{2.25}$$

Equation (2.25) now becomes a standard classical regression model that satisfies the Gauss- Markov properties. Thus, this transformed model now has a diagonal variance-covariance matrix with all its off-diagonal elements being zero. This is evident from the following reasoning;

$$E\left(\varepsilon^* \varepsilon^{*/}\right) = E\left(P\varepsilon\varepsilon^/ P^/\right)$$

$$= PE\left(\varepsilon\varepsilon^/\right)P^/$$

$$= P\Omega P^/ = I_m$$

Thus,

$$E\left(\varepsilon^* \varepsilon^{*/}\right) = I_m \tag{2.26}$$

This is a typical form of the Weighted Least Squares (WLS) methodology.

Applying OLS to estimate (2.25) would, however provide the Best Linear Unbiased Estimate (BLUE) of all the regression parameters of the system since it would now use the contemporaneous variance-covariance of all the disturbances in the whole system of equations as postulated by Zellner (1962). This is obvious since the OLS estimator of (2.25) now becomes

$$\hat{\beta}_{GLS} = \left(X^{*/} X^*\right)^{-1} X^{*/} Y^*$$

$$\rightarrow \qquad \hat{\beta}_{GLS} = \left(X^/ P^/ P X\right)^{-1} X^/ P^/ P Y$$

and from the relation in (2.24) we obtained

$$\Omega = P^{-1} P^{-1/} \Leftrightarrow \Omega^{-1} = P^/ P \tag{2.27}$$

which can then be applied on the above estimator to have its final form of

$$\hat{\beta}_{GLS} = \left(X^/ \Omega^{-1} X\right)^{-1} X^/ \Omega^{-1} Y$$

as earlier provided in (2.21).

In practice, the value of the variance-covariance matrix Ω is not always known. The popular procedure which is generally referred to as three stage least squares (3SLQ) method is as follows;

i) Estimate separately each of the equations in the system given in (2.12) and from each estimate determine the residual vectors $\hat{\varepsilon}_{i(n^/)} = y_i - X_i \hat{\beta}_i$ for $i = 1, 2, \ldots, m$ and $n^/ = 1, 2, \ldots, n$.

ii) From the estimated residuals $\hat{\varepsilon}_{i(n')}$ from (i) compute the initial estimates of the elements of contemporaneous variance-covariance Ω in (2.20), denoted by $\hat{\Omega}$. By this procedure, the estimate of the diagonal elements of Ω (the variance of the error term for each equation, σ_{ii}^2) can be computed by

$$\sigma_{ii}^2 = \frac{1}{n - p_i - 1}\sum_{n'}^{n} \varepsilon_{i(n')}^2 \tag{2.28}$$

and for $i, j = 1,2, \ldots, m,\; i \neq j$, the estimate of the off-diagonal elements of Ω (the covariances between i^{th} and j^{th} equations, σ_{ij}^2) can equally be computed by

$$\sigma_{ij}^2 = \frac{1}{n - \max(p_i, p_j) - 1}\sum_{n'}^{n} \varepsilon_{i(n')}\varepsilon_{j(n')} \tag{2.29}$$

iii) Replace Ω by the estimated variance-covariance $\hat{\Omega}$ in the GLS estimator (2.21) of the stacked equation (2.14) and carry out the SUR estimation of the regression equations now using estimator

$$\hat{\beta}_{GLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}Y \tag{2.30}$$

This procedure is sometimes called the Feasible Generalised Least Squares (FGLS) method.

iv) The variance-covariance matrix $\hat{\Omega}$ in (2.30) can then be re-estimated using the SUR residuals in step (iii) and continue iterating the procedure until convergence is achieved. This is the Iterated SUR (ITSUR) estimation procedure.

However, the inverse of the positive definite variance-covariance matrix $\hat{\Omega}$, $\hat{\Omega}^{-1}$ is simply achievable through the use of Cholesky decomposition (Horn & Johnson,1985, Julier & Uhlmann, 1997) on the matrix.

Both the FGLS and ITSUR are maximum likelihood estimators that are capable of providing consistent, more efficient and Best Linear Unbiased Estimation of a SUR system provided such system of equations is adequately identified. Omission of important variable(s) in one of the equations may introduce bias into the estimates of all the other regression coefficients. While it is possible to obtain unique structural coefficients from the estimates of the stacked (or reduced) form of the system of regression when the system is exactly identified, more than one unique numerical values are provided for the structural coefficients when the system is over-identified while it is impossible to get the structural coefficient values from the stack equation if the system is under-identified. Therefore, care must be taken to ensure that the system of equations is properly identified when using the FGLS or ITSUR procedure.

However, both the OLS and the GLS (FGLS or ITSUR) methods would provide the same efficient and unbiased estimate of the whole system of regression equations in (2.12) or (2.13) if;

i) $E(\varepsilon\varepsilon') = \{\sigma_{ij}^2\} I_n$ where, for $i \neq j =$ 1,2, …, m, $\sigma_{ij}^2 = 0$. That is, the contemporaneous covariances (the off-diagonal elements of Ω) are all zero. In other words, the error terms across all the equations are contemporaneously independent (uncorrelated).

ii) all the equations in the system in (2.12) have the same values for the independent variables. That is, $X_1 = X_2 = \cdots = X_m$. In this case, the regression equations in (2.12) are not just 'seemingly' but are actually related.

3. Simulation Studies

Our simulation work considers a system of SUR equations containing three distinct linear regression equations with one of them having a pair of correlated covariates. Let the three equations be distributed as follows; $(Y_1 \mid X_1) \sim N_n(X_1\beta_1, \sigma_{11}^2)$, $(Y_2 \mid X_2) \sim N_n(X_2\beta_2, \sigma_{22}^2)$ and $(Y_3 \mid X_3) \sim N_n(X_3\beta_3, \sigma_{33}^2)$. Thus, with m = 3 in (2.12) we have

$$\left.\begin{aligned} y_1 &= X_1'\beta_1 + \varepsilon_1 \\ y_2 &= X_2'\beta_2 + \varepsilon_2 \\ y_3 &= X_3'\beta_3 + \varepsilon_3 \end{aligned}\right\} \tag{3.10}$$

and the whole system in (3.10) is assumed to be distributed, according to (2.20), as

$$(Y \mid X) \sim N_{3n}\left(X\beta,\ \Sigma \otimes I_n \right) \tag{3.11}$$

where,

$$X = \begin{bmatrix} X_1^{/} & 0 & 0 \\ 0 & X_2^{/} & 0 \\ 0 & 0 & X_3^{/} \end{bmatrix} \in \Re^{3\times 3n} \;,\quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \in \Re^{3n} \qquad (3.12)$$

The covariates and parameters in (3.12) are therefore, defined as

$$\left.\begin{array}{ll} X_1^{/} = \left(X_{10} \quad X_{11} \quad X_{12}\right) & \beta_1^{/} = \left(\beta_{10} \quad \beta_{11} \quad \beta_{12}\right) \\ X_2^{/} = \left(X_{20} \quad X_{21}\right) & \beta_2^{/} = \left(\beta_{20} \quad \beta_{21}\right) \\ X_3^{/} = \left(X_{30} \quad X_{31}\right) & \beta_3^{/} = \left(\beta_{30} \quad \beta_{31}\right) \end{array}\right\} \qquad (3.13)$$

with $X_{10} = X_{20} = X_{30} = 1$ in (3.13) for the intercepts of the three equations.

