J. Ind. Soc. Agril. Statist.
62(1), 2008 : 34-41

Prediction for Seemingly Unrelated Regressions with Autocorrelated Errors

Amitava Dey, Himadri Ghosh and V.K. Sharma
Indian Agricultural Statistics Research Institute, New Delhi
(Received : October 2007)

SUMMARY

In regression models when the errors are correlated, the sample residuals contain some information about the future observations. This information, which is generally ignored, has been used in this paper to improve the precision of predicting the post-sample observations. The best linear unbiased predictor for an m-equation linear SURE model has been obtained under the assumption that the errors in each equation follow first-order autoregressive scheme. The gain in efficiency of the proposed predictor over the usual generalized least squares predictor has been obtained and the method is illustrated for a two-equation acreage response model. Small sample properties of the predictor have been studied by using a Monte-Carlo experiment.

Key words: Seemingly unrelated regression equations, Best linear unbiased predictor, GLS predictor, Autoregressive errors.

1. INTRODUCTION

Seemingly Unrelated Regression Equations (SURE), introduced by Zellner (1962), are a set of interrelated statistical equations that have wide applicability in the analysis of data in social sciences and other fields. These models take care of hidden interactions that are present in different equations of the model each one of which describes a certain aspect of behavior (Srivastava and Giles 1987). The equations of the model are linked statistically, through the joint distribution of the error terms in different equations and through the non-diagonality of the associated variance-covariance matrix. This possibility of nonzero covariances between error terms of different equations of the model was first visualized by Zellner (1962), who coined the term "Seemingly Unrelated Regression Equations" (SURE) to reflect the fact that equations which are apparently not connected or related structurally are, in fact, related to each other statistically.

On one hand, simultaneous estimation of the non-apparently related regression equations of the model improves the precision of the estimation of regression parameters over the situations in which the individual equations are estimated independently of each other. On the other hand, this also ensures the validity of test procedures. It was recognition of this fact, especially with regard to the asymptotic efficiency of the estimators of the parameters of SURE model, which motivated Zellner's original work.

The i^{th} regression equation of a SURE model consisting of m equations, may be written as

$$y_i = X_i\beta_i + \varepsilon_i,\ i = 1, 2, ..., m \qquad (1.1)$$

where y_i is T × 1 vector of time series observations on the dependent variable, X_i is T × p_i matrix of observations on non-stochastic explanatory variables with full column rank, β_i is the vector of parameters of order p_i, and ε_i is the error term of order T × 1 with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i \varepsilon'_j) = \Omega_{ij,}$ a matrix of order T × T; Ω_{ii} being positive definite symmetric matrices (i, j = 1, 2, ..., m).

In compact form the model (1.1) can be written as

$$y = X\beta + \varepsilon \qquad (1.2)$$

where $y = (y'_1\ y'_2 \ldots y'_m)'$ is an mT dimensional column vector.

$X = \text{diag}\ (X_1, X_2, \ldots X_m)$ is the matrix of order

$$mT \times \sum_{i=1}^{m} pi$$

$\beta = (\beta_1' \ \beta_2' \ \ldots \beta_m')'$ is a column vector of dimension

$$\sum_{i=1}^{m} p_i$$

$\varepsilon = (\varepsilon_1' \ \varepsilon_2' \ \ldots \varepsilon_m')'$ is a column vector of dimension mT, with $E(\varepsilon) = 0$

and $E(\varepsilon\varepsilon') = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \ldots & \Omega_{1m} \\ \Omega_{21} & \Omega_{22} & \ldots & \Omega_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ \Omega_{m1} & \Omega_{m2} & \ldots & \Omega_{mm} \end{bmatrix} = \Omega$ (say), is a positive definite symmetric matrix of dimension mT × m T.

