The Correlation of the Residuals in a Multivariate Regression Model*

Takeshi Amemiya

1. Introduction: Problem Stated

Zellner [5] showed that the efficiency of estimation of the parameters of a multivariate regression may be asymptotically improved if Aitken's generalized least squares method is applied to a whole set of equations instead of estimating each equation by least squares. Zellner's procedure is as follows:

Let the regression system be, following Zellner's notation,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots\cdots & 0 \\ 0 & X_2 & \cdots\cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots\cdots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix} \tag{1}$$

where y_i is a T-component vector of the i-th dependent variable, X_i a $T \times K_i$ matrix with rank K_i of observations on K_i nonstochastic variables, β_i a K_i-component vector of regression coefficients and u_i a T-component vector of random error terms, each with mean zero. Zellner assumes u_i is a vector of independent and stationary random variables but within the same period u_i and u_j are correlated.

Write (1) simply as

$$y = X\beta + u \tag{2}$$

Then

$$Euu'(=\Sigma) = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots\cdots & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I & \cdots\cdots & \sigma_{2M}I \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}I & \sigma_{M2}I & \cdots\cdots & \sigma_{MM}I \end{bmatrix} \tag{3}$$

where $\sigma_{ij} = \frac{1}{T} Eu'_i u_j$. Zellner's estimator β^* of β is obtained by first obtaining the least squares estimates $\hat{\beta}$, secondly calculating the residuals by $\hat{u} = y - X\hat{\beta}$, thirdly estimating σ_{ij} by $\hat{\sigma}_{ij} = \frac{1}{T} \hat{u}_i' \hat{u}_j$, and finally calculating

$$\beta^* = (X'\hat{\Sigma}^{-1}X)^{-1}(X'\hat{\Sigma}^{-1}y) \tag{4}$$

where $\hat{\Sigma}$ is obtained by replacing σ_{ij} by $\hat{\sigma}_{ij}$ in Σ.

Zellner obtains the asymptotic distribution of β^* and shows that it is asymptotically more efficient than $\hat{\beta}$.

The question we want to pose in this paper is, What happens to the superiority of Zellner's estimator over the simple least squares estimator if the elements of u_i are correlated? In that case the covariance matrix of u cannot be expressed as $M \times M$ blocks of diagonal matrices as in (3) and the elements of Σ specified to be zero in (3) become non-zero. Thus, both the least squares and Zellner's method assume wrong covariance matrices. Is Zellner's method still better than the least squares method in this case, because the covariance matrix on which Zellner's method is based is "closer" to the true one than that of the least squares method? Or, is the least squares better, for it may be sometimes better to be wholly wrong than to be only half wrong?

To answer the above question we consider for simplicity the case $M=2$ and where the autocorrelation of u is a bivariate first-order autoregression. That is,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{5}$$

(or simply $y = X\beta + u$)

and

$$\begin{aligned} u_{1t} &= r_{11}u_{1,t-1} + r_{12}u_{2,t-1} + e_{1t} \\ u_{2t} &= r_{21}u_{1,t-1} + r_{22}u_{2,t-1} + e_{2t} \end{aligned} \tag{6}$$

where $\{e_{1t}\}$ and $\{e_{2t}\}$ are each independent and stationary with mean zero and finite covariance matrix and mutually independent. Assume also that the

* The research on which this paper was based was carried out in part with the support of Grant GS-986 from the National Science Foundation to Stanford University and in part with that of a grant from Ministry of Education of the Japanese Government to Joint Research Project on Quantitative Economics at the Institute of Economic Research, Hitotsubashi University. The author is grateful to the statistical staff of the Institute for computational help.

roots of determinantal equation $\begin{vmatrix} r_{11}-\lambda & r_{12} \\ r_{21} & r_{22}-\lambda \end{vmatrix}=0$ are less than unity in modulus.

We want to compare the asymptotic efficiency of the following four estimators of the parameters β's of model (5)-(6).

(I) *The least squares method*: This is simply $(X'X)^{-1}X'y$.

(II) *Zellner's method*: This has been defined above.

(III) *The least squares method after a quasi first-difference transformation*: Assume each of u_1 and u_2 follows a first-order autoregression but is independent of the other. Estimate the parameters of the autoregression consistently from the calculated residuals. Transform the variables using these estimates and apply the least squares method to the transformed regression equations.

(IV) *Zellner's method after a quasi first-difference transformation*: After transformation described in (III), apply Zellner's method.

Each of the above four estimates can be written in the form $(X'\hat{C}_i^{-1}X)^{-1}(X'\hat{C}_i^{-1}y)$, $i=1, 2, 3, 4$, for some estimate of the true covariance matrix Σ. If the probability limit of $\hat{C}_i$ is C_i,[1] the asymptotic covariance matrix of the i-th estimates is $(X'C_i^{-1}X)^{-1}X'C_i^{-1}\Sigma C_i^{-1}X(X'C_i^{-1}X)^{-1}$ which reaches its minimum (in matrix sense)

$$(X'\Sigma^{-1}X)^{-1}$$

when $C_i=\Sigma$. Hence we may define the asymptotic efficiency of the i-th estimates by

$$\mathrm{Eff}(i)=\frac{|X'C_i^{-1}X|^2}{|X'C_i^{-1}\Sigma C_i^{-1}X||X\Sigma^{-1}X|} \qquad (7)$$

But this itself is not a good criterion for comparison because it depends on X. A natural procedure then is to consider a lower bound of (7) as X varies within a certain class. We will use one such lower bound proposed by Watson [3], namely,

$$\mathrm{LBE}(i)=\frac{4\bar{\lambda}_i\cdot\underline{\lambda}_i}{(\bar{\lambda}_i+\underline{\lambda}_i)^2} \qquad (8)$$

where $\bar{\lambda}_i, \underline{\lambda}_i$ are the largest and the smallest characteristic root of $C_i^{-1}\Sigma$ respectively. Before obtaining the formulas for λ_i as functions of the parameters in model (5)-(6), we will, in next section, briefly discuss Watson's aforementioned work.