When $m = 3$ in (2.17), Σ is then a positive definite 3×3 variance-covariance matrix of the errors ε_i, $i = 1, 2, 3,$ defined by

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 \end{pmatrix} \qquad (3.14)$$

In the Monte Carlo simulation work, we considered the two covariates X_{11} and X_{12} in model Y_1 as two correlated bivariate standard normal random variables with density function given by

$$P(X_{11}, X_{12}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left(X_{11}^2 + X_{12}^2 - 2\rho X_{11}X_{12}\right)\right] \quad (3.15)$$

where;

$$\rho = \rho_{X_{11},X_{12}} = \frac{Cov(X_{11}, X_{12})}{\sqrt{Var(X_{11}) \times Var(X_{12})}} \qquad (3.16)$$

Therefore, the structural form of the SUR equations for our simulation is

$$\left.\begin{array}{l} y_1 = 0.8 + 0.5X_{11} + 0.4X_{12} + \varepsilon_1 \\ y_2 = -0.6 + 0.6X_{21} + \varepsilon_2 \\ y_3 = 2.6 - 0.5X_{31} + \varepsilon_3 \end{array}\right\} \qquad (3.17)$$

as used in Adebayo(2003) and Lang *et al*(2003) with little modifications where $X_{21} \sim U(-3,3)$, $X_{31} \sim U(-3,3)$ and the distribution of X_{11} and X_{12} is as given in (3.15). The error terms, $\varepsilon = (\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3)^{/} \overset{iid}{\sim} N_{3n}(0, \quad \Sigma \otimes I_n)$ is independently and identically distributed as defined in (3.14). The true value of the variance-covariance Σ used in ε for this simulation is

$$\hat{\Sigma} = \begin{pmatrix} 1 & 0.6 & 0.8 \\ & 1 & 0.7 \\ & & 1 \end{pmatrix} \qquad (3.18)$$

The variance-covariance matrix (3.18) is a Hermittan positive-definite non-singular symmetric matrix whose Cholesky decomposition is computed as

$$K = \begin{pmatrix} 1 & 0 & 0 \\ 0.6 & 0.8 & 0 \\ 0.8 & 0.275 & 0.5333 \end{pmatrix} \qquad (3.19)$$

Therefore, in order to establish a strong contemporaneous relationship between the three sets of equations in (3.17) through their error terms, we have used in our simulation work, a new 3×1 error vector $\varepsilon^* = \left(\varepsilon_1^* \quad \varepsilon_2^* \quad \varepsilon_3^*\right)'$ in place of $\varepsilon = \left(\varepsilon_1 \quad \varepsilon_2 \quad \varepsilon_3\right)'$ whose elements are determined by the product

$$\varepsilon^* = K \times \varepsilon = \begin{pmatrix} \varepsilon_1 \\ 0.6\varepsilon_1 + 0.8\varepsilon_2 \\ 0.8\varepsilon_1 + 0.275\varepsilon_2 + 0.5333\varepsilon_3 \end{pmatrix} \qquad (3.20)$$

Also, non-orthogonality (collinearity) between the two covariates X_{11} and X_{12} in model Y_1 were established and incorporated in all our simulation studies by generating both covariates with correlation values: ±0.9, ±0.8, ±0.7, ±0.6, ±0.5, ±0.4, ±0.3, ±0.2, ±0.1, and at a correlation level of zero when the two of them are assumed to be purely uncorrelated. The values of the other covariates, X_{21} and X_{31} were equally generated based on their indicated distributions. The entire model was, however, simulated for various sample sizes (n); n =1000, 500, 200, 100 and 50 with 100 replicates in each case. This is to enable us assess the performances of both SUR and OLS estimators at various sample sizes in the presence or absence of multicolinearity. Various programs were written and exported into STATA do-file editor and were all executed using STATA/SE software package, 8.0 version (Stata Corporation, Texas, USA).

The quality of performance of each of these estimators (SUR and OLS) was assessed by their average mean square errors (AMSE) over 100 replications. This permits their assessments in terms of variations and unbiasedness. The MSE of an estimator is the expected value of the square of the error while the error is the amount by which estimator differs from the true value to be estimated. However, these differences may occur because of randomness (*which is applicable to both estimators*) and/or because the estimator does not account for the information that could produce a more accurate estimate.

Although, simulations were performed with n =1000, 500, 200, 100 and 50 using 100 replicates in each case, we chose to present findings for n = 500 and 50. *Tables 1a & b* to *Tables 5a & b* in *Appendices* show the results of our simulations. *Tables 1a & b* and *Tables 2a & b* present the output of simulation results for SUR and OLS estimators and their average mean square errors (AMSE) for model Y_1 at various correlation levels between the regressors X_{11} and X_{12} for n = 500 and 50. The simulation results for SUR and OLS estimators for other two models Y_2 and Y_3 are presented in *Tables 3a & b* and *Tables 4a & b* with the corresponding AMSE values. As discuss in the next Section, the correlation between X_{11} and X_{12} in model Y_1 did not affect the efficiency of both SUR and OLS estimators for models Y_2 and Y_3.

4. Discussion of Results

Relevant discussions of the simulation results obtained in this work are presented in this section. The results of the simulation studies (some of which are presented in *Appendix A & B*) generally show that both SUR and OLS estimators are increasingly more efficient as the sample size increases irrespective of the existence of multicollinearity among the pair of covariates in the models. This is evident from the decreasing trend in the average mean square errors (AMSE) in model Y_1 for both estimators as the sample size increases (see *Tables 1a & b* and *Tables 2a & b*). These appreciating performance by both estimators due to increase in sample size can be found in *Tables 3a & b* and *4a & b*, the respective table of results for models Y_2 and Y_3. These results are again presented by the plot of AMSE values against the various sample sizes for both estimators as shown in *Figs 1a & b* (see *Appendix C*) for models Y_2 and Y_3 in which it is observed that all the AMSE values for both estimators decrease as sample size increases with non-orthogonal covariates X_{11} and X_{12} in model Y_1.

However, despite the gain in efficiency by OLS estimators due to increased sample size, SUR estimators are still consistently more efficient than OLS in all cases as could be seen from their respective AMSE plots in *Figs 1a & b*. The AMSE values for SUR estimators are considerably lower than that of OLS in all cases (see *Tables1a & b* to *2a & b*). These are clearly presented in the Box-and-Whisker plots in *Figs 2a & b* for models Y_2 and Y_3. On the two plots, the median AMSE values for SUR estimators are considerably smaller than that of the OLS.

Our results here further reveal that the efficiency of SUR estimator continues to increase in the presence of highly correlated covariates (multicollinearity). In model Y_1 for example, incredible gains in efficiency of SUR model were recorded when the sample size $n \geq 500$ even at a higher correlation level between X_{11} and X_{12}. It is therefore found that, at any sample size $n \geq 500$ with correlation level between a pair of predictors as high as ±0.9, SUR estimators is still more efficient than its (SUR) estimators at smaller sample sizes even if the predictors in the model are not correlated. This can be observed on *Tables 1a & 1b* where the AMSE values for SUR estimator $\hat{\beta}_{11}$ are given as 0.0039 (at $\rho_{X_{11},X_{12}} = 0.9$, *Table 1a*) and 0.0038 (at $\rho_{X_{11},X_{12}} = -0.9$, *Table 1b*) at $n = 500$ in each case. These AMSE values are relatively smaller than their respective values computed for the same SUR estimator $\hat{\beta}_{11}$ at smaller sample size $n = 50$ as shown in *Table 2a* ($AMSE_{SUR}$ = 0.0083) and *Table 2b* ($AMSE_{SUR}$ = 0.0068) for $\hat{\beta}_{11}$ at $\rho_{X_{11},X_{12}} = 0.0$. This better performance of SUR estimators at higher sample sizes and collinearity is also observed at $n \geq$ 500 relative to other lower samples n = 100 and 200. All these results are shown in *Figs 3a & b*. The Box-and-Whisker plot of the AMSEs of SUR estimator at various sample sizes as shown in *Figs 4a & b* equally confirmed that SUR estimator increases in efficiency as the sample size increases irrespective the level of correlation between the covariates.

As stated in Section 1, the admissible correlation level (TNCP) among the predictors in separate equation of SUR system is to be determined in this study. Based on our AMSE criteria, we consider the TNCP point as the correlation point (or a range of correlation values) at which AMSE of SUR estimator with non-orthogonal (correlated) covariates (AMSE$_{\text{Nonothogonal}}$) do not significantly differ from its AMSE values (AMSE$_{\text{Othogonal}}$) at which the respective covariates are purely orthogonal (uncorrelated). This study found the TNCP to fall within the two end points -0.2 and +0.2. That is, $|TNCP| \leq 0.2$. This is the admissible correlation range of values that could exist between any pair of predictors in SUR system of equations at which the efficiency of SUR estimators would still be preserved at all sample sizes. This is particularly true for SUR models that satisfied the conditions established for the simulation studies in this work. The details of how we found this result are given below.