Zellner's work stimulated extensive theoretical work and many empirical applications in econometrics and other areas (see e.g., Richard and Steel 1988, Sharma 1993, Chib and Greenberg 1995, Smith and Kohn 2000, Deo and Rong 2000, Srivastava and Wan 2002). Percy (1992) studied the problem of prediction for SURE model from the Bayesian perspective. Using Jeffreys' invariant prior Percy obtained the posterior density function of post sample observations through Gibbs sampling procedure.

If errors in each equation follow a first order autoregressive process, then

$$\varepsilon_{it} = \rho_i \varepsilon_{i(t-1)} + e_{it}, \ |\rho_i| < 1$$

$$i = 1, \ldots, m \text{ and } t = 2, \ldots, T \qquad (1.3)$$

where ε_{it} is the t^{th} element of ε_i, e_{it} is white noise such that

$$E(e_{it}) = 0, \ i = 1, \ldots, m \text{ and } t = 1, \ldots, T$$

and $E(e_{it}\, e_{js}) = \begin{cases} \sigma_{ij} & \text{for } t = s;\ i, j = 1, \ldots, m \\ 0 & \text{for } t \neq s;\ i, j = 1, \ldots, m \end{cases}$

It is easily seen that

$$E(\varepsilon_i \varepsilon_j') = \Omega_{ij} = \sigma_{ij}^* \begin{bmatrix} 1 & \rho_j & \ldots & \rho_j^{T-1} \\ \rho_i & 1 & \ldots & \rho_j^{T-1} \\ \ldots & \ldots & \ldots & \ldots \\ \rho_i^{T-1} & \rho_i^{T-2} & \ldots & 1 \end{bmatrix} \qquad (1.4)$$

where $\sigma_{ij}^* = \dfrac{\sigma_{ij}}{1 - \rho_i \rho_j}$

When the observations are correlated, the least-squares residuals from (1.1) contain information about the future observations. This information, which is generally ignored, has been exploited here in predicting post-sample observations. Best linear unbiased predictor has been obtained from classical theory point of view assuming that the errors in each equation follow AR (1). The gain in efficiency of the proposed predictor over the usual generalized least-squares predictor has been obtained. The method of obtaining prediction is illustrated using a two-equation acreage response model. To observe the effect of magnitude of correlation between explanatory variables across the equations and of covariance between the error terms of the two equations on the prediction efficiency, a Monte-Carlo experiment has also been carried out using a two-equation model given in Kmenta and Gilbert (1968) incorporating first order autocorrelation structure in errors.

2. BEST LINEAR UNBIASED PREDICTION

Our problem is to predict a single drawing of the response y_i, for given x_i^*, the row vector of regressors with p_i components (i = 1, 2, ..., m).

Denoting the (T + 1)th observation from ith regression equation as y_i^*, i = 1,..., m from equation (1.1), this can be modeled as

$$y_i^* = x_i^* \beta_i + \varepsilon_i^*; \ i = 1, \ldots, m \qquad (2.1)$$

where ε_i^* is the scalar value of the prediction disturbance with

$$E(\varepsilon_i^*) = 0$$

$$E(\varepsilon_i^* \ \varepsilon_j^*) = \frac{\sigma_{ij}}{1-\rho_i\rho_j} = \sigma_{ij}^*, \ i, j = 1, ..., m \qquad (2.2)$$

In general, it is not reasonable to assume that the out of sample error is independent of the sample errors. A more reasonable assumption is one which allows the out of sample disturbance, ε_i^* to be correlated with the sample disturbances ε_j in (1.3) in the same fashion as in the sample.

It can be seen that

$$Cov(\varepsilon_j, \ \varepsilon_i^*) = \frac{\sigma_{ij}}{1-\rho_i\rho_j}\begin{bmatrix} \rho_j^T \\ \rho_j^{T-1} \\ \vdots \\ \rho_j \end{bmatrix} = \omega_{ij}, \text{ say; } i, j = 1, ..., m \qquad (2.3)$$

Here, ω_{ij}, (i, j = 1, ..., m) represent T × 1 vectors of covariance between post-sample error and sample errors.