From *Tables 2a & b*, the AMSE$_{\text{Othogonal}}$ for β_{11} in model Y_1 from SUR estimators at n = 50 are 0.0083 and 0.0068 respectively. Also, the respective AMSE$_{\text{Orthogonal}}$ values for β_{12} from SUR estimators are 0.0060 and 0.0109. These are their AMSE values at $\rho_{X_{11},X_{12}} = 0.0$. Similarly, from *Tables 2a & b,* the AMSE$_{\text{Nonorthogonal}}$ values for β_{11} from SUR estimators at $\rho_{X_{11},X_{12}} = \pm 0.2$ are 0.0085 and 0.0071 respectively. The corresponding AMSE$_{\text{Nonorthogonal}}$ values of SUR estimators for β_{12} are 0.0063 and 0.0114. Therefore, the simple differences (AMSE$_{\text{Nonorthogonal}}$ - AMSE$_{\text{Orthogonal}}$) between these two sets of AMSE for each of the β_{11} and β_{12} yield 0.0002 (for β_{11}, *Table 2a*), 0.0003 (for β_{11}, *Table 2b*), 0.0003 (for β_{12}, *Table 2a*) and 0.0005 (for β_{12} *Table 2b*). In all the cases, it is clear that AMSE$_{\text{Nonorthogonal}}$ at $\rho_{X_{11},X_{12}} = \pm 0.2$ agrees with the AMSE$_{\text{Orthogonal}}$ up to 3 decimal places for SUR estimators. There seems a perfect agreement at this point. However, apparent loss in the efficiency of SUR estimator is observed when we take the differences between the AMSE$_{\text{Nonorthogonal}}$ and AMSE$_{\text{Orthogonal}}$ values at other higher correlation levels beyond the range $|TNCP| \leq 0.2$. For example, at $\rho_{X_{11},X_{12}} = 0.3$, the difference (AMSE$_{\text{Nonorthogonal}}$ - AMSE$_{\text{Orthogonal}}$) for SUR estimator $\hat{\beta}_{11}$ from *Table 2a* yields $0.0006 \approx 0.001$ when rounded up to 3 decimal places. This difference is ≈ 0.002 at $\rho_{X_{11},X_{12}} = 0.5$ while it is ≈ 0.011 at $\rho_{X_{11},X_{12}} = 0.8$. Thus, there is an increasing sequence of loss in the efficiency of SUR estimator as the collinearity levels move away from the TNCP range at n = 50. These results are better explained by the plot of the AMSE values of SUR estimators for model Y_1 against the various correlation points at n = 50 (*Figs 5a & b*). It can be observed from the graphs that the AMSE of SUR estimators stay the same at all the correlations between X_{11} and X_{12} starting from 0.0, ±0.1 and ±0.2 after which an up-ward rise in these AMSE values started to manifest at other higher correlation values.

In order to give credence to the choice of our TNCP range of values, the differences between the $AMSE_{Nonorthogonal}$ and $AMSE_{Orthogonal}$ values for SUR estimators $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ in model Y_1 were similarly examined at other samples sizes (n = 100, 200, 500 and 1000) and correlation. All the results obtained supported the TNCP threshold levels of ±0.2. These results are presented by the plot of AMSE values for SUR estimators at various sample sizes and correlations as shown in *Figs 6a & b*.

However, it can be observed from *Figs 6a & b* that there are perfect agreements between the $AMSE_{Nonorthogonal}$ and $AMSE_{Orthogonal}$ even beyond the TNCP threshold levels of ±0.2, but at higher sample sizes n = 500 and 1000. This result pointed at the remedy of multicollinearity provided by working with large sample size. However, the higher correlation levels beyond ±0.2 at which such perfect agreement was achieved are ignored. This is to avoid running into small sample problems. For instance, the SUR estimators remain efficient (both $AMSE_{Nonorthogonal}$ and $AMSE_{Orthogonal}$ remain the same) up to the collinearity level of ±0.4 at n = 500 and 1000. This apparent efficiency is not, however, sustained at smaller sample sizes n = 50 and 100 as an up-ward movements of the curves are noticed immediately after the collinearity level of ±0.2 at each of these sample sizes.

Since the SUR estimator is found to be efficient in the presence of collinearity levels up to ±0.2 and at a small sample size as low as 50, this shows that its efficiency will obviously improve at higher sample sizes even at collinearity levels beyond the TNCP threshold values. This is clearly demonstrated by the various graphs of AMSE in *Figs 6a & b*. *Tables 5a & b* in *Appendix A & B* however, showed all the $AMSE_{Nonorthogonal}$ and $AMSE_{Orthogonal}$ values for SUR estimators as well as their respective differences at TNCP = ±0.2 at all the chosen sample sizes. It is obvious from these tables that the difference ($AMSE_{Nonorthogonal}$ -

$AMSE_{Orthogonal}) \rightarrow 0$ as $n \rightarrow \infty$. If these differences are rounded up to 3 decimal place, both the $AMSE_{Nonorthogonal}$ and $AMSE_{Orthogonal}$ are exactly the same at all the sample sizes. However, the equality or otherwise of these two AMSE is further established at TNCP = ± 0.2 using the Bartlett's test for homogeneity of variances. The result are again provided in *Tables 5a & b*. At all the indicated sample sizes, the p-values > 0.05 an indication that the two AMSEs are the same at the chosen TNCP point.

In *Figs 5 & 6*, the $AMSE_{Orthogonal}$ which is read along the vertical axis, is the AMSE values of SUR estimator at zero collinearity (correlation) level between the covariates X_{11} and X_{12}. The $AMSE_{Nonorthogonal}$ on the other hand represented the AMSE for SUR estimators at other collinearity level different from zero which is equally read along the vertical axis.

This study reveals the efficiency of SUR estimator over OLS method at different sample sizes. Therefore, taking cognisance of the contemporaneous correlation among the response variables surely improves the efficiency of estimators from SUR models over OLS. This work further reveals that efficiency of SUR estimator over the OLS is more of a function of contemporaneous correlation and the sample size rather than the presence of multicollinearity alone which is the most commonly cited scenario.

It is important to remark that, the various collinearity levels considered for the two covariates X_{11} and X_{12} of model Y_1 only affected the regression estimates for model Y_1 while the estimates of models Y_2 and Y_3 remain the same under specific cases considered. This is because the two covariates that are collinear belong to model Y_1 only. Therefore, whatever influence brought about by the changes in the levels of their collinearity can only affect the regression estimates of model Y_1.

5. Concluding Remarks

In this work, we reaffirm that SUR estimator is consistently better than the equation-by-equation method of OLS in estimating a system of regression equations, which are related by their disturbance terms. Though, the two estimators increase in their efficiencies as the sample size increases, nonetheless SUR supremacy over the OLS is maintained whether or not some pairs of covariates in the system are correlated.

It was equally established that no matter the level of correlation (collinearity) among the pair of covariates in a given SUR system of equations, SUR estimator would still be efficient at a large simple size $n \geq 500$. In other words, multicollinearity effect that seemed to be a peculiar challenge with SUR can actually be overcome with larger sample size.

In addition, it is found that if the collinearity level (correlation) between some pair(s) of independent variables in any of the separate equations in a given SUR system fall within the end-points ±0.2, SUR estimators would still be efficient. This is particularly true if the conditions under which the simulation studies carried out in this work are present. We have called this correlation range the *Tolerable Non-orthogonal Correlation Points* (TNCP). A correlation level of ±0.2 between a pair of covariates has been found to be significant throughout our simulation studies. Therefore, discovering these as the end points of our TNCP range has further revealed the significance of this correlation values within the context treated here.

In our future study we shall try to examine the influence of collinear predictors on the performance of SUR within the context of non-linear models.

A flexible test procedure on the variance-covariance matrix Σ in (2.16) would also be developed to easily establish the existence or otherwise of SUR criteria in a given system of equation. This will enable us to carry out relevant directional or non-directional hypothesis testing of the form

$$H_0 : \Sigma = I_m \quad \text{vs} \quad H_a : \Sigma \neq I_m \tag{4.10}$$

on Σ for the off-diagonal elements where I_m is a $m \times m$ diagonal variance matrix of the error ε with all its off-diagonal elements being zero. This may be necessary, since if H_0 is not rejected by such test, then the use of SUR estimators would merely be a waste of time and efforts. The OLS method would be a better option under such a situation.

Finally, a possible extension of this study to Non-parametric, Semi-parametric and Geoadditive SUR model frameworks would also be considered, especially when the Gaussian distribution assumption on the predictors are violated.

Acknowledgements: The authors would like to thank Steven Lane for his comments on the draft of this work. These have helped in the production of this final version.

Appendix A

Table of Results for Positive Collinearity

Table 1a: *Simulation results of model Y_1 for SUR and OLS estimators with X_{11} and X_{12} positively correlated (n = 500)*

Sample size n = 500															
$\rho_{x_{11},x_{12}}$	Model Y_1 True coefficients			SUR Estimations			OLS Estimations			SUR Average MSE			OLS Average MSE		
	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}
0.9	0.8	0.5	0.4	0.8011	0.4900	0.4069	0.8011	0.4931	0.4112	0.0020	0.0039	0.0036	0.0020	0.0108	0.0101
0.8	0.8	0.5	0.4	0.8011	0.4922	0.4050	0.8011	0.4966	0.4081	0.0020	0.0021	0.0019	0.0020	0.0058	0.0053
0.7	0.8	0.5	0.4	0.8011	0.4932	0.4042	0.8011	0.4983	0.4068	0.0020	0.0015	0.0013	0.0020	0.0042	0.0038
0.6	0.8	0.5	0.4	0.8011	0.4939	0.4038	0.8011	0.4995	0.4061	0.0020	0.0012	0.0012	0.0020	0.0034	0.0030
0.5	0.8	0.5	0.4	0.8011	0.4944	0.4035	0.8011	0.5003	0.4056	0.0020	0.0011	0.0009	0.0020	0.0029	0.0026
0.4	0.8	0.5	0.4	0.8011	0.4948	0.4033	0.8011	0.5010	0.4053	0.0020	0.0009	0.0008	0.0020	0.0026	0.0023
0.3	0.8	0.5	0.4	0.8011	0.4952	0.4032	0.8011	0.5016	0.4051	0.0020	0.0009	0.0008	0.0020	0.0024	0.0021
0.2	0.8	0.5	0.4	0.8011	0.4955	0.4031	0.8011	0.5021	0.4050	0.0020	0.0008	0.0007	0.0020	0.0023	0.0020
0.1	0.8	0.5	0.4	0.8011	0.4959	0.4030	0.8011	0.5026	0.4049	0.0020	0.0008	0.0007	0.0020	0.0022	0.0019
0.0	0.8	0.5	0.4	0.8011	0.4962	0.4030	0.8011	0.5031	0.4049	0.0020	0.0008	0.0007	0.0020	0.0021	0.0019

Table 2a: *Simulation results of model Y_1 for SUR and OLS estimators with X_{11} and X_{12} positively correlated (n = 50)*

Sample size n = 50															
$\rho_{x_{11},x_{12}}$	Model Y_1 True coefficients			SUR Estimations			OLS Estimations			SUR Average MSE			OLS Average MSE		
	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}	β_{10}	β_{11}	β_{12}
0.9	0.8	0.5	0.4	0.7968	0.4805	0.4183	0.7945	0.4474	0.4455	0.0194	0.0343	0.0317	0.0216	0.0975	0.0896
0.8	0.8	0.5	0.4	0.7968	0.4864	0.4133	0.7945	0.4619	0.4331	0.0194	0.0192	0.0167	0.0216	0.0546	0.0473
0.7	0.8	0.5	0.4	0.7968	0.4892	0.4112	0.7945	0.4689	0.4278	0.0194	0.0142	0.0118	0.0216	0.0405	0.0334
0.6	0.8	0.5	0.4	0.7968	0.4910	0.4100	0.7945	0.4735	0.4248	0.0194	0.0118	0.0094	0.0216	0.0335	0.0334
0.5	0.8	0.5	0.4	0.7968	0.4924	0.4092	0.7945	0.4769	0.4229	0.0194	0.0104	0.0080	0.0216	0.0295	0.0227
0.4	0.8	0.5	0.4	0.7968	0.4935	0.4087	0.7945	0.4797	0.4217	0.0194	0.0095	0.0072	0.0216	0.0270	0.0203
0.3	0.8	0.5	0.4	0.7968	0.4945	0.4084	0.7945	0.4821	0.4208	0.0194	0.0089	0.0066	0.0216	0.0253	0.0183
0.2	0.8	0.5	0.4	0.7968	0.4954	0.4082	0.7945	0.4843	0.4203	0.0194	0.0085	0.0063	0.0216	0.0242	0.0177
0.1	0.8	0.5	0.4	0.7968	0.4962	0.4080	0.7945	0.4864	0.4200	0.0194	0.0083	0.0061	0.0216	0.0236	0.0172
0.0	0.8	0.5	0.4	0.7968	0.4970	0.4080	0.7945	0.4884	0.4199	0.0194	0.0083	0.0060	0.0216	0.0234	0.0170

Table 3a: *Simulation results for model Y_2 for SUR and OLS estimators with X_{11} and X_{12} positively correlated at various sample sizes*

$\rho_{x_{11},x_{12}}$ = 0.0, 0.1, 0.2, … , 0.9										
Sample sizes (n)	Model Y_2 True coefficients		SUR Estimations		OLS Estimations		SUR Average MSE		OLS Average MSE	
	β_{20}	β_{21}	β_{20}	β_{21}	β_{20}	β_{21}	β_{20}	β_{21}	β_{20}	β_{21}
1000	-0.6	0.6	-0.5939	0.5993	-0.5938	0.5975	0.0010	0.0002	0.0010	0.0003
500	-0.6	0.6	-0.5986	0.5963	-0.5986	0.5970	0.0020	0.0004	0.0020	0.0007
200	-0.6	0.6	-0.5962	0.6040	-0.5962	0.6040	0.0050	0.0008	0.0051	0.0016
100	-0.6	0.6	-0.6053	0.6039	-0.6054	0.6052	0.0101	0.0017	0.0103	0.0034
50	-0.6	0.6	-0.5870	0.6071	-0.5868	0.6063	0.0192	0.0034	0.0202	0.0068

Table 4a: *Simulation results for model Y_3 for SUR and OLS estimators with X_{11} and X_{12} positively correlated at various sample sizes*

$\rho_{x_{11},x_{12}}$ = 0.0, 0.1, 0.2, … , 0.9										
Sample sizes (n)	Model Y_3 True coefficients		SUR Estimations		OLS Estimations		SUR Average MSE		OLS Average MSE	
	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}
1000	2.6	-0.5	2.6050	-0.5012	2.6049	-0.5020	0.0010	0.0001	0.0010	0.0004
500	2.6	-0.5	2.6054	-0.5013	2.6054	-0.5009	0.0020	0.0002	0.0020	0.0007
200	2.6	-0.5	2.6023	-0.5015	2.6023	-0.5024	0.0050	0.0005	0.0051	0.0017
100	2.6	-0.5	2.5995	-0.4982	2.5990	-0.5013	0.0099	0.0011	0.0101	0.0038
50	2.6	-0.5	2.6103	-0.5028	2.6105	-0.5055	0.0188	0.0024	0.0196	0.0080

Table 5a: *Average Mean Square Error, $AMSE_{Orthogonal}$ and Average Mean Square Error, $AMSE_{Nonorthogonal}$ at TNCP = 0.2 for SUR estimators $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ of model Y_1 and their respective differences. The estimates of the chi-square statistics and p-values for the Bartlett Tests establishing the equality of the two AMSE are also reported.*

Sample size (n)	SUR $AMSE_{Orthogonal}$	SUR $AMSE_{Nonorthogonal}$	*Difference	Bartlett Test		SUR $AMSE_{Orthogonal}$	SUR $AMSE_{Nonorthogonal}$	*Difference	Bartlett Test	
	β_{11}	β_{11}		X^2-values	P-values	β_{12}	β_{12}		X^2-values	P-values
50	0.0083	0.0085	0.0002	0.014	0.905	0.0060	0.0063	0.0003	0.020	0.887
100	0.0043	0.0044	0.0001	0.038	0.846	0.0036	0.0038	0.0002	0.041	0.839
200	0.0020	0.0022	0.0002	0.295	0.587	0.0020	0.0021	0.0001	0.083	0.773
500	0.0008	0.0008	0.0000	0.520	0.471	0.0007	0.0007	0.0000	0.208	0.649
1000	0.0004	0.0004	0.0000	0.758	0.384	0.0004	0.0004	0.0000	0.416	0.519