Defining ε* as an m × 1 vector of ε_i^*'s, it is easy to see that $E(\varepsilon^*) = 0$ and

$$E(\varepsilon^*\varepsilon^{*\prime}) = \begin{bmatrix} \sigma_{11}^* & \sigma_{12}^* & \cdots & \sigma_{1m}^* \\ \sigma_{21}^* & \sigma_{22}^* & \cdots & \sigma_{2m}^* \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m1}^* & \sigma_{m2}^* & \cdots & \sigma_{mm}^* \end{bmatrix} = \Omega^{*m\times m}$$

and

$$E(\varepsilon\varepsilon^{*\prime}) = E\left[\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}(\varepsilon_1^* \ \ \varepsilon_2^* \ \ ... \ \ \varepsilon_m^*)\right]$$

$$= \begin{bmatrix} \omega_{11}^{T\times1} & \omega_{12}^{T\times1} & \cdots & \omega_{1m}^{T\times1} \\ \omega_{21}^{T\times1} & \omega_{22}^{T\times1} & \cdots & \omega_{2m}^{T\times1} \\ \vdots & \vdots & \vdots & \vdots \\ \omega_{m1}^{T\times1} & \omega_{m2}^{T\times1} & \cdots & \omega_{mm}^{T\times1} \end{bmatrix} = W^{Tm\times m} \text{ (say)} \qquad (2.4)$$

where superscripts in $\Omega^{*m\times m}$ and $W^{Tm\times m}$ denote the dimensions of the matrices.

Let $y_{pi} = c_{i1}'y_1 + c_{i2}'y_2 + ... + c_{im}'y_m; \ (i = 1, ..., m)$

be an unbaised predicator for y_i^*, where c_{ij}'s are T × 1 vector of constants. Now, from (1.1)

$$y_{pi} = \sum_{j=1}^{m} c_{ij}'X_j\beta_j + \sum_{j=1}^{m} c_{ij}'\varepsilon_j \qquad (2.5)$$

This gives

$$E(y_{pi}) = \sum_{j=1}^{m} c_{ij}'X_j\beta_j$$

$$= c_{ii}'X_i\beta_i + \sum_{j(\neq i)=1}^{m} c_{ij}'X_j\beta_j \qquad (2.6a)$$

$$= x_i^*\beta_i \forall \beta_i, \text{ iff } c_{ii}'X_i = x_i^*$$

$$\text{and } c_{ij}'X_j = 0 \forall j(\neq i) = 1,..., m \qquad (2.6b)$$

Thus for unbiased predictions, the prediction errors are

$$y_{pi} - y_i^* = \sum_{j=1}^{m} c_{ij}'\varepsilon_j - \varepsilon_i^*, \ \ i = 1,...,m \qquad (2.7)$$

Stacking in equations in (2.7) together, we get

$$y_p - y^* = \begin{bmatrix} c_{11}' & c_{12}' & ... & c_{1m}' \\ c_{21}' & c_{22}' & ... & c_{2m}' \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1}' & c_{m2}' & ... & c_{mm}' \end{bmatrix}\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} - \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_m^* \end{bmatrix}$$

$$= C\varepsilon - \varepsilon^* \qquad (2.8)$$

where $y_p = (y_{p1} \ y_{p2} \ \cdots \ y_{pm})'$

and $y^* = (y_1^* \ y_2^* \ \cdots \ y_m^*)'$

From (2.8) the variance-covariance matrix of the prediction error is

$$\begin{aligned}\Omega_p &= \text{var}(y_p - y^*) \\ &= E(C\varepsilon - \varepsilon^*)(C\varepsilon - \varepsilon^*)' \\ &= C\Omega C' - CW - W'C' + \Omega^* \end{aligned} \quad (2.9)$$