**Note that when the differences is rounded up to 3 decimal places the equivalence of $AMSE_{Orthogonal}$ and $AMSE_{Nonorthogonal}$ at collinearity level of 0.2, which is our TNC point is established*

Appendix B

Table of Results for Negative Collinearity

Table 1b: *Simulation results of model Y_1 for SUR and OLS estimators with X_{11} and X_{12}*

Sample size n = 500															
$\rho_{x_{11},x_{12}}$	Model Y_1 True coefficients			SUR Estimations			OLS Estimations			SUR Average MSE			OLS Average MSE		
	β10	β11	β12	β10	β11	β12	β10	β11	β12	β10	β11	β12	β10	β11	β12
-0.9	0.8	0.5	0.4	0.8018	0.4951	0.3936	0.8018	0.5005	0.3935	0.0020	0.0038	0.0037	0.0020	0.0106	0.0104
-0.8	0.8	0.5	0.4	0.8018	0.4971	0.3954	0.8018	0.5026	0.3953	0.0020	0.0020	0.0020	0.0020	0.0057	0.0055
-0.7	0.8	0.5	0.4	0.8018	0.4981	0.3961	0.8018	0.5036	0.3961	0.0020	0.0014	0.0014	0.0020	0.0040	0.0039
-0.6	0.8	0.5	0.4	0.8018	0.4987	0.3965	0.8018	0.5042	0.3965	0.0020	0.0012	0.0011	0.0020	0.0032	0.0032
-0.5	0.8	0.5	0.4	0.8018	0.4992	0.3968	0.8018	0.5047	0.3968	0.0020	0.0010	0.0010	0.0020	0.0028	0.0026
-0.4	0.8	0.5	0.4	0.8018	0.4996	0.3970	0.8018	0.5051	0.3969	0.0020	0.0009	0.0008	0.0020	0.0025	0.0023
-0.3	0.8	0.5	0.4	0.8018	0.4999	0.3971	0.8018	0.5054	0.3970	0.0020	0.0008	0.0008	0.0020	0.0023	0.0022
-0.2	0.8	0.5	0.4	0.8018	0.5002	0.3972	0.8018	0.5057	0.3971	0.0020	0.0008	0.0007	0.0020	0.0022	0.0021
-0.1	0.8	0.5	0.4	0.8018	0.5005	0.3972	0.8018	0.5060	0.3972	0.0020	0.0008	0.0007	0.0020	0.0022	0.0020
-0.0	0.8	0.5	0.4	0.8018	0.5008	0.3972	0.8018	0.5063	0.3972	0.0020	0.0008	0.0007	0.0020	0.0021	0.0020

negatively correlated (n = 500)

Table 2b: *Simulation results of model Y_1 for SUR and OLS estimators with X_{11} and X_{12} negatively correlated (n = 50)*

Sample size n = 50															
$\rho_{x_{11},x_{12}}$	Model Y_1 True coefficients			SUR Estimations			OLS Estimations			SUR Average MSE			OLS Average MSE		
	β10	β11	β12	β10	β11	β12	β10	β11	β12	β10	β11	β12	β10	β11	β12
-0.9	0.8	0.5	0.4	0.8027	0.5449	0.4514	0.7968	0.5352	0.4656	0.0200	0.0522	0.0574	0.0233	0.1463	0.1625
-0.8	0.8	0.5	0.4	0.8027	0.5285	0.4374	0.7968	0.5143	0.4477	0.0200	0.0255	0.0303	0.0233	0.0715	0.0858
-0.7	0.8	0.5	0.4	0.8027	0.5206	0.4314	0.7968	0.5042	0.4400	0.0200	0.0167	0.0214	0.0233	0.0473	0.0488
-0.6	0.8	0.5	0.4	0.8027	0.5154	0.4280	0.7968	0.4976	0.4357	0.0200	0.0125	0.0170	0.0233	0.0356	0.0482
-0.5	0.8	0.5	0.4	0.8027	0.5116	0.4259	0.7968	0.4927	0.4330	0.0200	0.0101	0.0145	0.0233	0.0290	0.0412
-0.4	0.8	0.5	0.4	0.8027	0.5084	0.4245	0.7968	0.4886	0.4312	0.0200	0.0086	0.0130	0.0233	0.0249	0.0368
-0.3	0.8	0.5	0.4	0.8027	0.5057	0.4235	0.7968	0.4852	0.4300	0.0200	0.0077	0.0120	0.0233	0.0224	0.0339
-0.2	0.8	0.5	0.4	0.8027	0.5032	0.4229	0.7968	0.4820	0.4292	0.0200	0.0071	0.0114	0.0233	0.0210	0.0322
-0.1	0.8	0.5	0.4	0.8027	0.5009	0.4225	0.7968	0.4790	0.4287	0.0200	0.0068	0.0110	0.0233	0.0203	0.0312
-0.0	0.8	0.5	0.4	0.8027	0.4986	0.4224	0.7968	0.4762	0.4286	0.0200	0.0068	0.0109	0.0233	0.0202	0.0309

Table 3b: *Simulation results for model Y_2 for SUR and OLS estimators with X_{11} and X_{12} negatively correlated at various sample sizes*

$\rho_{x_{11},x_{12}}$ = - 0.0, - 0.1, - 0.2, …, - 0.9										
Sample sizes (n)	Model Y_2 True coefficients		SUR Estimations		OLS Estimations		SUR Average MSE		OLS Average MSE	
	β20	β21	β20	β21	β20	β21	β20	β21	β20	β21
1000	-0.6	0.6	-0.5976	0.6008	-0.5977	0.6002	0.0010	0.0002	0.0010	0.0003
500	-0.6	0.6	-0.5911	0.6005	-0.5911	0.6002	0.0020	0.0003	0.0021	0.0007
200	-0.6	0.6	-0.5931	0.6020	-0.5930	0.5996	0.0049	0.0009	0.0050	0.0018
100	-0.6	0.6	-0.6017	0.5945	-0.6013	0.5980	0.0095	0.0017	0.0098	0.0033
50	-0.6	0.6	-0.5918	0.6024	-0.5943	0.6113	0.0193	0.0035	0.0203	0.0073

Table 4b: *Simulation results for model Y_3 for SUR and OLS estimators with X_{11} and X_{12} negatively correlated at various sample sizes*

$\rho_{x_{11},x_{12}}$ = - 0.0, - 0.1, - 0.2, … , - 0.9										
Sample sizes (n)	Model Y_3 True coefficients		SUR Estimations		OLS Estimations		SUR Average MSE		OLS Average MSE	
	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}	β_{30}	β_{31}
1000	2.6	-0.5	2.5989	-0.4998	2.5988	-0.5016	0.0010	0.0001	0.0012	0.0003
500	2.6	-0.5	2.6017	-0.5001	2.6017	-0.5005	0.0020	0.0002	0.0020	0.0007
200	2.6	-0.5	2.6010	-0.5006	2.6004	-0.5062	0.0050	0.0005	0.0050	0.0017
100	2.6	-0.5	2.6060	-0.4983	2.6050	-0.5019	0.0096	0.0010	0.0099	0.0033
50	2.6	-0.5	2.5967	-0.5057	2.5969	-0.5082	0.0193	0.0018	0.0201	0.0061

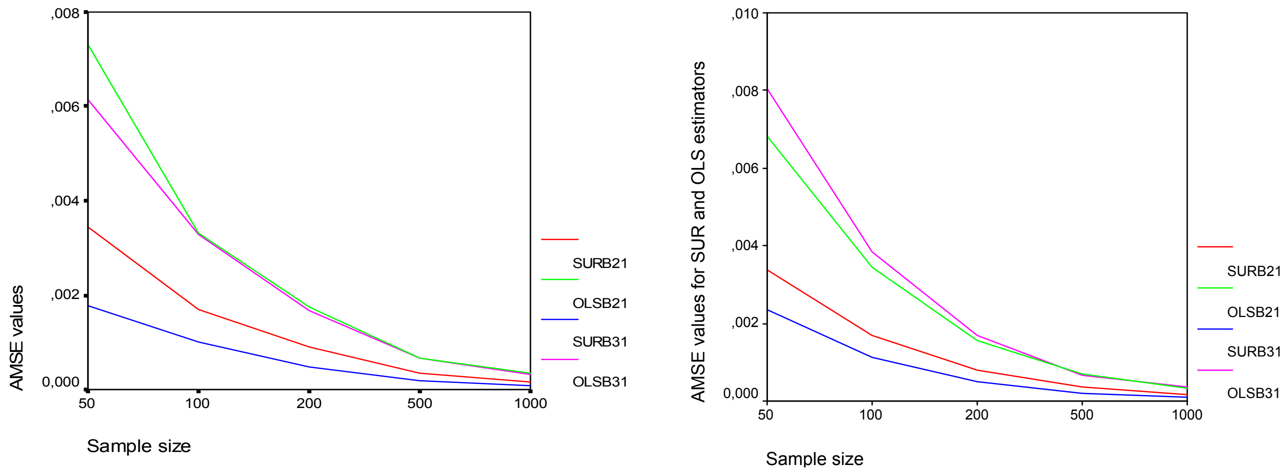
Table 5b: *Average Mean Square Error $AMSE_{Orthogonal}$ and Average Mean Square Error $AMSE_{Nonorthogonal}$ at TNCP = -0.2 for SUR estimators $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ of model Y_1 and their respective differences. The estimates of the chi-square statistics and P-values for the Bartlett Tests establishing the equality of the two AMSE are also reported.*

Sample size (n)	SUR $AMSE_{Orthogonal}$	SUR $AMSE_{Nonorthogonal}$	*Difference	Bartlett Test		SUR $AMSE_{Orthogonal}$	SUR $AMSE_{Nonorthogonal}$	*Difference	Bartlett Test	
	β_{11}	β_{11}		X^2-values	P-values	β_{12}	β_{12}		X^2-values	P-values
50	0.0068	0.0071	0.0003	0.030	0.862	0.0109	0.0114	0.0005	0.020	0.887
100	0.0041	0.0043	0.0002	0.066	0.796	0.0037	0.0039	0.0002	0.040	0.839
200	0.0019	0.0020	0.0001	0.113	0.737	0.0018	0.0019	0.0001	0.083	0.774
500	0.0008	0.0008	0.0000	0.228	0.633	0.0007	0.0007	0.0000	0.208	0.648
1000	0.0004	0.0004	0.0000	0.490	0.484	0.0004	0.0004	0.0000	0.416	0.519

**Note that when the differences is rounded up to 3 decimal places the equivalence of* $AMSE_{Orthogonal}$ *and* $AMSE_{Nonorthogonal}$ *at collinearity level of -0.2, which is our TNCP point is established*

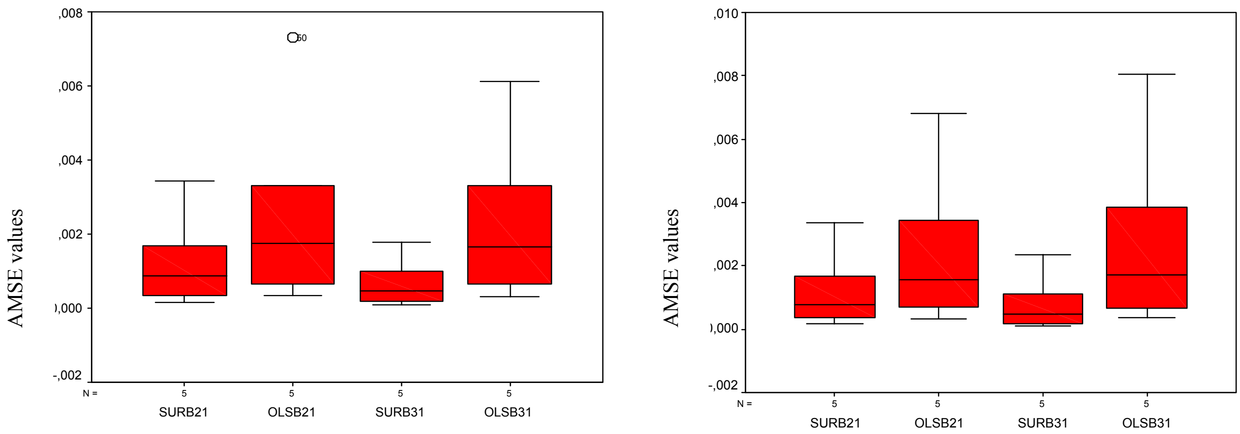
Appendix C

The list of figures



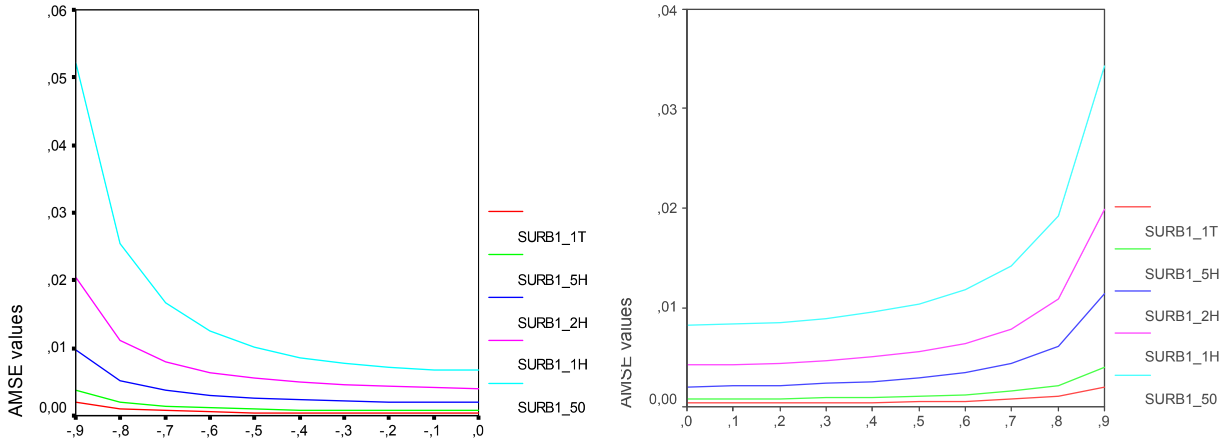
a. For Negative collinearity in model Y_1 *b. For Positive collinearity in model* Y_1

Fig 1: The plot of AMSE values against different sample sizes for SUR and OLS estimators for models Y_2 and Y_3. The various plots revealed appreciable increase in efficiency (lower AMSE) of the two estimators as sample size increases with SUR estimator showing better performance over OLS. SURB21 = SUR estimator for beta21, SURB31 = SUR estimator for beta31, OLS21 = OLS estimator for beta21, OLS31 = OLS estimator for beta31



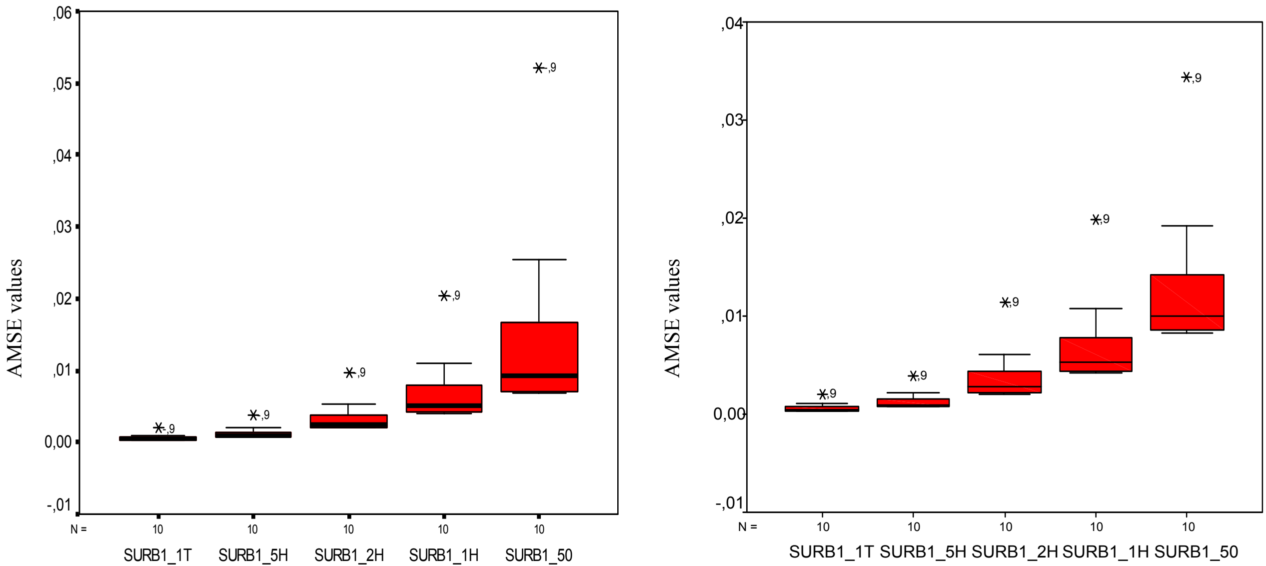
a. With negative collinearity in model Y_1 *b. With positive collinearity in model* Y_1