Then trace of Ω_p is

$$\begin{aligned}\text{tr}(\Omega_p) &= \text{tr}(C\Omega C') - 2\text{tr}(CW) - \text{tr}(\Omega^*) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} c'_{ik}\Omega_{kj}c_{ij} \\ &\quad - 2\sum_{i=1}^{m}\sum_{j=1}^{m} c'_{ij}\omega_{ij} + \sum_{i=1}^{m}\sigma^*_{ii} \end{aligned} \quad (2.10)$$

Now to obtain the best predictor in the sense of minimizing $\text{tr}(\Omega_p)$, we minimize (2.10) subject to the conditions (2.6b). That is, for best linear unbiased predictor we minimize the function

$$\begin{aligned}\Phi = &\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} c'_{ik}\Omega_{kj}c_{ij} - 2\sum_{i=1}^{m}\sum_{j=1}^{m} c'_{ij}\omega_{ij} \\ &+\sum_{i=1}^{m}\sigma^*_{ii} - 2\sum_{i=1}^{m}\lambda_{ii}'(X_i'c_{ii} - x_i^*) - 2\sum_{i=1}^{m}\sum_{j(\neq i)=1}^{m}\lambda'_{ij}X_j'c_{ij}\end{aligned}$$

where λ_{ij}'s (i, j = 1, ..., m) are $p_j \times 1$ vectors of Lagrangian multipliers. The first order conditions for minimization of ϕ with respect to c_{ij}'s and λ_{ij}'s, after re-arrangement of terms, can be written as

$$\begin{bmatrix}\Omega_d & X_d \\ X_d' & 0\end{bmatrix}\begin{bmatrix}\hat{\theta} \\ -\hat{\lambda}\end{bmatrix} = \begin{bmatrix}\underline{\omega} \\ \underline{x}^*\end{bmatrix} \quad (2.11)$$

where

$\Omega_d = I_m \otimes \Omega$, ($I_m$ being an indentity matrix of order m)

$X_d = I_m \otimes X$

$\underline{\omega} = (\omega_1' \ \omega_2' \ldots \omega_m')'$, ω_i being the i^{th} column of W

$\underline{x}^* = \left(x^{*'}_{(1)} \ x^{*'}_{(2)} \ \ldots \ x^{*'}_{(m)}\right)'$

(where $x^*_{(i)}$ is the i^{th} column of

$X^* = \text{diag}\left(x_1^{*'} \ x_2^{*'} \ \ldots \ x_m^{*'}\right)$

$\hat{\theta} = (\hat{c}'_{11} \ldots \hat{c}'_{1m} \ \hat{c}'_{21} \ldots \hat{c}'_{2m} \ldots \hat{c}'_{m1} \ldots \hat{c}'_{mm})$ is $m^2T \times 1$ vector, and

$\hat{\lambda} = \left(\hat{\lambda}'_{11} \ldots \hat{\lambda}'_{1m} \ \hat{\lambda}'_{21} \ldots \hat{\lambda}'_{2m} \ldots \hat{\lambda}'_{m_1} \ldots \hat{\lambda}'_{mm}\right)$ is

$m\left(\sum_{i=1}^{m} p_i\right) \times 1$ vector.

Solution of (2.11) yields

$$\begin{aligned}\hat{\theta} &= \Omega_d^{-1}X_d(X_d'\Omega_d^{-1}X_d)^{-1}\underline{x}^* \\ &\quad +\Omega_d^{-1}\left[1 - X_d\left(X_d'\Omega_d^{-1}X_d\right)^{-1}X_d'\Omega_d^{-1}\right]\underline{\omega} \\ &= \begin{bmatrix}\Omega^{-1}X(X'\Omega^{-1}X)^{-1}x^*_{(1)} + \Omega^{-1}\left[1 - X(X'\Omega^{-1}X)^{-1}\right]\omega_1 \\ \Omega^{-1}X(X'\Omega^{-1}X)^{-1}x^*_{(2)} + \Omega^{-1}\left[I - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\omega_2 \\ \Omega^{-1}X(X'\Omega^{-1}X)^{-1}x^*_{(m)} + \Omega^{-1}\left[I - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\omega_m\end{bmatrix}\end{aligned} \quad (2.12)$$