Fig 2: The Box- and-Whisker plot of AMSE for SUR and OLS estimators for models Y_2 and Y_3 which clearly showed lower median AMSE values for SUR estimators relative to OLS estimators. SURB21 = SUR estimator for beta21, SURB31 = SUR estimator for beta31, OLS21 = OLS estimator for beta21, OLS31 = OLS estimator for beta31



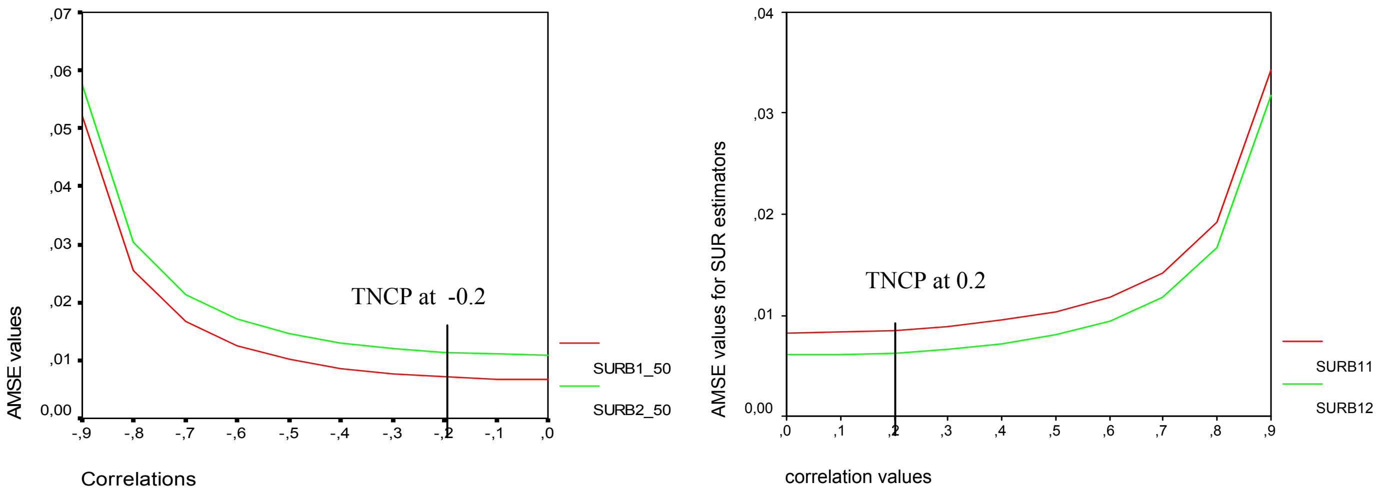
a. With negative collinearity in model Y_1 *b. With positive collinearity in model Y_1*

Fig 3: The plot of AMSE values against various correlation (collinearity) levels for SUR estimator $\hat{\beta}_{11}$ in model Y_1 at various sample sizes. The plots showed better performance of SUR estimator as sample size increases irrespective the level of collinearity among the predictors. SURB1_1T = SUR estimator for beta11 at n = 1000, SURB1_5H = SUR estimator for beta11 at n = 500, SURB1_2H = SUR estimator for beta11 at *n = 200, SURB1_1H = SUR estimator for beta11 at n = 100, SURB1_50 = SUR estimator for beta11 at n* = 50.



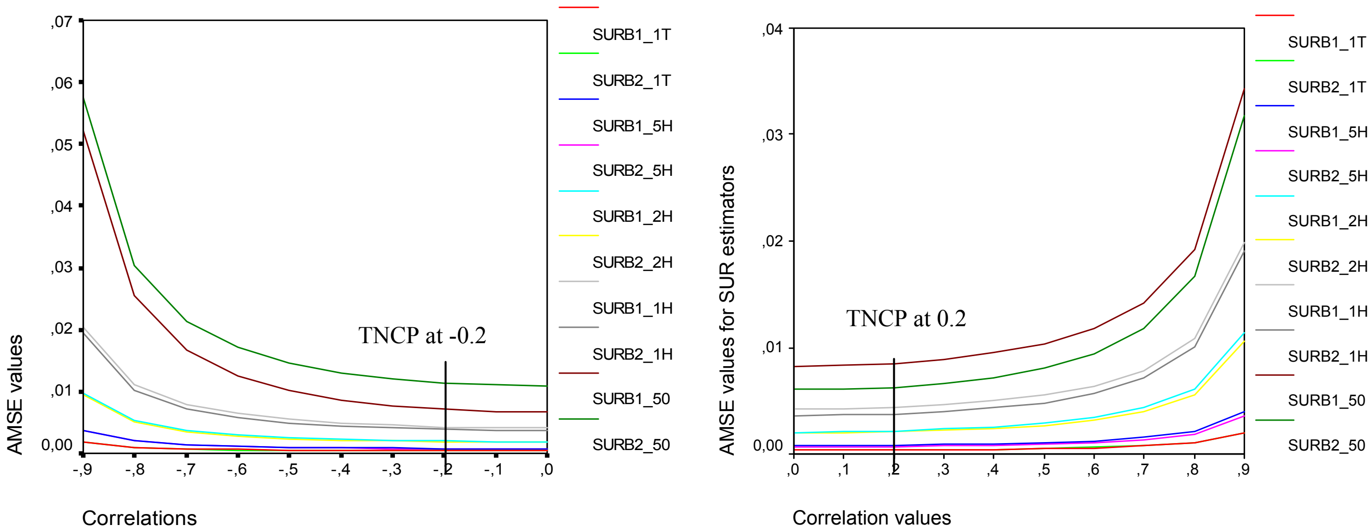
a. With negative collinearity in model Y_1 *b. With positive collinearity in model Y_1*

Fig 4: The Box- and-Whisker plot of AMSE for SUR estimator at different sample sizes for model Y_1 in which a decreasing trend in the median AMSE is observed as sample size increases for SUR estimator. Ten correlation points were present in each sample. SURB1_1T = SUR estimator for beta11 at n = 1000, SURB1_5H = SUR estimator for beta11 at n = 500, SURB1_2H = SUR estimator for beta11 at n = 200, SURB1_1H = SUR estimator for beta11 at n = 100, SURB1_50 = SUR estimator for beta11 at n = 50.



a. With negative collinearity in model Y_1 *b. With positive collinearity in model Y_1*

Fig 5: The plot of AMSE values against the correlation points for SUR estimators β_{11} and β_{12} at sample size of 50. The TNCP point of ±0.2 is shown after which up-ward movements in the graphs started. SURB11 = SURB1_50 = SUR estimator for beta11 at n = 50, SURB12 = SURB2_50 = SUR estimator for beta12 at n = 50.



a. With negative collinearity in model Y_1 *b. With positive collinearity in model Y_1*

Fig 6: The plot of AMSE values against correlation points for SUR estimators $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ at all the sample sizes. The TNCP point of ±0.2 is shown after which up-ward movements in the graphs started. SURB1_1T = SUR estimator for beta11 at n = 1000, SURB2_1T = SUR estimator for beta12 at n = 1000, SURB1_5H = SUR estimator for beta11 at n = 500, SURB2_5H = SUR estimator for beta12 at n = 500, SURB1_2H = SUR estimator for beta11 at n = 200, SURB2_2H = SUR estimator for beta12 at n = 200, SURB1_1H = SUR estimator for beta11 at n = 100, SURB2_1H = SUR estimator for beta12 at n = 100, SURB1_50 = SUR estimator for beta11 at n = 50, SURB2_50 = SUR estimator for beta12 at n = 50.