Therefore, the best linear unbiased predictor for y_i^* is given by

$$y_{pi} = \hat{c}_i'y \quad (2.13)$$

where $\hat{c}_i = (\hat{c}'_{i1} \ldots \hat{c}'_{im})'$ is the i^{th} sub-vector of $\hat{\theta}$. Substitution of the value of $\hat{c}_i$ from (2.12), yields

$$\begin{aligned}y_{pi} &= x^*_{(i)}\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y + \omega_i'\Omega^{-1}y \\ &\quad - \omega_i'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \\ &= x^*_{(i)}\hat{\beta}_G + \omega_i'\ \Omega^{-1}\hat{e}_G \end{aligned} \quad (2.14)$$

where $\hat{\beta}_G = (X'\Omega^{-1}X)X'\Omega^{-1}y$ is the generalized least-squares (GLS) estimator and $\hat{e}_G$ is the $mT \times 1$ vector of sample residuals from the GLS regression.

Substituting the values of $\hat{c}_i$ (i = 1, ..., m) (2.9) yields the error variance due to y_{pi} as

$$\sigma^2_{pi} = \hat{c}_i'\Omega\hat{c}_i + \sigma^*_{ii} - 2\hat{c}_i\omega_i$$

$$= x^{*}_{(i)}{}'\left(X'\Omega^{-1}X\right)^{-1}x^*_{(i)} + \sigma^*_{ii}$$

$$-2x^*_{(i)}(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\omega_i$$

$$-\omega_i'\left[\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\omega_i \quad (2.15)$$

3. GAIN IN EFFICIENCY

The conventional linear unbiased predictor ignoring the correlation between sampled and post-sample unobserved error is

$$\tilde{y}_{pi} = x^{*}_{(i)}{}'\hat{\beta}_G \qquad i = 1, ..., m \quad (3.1)$$

Now, we work out the gain in efficiency of the proposed predictor, y_{pi}, over the predictor $\tilde{y}_{pi}$. Considering, $\tilde{y}_{pi} = \tilde{c}_i'y$ (3.1) amounts to choosing $\tilde{c}_i$ as

$$\tilde{c}_i = \Omega^{-1}X(X'\Omega^{-1}X)^{-1}x^*_{(i)} \quad (3.2)$$

and the error variance due to the predictor $\tilde{y}_{pi}(i = 1, ..., m)$ is seen to be

$$\sigma^2_{\tilde{y}_{pi}} = \tilde{c}_i'\Omega\tilde{c}_i - 2\tilde{c}_i'\omega_i + \sigma^*_{ii}$$

$$= x^{*'}_{(i)}\left(X'\Omega^{-1}X\right)^{-1}x^*_{(i)} + \sigma^*_{ii} - 2x^{*'}_{(i)}\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\omega_i$$

$$= \sigma^2_{pi} + w^*_i\left[\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1})\right]\omega_i$$

$$= \sigma^2_{pi} + \zeta\Omega^{-1}\zeta_i \quad (3.3)$$

where $\zeta_i = \left[I - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]\omega_i$ is the column vector with mT elements. The second term in (3.3) indicates the gain in efficiency over $\tilde{y}_{pi}$ (i = 1, ..., m) as this term is seen to be non-negative.

When σ_{ij}'s and ρ_i's are known, we obtain the best linear unbiased predictions along with their variances. The parameters σ_{ij}'s and ρ_i's when unknown, may be replaced by their consistent estimates in (2.14) and (2.15) to yield predictions that are asymptotically efficient.

Kmenta and Gilbert (1970) examined the small-sample efficiency of four different methods of estimation of regression and autocorrelation coefficients by conducting a Monte Carlo experiment and found that the Joint Nonlinear Estimation method performs better for small samples in their setup.

In the following section, an experiment has been carried out using a two-equation model given in Kmenta and Gilbert (1968) incorporating first-order autoregressive structure in errors, to observe the effect of magnitude of correlation between explanatory variables across the equations on the prediction efficiency. Estimates of σ_{ij}'s and ρ_i's obtained by the Joint Nonlinear Estimation method have been used in (2.14) and (2.15) for obtaining the predictions and their estimated variances. Thus, the present paper extends the scope of Kmenta and Gilbert work in the context of joint predictions.

4. ILLUSTRATION

We obtain predictions from a two-equation acreage response model for two competing winter crops viz., rapeseed and wheat (data given in Appendix) using the results of Section 2 for Haryana State of India. The current acreages under rapeseed (AR_t) and wheat (AW_t) have been considered as the dependent variables and time series data have been taken from 1983-84 to 1998-99 from various published sources. For rapeseed crop explanatory variables were: one year lagged yield (YR_{t-1}) and one year lagged price (PR_{t-1}). The explanatory variables for the wheat crop were also taken as one year lagged yield (YW_{t-1}) and one year lagged price (PW_{t-1}). The proposed two-equation acreage response SURE model for rapeseed and wheat is

$$AR_t = \alpha_0 + \alpha_1 YR_{t-1} + \alpha_2 PR_{t-1} + \varepsilon_{1t}$$

$$AW_t = \beta_0 + \beta_1 YW_{t-1} + \beta_0 PW_{t-1} + \varepsilon_{2t} \quad (4.1)$$

where suffix t denotes the value of the variable in year t, and ε_{it} (i =1, 2) are the stochastic error terms in the

two equations which follow first order autoregressive process given in (1.3).

As the first step, ordinary least squares method was applied to each of the equations in (4.1) to obtain the initial estimates of $\rho_1, \rho_2, \sigma_{11}, \sigma_{22}$ and $\sigma_{12} (= \sigma_{21})$.Then, Joint Nonlinear Estimation method (Kmenta and Gilbert, 1970) was used to estimate the parameters of the model (4.1) by replacing the quantities in with their respective initial estimates. The final estimates obtained were

$\hat{\rho}_1 = 0.612,\ \hat{\rho}_2 = -0.082,\ (\hat{\alpha}_0,\ \hat{\alpha}_1,\ \hat{\alpha}_2,\ \hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\beta}_2)' =$ (11.517, 1.260, 0.026, 142.562, 0.906, 0.062)

$\hat{\sigma}_{11} = 36.657,\ \hat{\sigma}_{22} = 22.954$ and $\hat{\sigma}_{12} = \hat{\sigma}_{21} = 15.760.$

The estimated values of σ^*_{ij}, viz. $\hat{\sigma}^*_{11}, \hat{\sigma}^*_{22}, \hat{\sigma}^*_{12}$ were computed by replacing the parametric values in (2.2) with their estimates which were found to be

$\hat{\sigma}^*_{11} = 58.608,\ \hat{\sigma}^*_{22} = 23.109,\ \hat{\sigma}^*_{12} = 15.006.$

For given $x^*_1 = (1, 16.5, 1200)$ and $x^*_2 = (1, 42, 520)$ vectors for the year 1999-2000, the predicted values from (2.14) are seen to be $y_{p1} = 600.40$ thousand hectares and $y_{p2} = 2103.50$ thousand hectares with estimated error variances from (2.15) as, 3974.10 and 3070.95 square thousand hectare respectively. In view of the available data, the predicted values based on the proposed predictors seem to be quite reasonable for all practical purposes.