References

[1] Adebayo, S. B., 2003, Semiparametric Bayesian Regression for Multivariate Response. *Hieronymus, Munich.*

[2] Aitken, A. C., 1934-35, On Least-Squares and Linear Combination of Observations. *Proceedings of The Royal Society of Edinburgh.* 55: 42 - 48.

[3] Amemiya, T ., 1977, A note on a Heteroscedastic Model. *Journal of Econometrics,* 6: 365 - 370

[4] Blattberg, R., and George, E. I., 1991, Shrinkage estimation of price and optional elasticities: Seemingly Unrelated Equations. *Journal of American Statistical Association*. 86: 304-315.

[5] Dwivedi, T. D. and Srivastava, V. K., 1978, Optimality of Least Squares in the Seemingly Unrelated Regression Equation Model. *Journal of Econometrics*. 7: 391-395.

[6] Foschi, P., 2004, Numerical Methods for estimating Linear Econometric Models. Ph. D. Thesis. Institute d'informatique, Université de Neuchâtel, Switzerland.

[7] Foschi, P., Belsley, D. A. and Kontoghiorghes, E. J., 2003, A comparative study of algorithms for solving Seemingly Unrelated Regressions Models. *Computational Statistics & Data Analysis. (In press)*

[8] Foschi, P., and Kontoghiorghes, E. J., 2002, Estimation of Seemingly Unrelated Regressions Models with unequal size of observations. *Computational Statistics & Data Analysis, 41(1): 211 - 229*

[9] Foschi, P., and Kontoghiorghes, E. J., 2003, Estimating Seemingly Unrelated Regressions Models with Orthogonal Regressors: Computational aspects. *Linear Algebra and Its Applications. (In press).*

[10] Horn, R. A. and Johnson, C. R., 1985, Matrix Analysis, Section 7.2. *Cambridge University Pres.*

[11] Jackson, J.E., 2002, A Seemingly Unrelated Regression Model for Analyzing Multiparty Elections. *Political Analysis*. 10: (1), 49-65.

[12] Julier, S. J. and Uhlmann, J.K., 1997, A new extension of the Kalman filter to nonlinear systems. *Proceedings of AeroSense: 11th International Symposium of Aerospace/Defence Sensing, Simulation and Controls.*182-193.

[13] Kakwani, N. C., 1967, The unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators. *Journal of the American Statistical Association,* 62: 141 – 142.

[14] Kmenta, J. and Gilbert, R. F., 1968, Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regression. *Journal of American Statistical Association*. 63: 1180-1200.

[15] Kontoghiorghes, E. J., 2003, Computational Methods for Modifying SUR Models. *Journal of Computational and Applied mathematics. (In press).*

[16] Kontoghiorghes, E. J. and Clerke, M. R. B. (1995): An Alternative Approach for the Numerical solution of SUR Equations Models. *Computational Statistics & Data Analysis, 199(4); 369 – 377.*

[17] Kontoghiorghes, E. J. and Dinenis, E., 1996, Solving Triangular SUR Equations Models on Massively Parallel systems. *In M. Gilli, editor, Computational Economic Systems: Models, Methods 7 Econometrics, Vol. 5 of Advances in Computational Economics, Kluwer Academic publishers, 191 – 201.*

[18] Kontoghiorghes, E. J. and Dinenis, E., 1997, Towards the parallel implementation of the SUR Model Estimation Algorithm. Journal of Mathematical Modeling and Scientific *Computing, 8: 335 – 341.*

[19] Kurata, H., 1999, On the efficiencies of several Generalised Least Squares Estimators in a Seemingly Unrelated Regression model and Heteroscedastic model. *Journal of Multivariate Analysis*. 70: 86-94.

[20] Lang, S., Adebayo, S. B., and Fahrmeir, L., 2002, Bayesian Semiparametric Seemingly Unrelated Regression. *In: Härdle, W. and Roenz, B. (eds): 195 – 200. Physika-Verlag, Heidelberg.*

[21] Lang, S., Adebayo, S. B., Fahrmeir, L. and Steiner, W. J., 2003, Bayesian Geoadditive Seemingly Unrelated Regression. *Computational Statistics, 18(2): 263 – 292.*

[22] Liu, J. S. and Wang, S. G., 1999, Two-Stage estimate of the parameters in Seemingly Unrelated Regression Model. *Progress in Natural Science. 9.*

[23] Maeshiro, A., 1980, New evidence on Small sample Properties of Estimators of SUR models with Autocorrelated Disturbances. *Journal of Econometrics*. 12: 177-187.

[24] Mandy, D. M. and Martins-Filho, C., 1993, Seemingly Unrelated Regressions Under Additive Heteroscedasticity: Theory and share equation applications. *Journal of Econometrics*, 58: 315 – 346.

[25] Mehta, J. S. and Swamy, P. A. V. B., 1976, Further Evidence on the Relative Efficiencies of Zellner's Seemingly Unrelated Regression Estimator. *Journal of American Statistical Association*. 71: 634-639

[26] Neter, J. and Wasserman, W., 1974, Applied Linear Statistical Models. *Homewood, Illinois, Irwin.*

[27] Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W., 1996, Applied Linear Regression Models. *3rd Ed. Richard D. Irwin.*

[28] O'Donnel, C. J. Shumway, C. R. and Ball, V. E., 1999, Inputs Demands and Inefficiency in U.S Agriculture. *American Journal of Agricultural Economics*. 81: 865-880.

[29] Parks, R. W., 1967, Efficient estimation of a system of Regression Equations when disturbances are Both Serially and Contemporaneously correlated. . *Journal of the American Statistical Association,* 62: 500 – 509.

[30] Powell, A. A., 1965, Aitken Estimators as a tool in Allocating Predetermined Aggregates. *Journal of the American Statistical Association*. 64: 913 – 922.

[31] Revankar, N. S. (1974): Some Finite Sample Results in the context of two Seemingly Unrelated Regression equations. *Journal of the American Statistical Association*. 69: 187-190.

[32] Revankar, N. S., 1976, Use of Restricted Residuals in SUR system: Some Finite Sample Results. *Journal of the American Statistical Association*. 71: 183-192.

[33] Silvey, S. D., 1969, Multi-collinearity and Imprecise Estimation. *Journal of Royal Statistical Society*, 31: 539 – 552.

[34] Smith, M. and Kohn, R., 2000, Nonparametric Seemingly Unrelated Regression. *Journal of Econometrics*, 98: 257 – 281.

[35] Sparks, R., 2004, SUR Models Applied To an Environmental Situation with Missing Data and Censored Values. *Journal of Applied Mathematics and Decision Sciences*. 8: (1), 15-32.

[36] Taylor, J. M. G., Schwartz, K. and Detels, R., 1986, The time from infection with human Immunodeficiency virus (HIV) to the onset AIDS. *Journal of Infectious Disease, 154, 694 – 697.*

[37] Taylor, J. M. G., Bass, A. S. M., Saah, A. J., Chmiel, J. S., Kingsley, L. A. and The Multicentre AIDS cohort study '*centres & investigators*' , 1990, Estimating the Distribution of times from HIV Seroconversion to AIDS using Multiple Imputation. *Statistics In Medicine, 9, 505 – 514.*

[38] Telser, L. G., 1964, Iterative estimation of a Set of Linear Regression Equations. *Journal of The American statistical association, 59, 845 – 862.*

[39] Wilde, P.E., McNamara, P. E., and Ranney, C. K., 1999, The Effect of Income and Food programs on Dietary Quality: A Seemingly Unrelated Regression Analysis

with Error components. *American Journal of Agricultural economics*. 81: (4), 959-971.

[40] Zellner, A., 1962, An efficient Method of Estimating Seemingly Unrelated Regression Equations and Test for Aggregation Bias. *Journal of the American Statistical Association, 57: 348 – 368.*

[41] Zellner, A. and Theil, H., 1962, Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica.* 30: 54 – 78.

[42] Zellner, A., 1963, Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results. *Journal of the American Statistical Association,* 58: *977 – 992.*