5. MONTE CARLO EXPERIMENT FOR PREDICTIONS

We considered the following two-equation model (Kmenta and Gilbert, 1968)

$$y_{1t} = 10 + 2x_{11t} + 5x_{12t} + \varepsilon_{1t}$$

$$y_{2t} = -10 + 6x_{21t} - 3x_{22t} + \varepsilon_{2t} \qquad (5.1)$$

with $\varepsilon_{1t} = 0.5\varepsilon_{1(t-1)} + e_{1t}$ and $\varepsilon_{2t} = -0.5\ \varepsilon_{2(t-1)} + e_{2t}$

Two different sets of X's were used. The first is the set given in Kmenta and Gilbert (1968, p.1186) for their experiment 1, and the second set is given in Kmenta and Gilbert (1968, p.1187) for their experiment 2. Ten observations are listed in each case for x_{11}, x_{12}, x_{21} and x_{22}; and these are repeated as many times as necessary to generate the required number of observations. The difference between the two sets of X's is in the correlation between x_{11} and x_{21}, and x_{12} and x_{22}. These correlations are low for the first set of X's and high for the second set of X's. Zellner and Huang (1962) have shown that the asymptotic variance of the Seemingly Unrelated Regression estimator is low when this correlation is low. As a result it may be worthwhile to use these two different sets of X's to see if it makes a difference in the present context. In all the cases the variances of e_1 and e_2 are taken as one, and three values of the correlation between e_1 and e_2 were used, namely, 0.3, 0.5 and 0.9. Four sample sizes were considered

T = 15, 20, 30 and 50 and x* was taken as

$x^*_1 = (1\ 4\ 1)$, and $x^*_2 = (1\ 5\ 1)$

For each given model specification (X matrix, value of σ_{12} and value of T) a sample from bi-variate normal distribution was generated using IML in SAS software. Using the autoregressive scheme in Section 2, we generated the series of it ε_{it}, i = 1, 2; t = 1, ..., T.

From each sample, autocorrelation coefficients were first calculated using the Joint Nonlinear Estimation (JOINTEST) method (Kmenta and Gilbert, 1970) and the estimated values of ρ_i, i = 1, 2; were used to generate the series using the autoregressive scheme in (5.1). The estimates of σ_{11}, σ_{12} and σ_{22} were calculated by using the sample variance-covariance matrix of e_1 and e_2 . The prediction and prediction variance were calculated as described in Section 2.

Table 1 gives the estimated variances of BLUP and GLS predictor. Zellner (1962) showed that the asymptotic variance of the SURE estimator was smaller when the correlation between the white noise it e_{it}'s of the two equations was larger. A glance at Table 1 reveals that this is also true for BLUP, as proposed in this paper. It has also been found that the BLUP is far more superior to GLS prediction estimator when errors in each equation follow AR(1) with value of auto-correlation 0.5 and –0.5 for 1st and 2nd equation

Table 1. Variances of BLUP, GLS prediction with autocorrelated errors

Correlation Between X's	Error covariance between equations (σ_{12})	No. of Observations (T)	Prediction variances × 100			
			BLUP for 1st Eq.	BLUP for 2nd Eq.	GLS for 1st Eq.	GLS for 2nd Eq.
			$(\sigma^2_{p_1})$	$(\sigma^2_{p_2})$	$(\sigma^2_{\tilde{y}_{p_1}})$	$(\sigma^2_{\tilde{y}_{p_2}})$
Low	0.9	15	89.9	77.5	106.0	104.2
		20	89.2	71.7	105.8	99.7
		30	86.5	70.1	103.8	99.5
		50	84.3	68.8	102.3	96.2
	0.5	15	96.4	90.6	112.0	113.5
		20	94.2	73.2	111.5	107.9
		30	92.4	71.1	108.9	101.5
		50	87.8	69.4	105.3	98.6
	0.3	15	97.9	93.8	113.4	114.3
		20	95.4	73.6	112.4	110.9
		30	92.8	71.3	110.1	108.5
		50	88.6	69.5	106.0	99.1
High	0.9	15	97.9	86.8	113.3	113.1
		20	97.5	84.7	113.1	112.7
		30	92.0	78.7	108.7	108.3
		50	87.5	73.9	105.2	104.9
	0.5	15	98.5	88.8	114.0	115.1
		20	96.0	86.1	112.0	113.9
		30	94.2	81.6	110.5	111.1
		50	88.9	75.6	106.2	106.6
	0.3	15	98.6	88.9	114.0	115.3
		20	96.4	86.5	112.3	113.3
		30	94.5	81.9	110.7	111.4
		50	89.0	75.8	106.4	106.8

respectively, in (5.1). The result is also similar to that reported by Zellner and Huang (1962) in case of estimation of a SURE model that with low correlation between X's across equations, the estimated variances of BLUP is smaller.

REFERENCES

Chib, S. and Greenberg, E. (1995). Hierarchical analysis of SURE models with extensions to correlated serial errors and time-varying parameter models. *J. Eco.*, **68**, 339- 360.

Dao de, G. and Rong, B.H. (2000). Some finite sample properties of Zellner estimator in the context of m seemingly unrelated regression equations. *J. Statist. Plann. Inf.*, **88**, 267-283.

Kmenta, J. and Gilbert, R.F. (1968). Small sample properties of alternative estimators of seemingly unrelated regression. *J. Amer. Statist. Assoc.,* **63**, 1180-1200.

Kmenta, J. and Gilbert, R.F. (1970). Estimation of seemingly unrelated regression with autoregressive disturbance. *J. Amer. Statist. Assoc.*, **65**, 186-196.

Percy, D.F. (1992). Prediction for seemingly unrelated regressions. *J. Roy. Statist. Soc.*, **B54**, 243-252.

Richard, J.F. and Steel, M.J.F. (1988). Bayesian analysis of system of SURE under a recursive extended natural conjugate prior density. *J. Eco.*, **38**, 7-37.

Sharma, V.K. (1993). Estimation of seemingly unrelated regressions with unequal numbers of observations. *Sankhya*, **B55**, 135-138.

Srivastava, V.K and Giles, D.E.A. (1987). *Seemingly Unrelated Regression Equation Models-Estimation and Inference.* Marcel Dekker, New York.

Srivastava, V.K. and Wam, A.T.K. (2002). Separate versus system methods of Stein-Rule estimation in seemingly unrelated regression models. *Comm. Statist.- Theory Methods*, **31**, 2007-2099.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *J. Amer. Statist. Assoc.*, **57**, 348-368.

Zellner, A. and Huang, D.S. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. *Intt. Rev.*, **3**, 300-313.

APPENDIX

Current area (thousand hectare), lagged* yield (quintal per hectare) and lagged* prices (rupees per quintal) of rapeseed and wheat in Haryana State

Year	Current area		Lagged yield		Lagged price	
	Rapeseed	Wheat	Rapeseed	Wheat	Rapeseed	Wheat
1983-84	307	1789	8.26	26.62	318.74	135.20
1984-85	328	1705	9.87	24.89	340.86	151.55
1985-86	345	1699	10.76	25.56	368.61	150.59
1986-87	342	1782	9.99	26.46	393.59	163.97
1987-88	276	1731	8.49	31.15	546.17	163.22
1988-89	328	1827	8.68	28.61	802.11	167.43
1989-90	383	1859	13.36	28.69	528.23	180.47
1990-91	429	1850	11.01	33.79	664.43	185.31
1991-92	474	1808	13.54	33.25	814.71	214.62
1992-93	638	1956	11.83	35.99	888.18	236.48
1993-94	567	1998	10.94	36.95	848.86	291.55
1994-95	579	1986	11.88	38.29	953.99	365.51
1995-96	580	2003	12.81	38.56	1043.77	359.35
1996-97	617	2020	12.90	38.85	959.76	382.65
1997-98	610	2060	15.54	40.80	1094.58	477.40
1998-99	625	2084	16.83	36.57	1174.39	517.62

* lagged by one year