2. The Lower Bound of Efficiency

Consider the regression model $y=X\beta+u$ where y is a T-component vector of dependent variables, X a $T\times K$ matrix of constants, β a K-component vector of parameters to be estimated, and u a T-component vector of stationary random variables with mean zero. If one uses Aitken's least squares estimator assuming the covariance matrix of u is A when it really is B, the efficiency of the estimator may be defined by

$$\mathrm{Eff}=\frac{|X'A^{-1}X|^2}{|X'A^{-1}BA^{-1}X|\cdot|X'B^{-1}X|} \qquad (9)$$

or

$$\mathrm{Eff}=\frac{|Z'Z|^2}{|Z'\Lambda Z|\cdot|Z'\Lambda^{-1}Z|} \qquad (10)$$

where $Z=A^{-1/2}X$ and $\Lambda=A^{-1/2}BA^{-1/2}$.

Let $z_1, z_2, \cdots\cdots, z_K$ be the column vectors of Z. Suppose that $z_{K-h+1}, \cdots\cdots, z_K$ are characteristic vectors of Λ associated with the characteristic roots $\lambda_{T-h+1}, \cdots\cdots, \lambda_T$. Without loss of generality we may assume that $\lambda_1<\lambda_2<\cdots\cdots<\lambda_{T-h}$. Then, according to theorem of Watson [3],

$$\mathrm{Eff}\geqq\frac{4\lambda_1\cdot\lambda_{T-h}}{(\lambda_1+\lambda_{T-h})^2}\cdot\frac{4\lambda_2\lambda_{T-h-1}}{(\lambda_2+\lambda_{T-h-1})^2}\cdots\cdots\frac{4\lambda_{K-h}\lambda_{T-K+1}}{(\lambda_{K-h}+\lambda_{T-K+1})^2} \qquad (11)$$

In the special case where all the columns of Z except one are characteristic vectors of Λ associated with the characteristic roots of other than the largest and smallest roots, the lower bound of efficiency is given by

$$\mathrm{LBE}=\frac{4\bar{\lambda}\cdot\underline{\lambda}}{(\bar{\lambda}+\underline{\lambda})^2} \qquad (12)$$

where $\bar{\lambda}$ and $\underline{\lambda}$ are the largest and the smallest characteristic roots of Λ respectively. We will use this formula in this paper, just as Watson and Hannan [4] do in theirs. The use of the simpler formula (12) rather than the more general (11) may be justified by observing that both (11) and (12) measure how widely scattered the characteristic roots of Λ are and hence one behaves somewhat similarly to the other.[2]

Consideration of a lower bound is in the spirit of minmax and may be justified for itself. But there is a possibility that estimator (a) is better than

1) In section 3 each $\hat{C}_i$ is defined in detail. Because of our assumptions stated after equation (6), the convergence of $\hat{C}_i$ to C_i in probability can be easily proved by means of Theorem 3A of Diananda [1].

estimator (b) if judged by the lower bound of efficiency whereas the latter is more efficient for most probable values of X. The smaller is the likelihood of such a possibility, the better measure the lower bound of efficiency is. Unfortunately we have not been able to justify Watson's lower bound fully in this respect: hence, we must use it with caution.

3. Derivation of Characteristic Roots

In this section we show how to obtain the characteristic roots required to calculate the lower bound of efficiency (8) for each of the four estimators we proposed to compare in section 1.

(I) *The least squares method*

Put $\sigma_{12}=0$ in formula (16) below. Thus, the characteristic roots of $C_1^{-1}\Sigma$ are, approximately,

$$\lambda_1(\omega)=\frac{1}{2}\left\{\frac{f_{11}(\omega)}{\sigma_{11}}+\frac{f_{22}(\omega)}{\sigma_{22}}\right.$$
$$\left.\pm\sqrt{\left(\frac{f_{11}(\omega)}{\sigma_{11}}+\frac{f_{22}(\omega)}{\sigma_{22}}\right)^2-\frac{4(f_{11}(\omega)f_{22}(\omega)-|f_{12}(\omega)|^2)}{\sigma_{11}\sigma_{22}}}\right\}$$
$$\omega=\frac{\pi}{T}j,\qquad j=0,1,\cdots\cdots,T-1 \qquad (13)$$

For the meaning of notation, see discussion of (II) below.

(II) *Zellner's method*

Partition Σ naturally as

$$\Sigma=E\binom{u_1}{u_2}(u_1',u_2')=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$$

We want to obtain the characteristic roots of

$$C_2^{-1}\Sigma=\begin{bmatrix}\sigma_{11}I & \sigma_{12}I\\ \sigma_{12}I & \sigma_{22}I\end{bmatrix}^{-1}\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$$
$$\text{say}=\begin{bmatrix}a\cdot I & c\cdot I\\ c\cdot I & b\cdot I\end{bmatrix}\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$$

where σ_{11},σ_{22}, and σ_{12} are the diagonal elements of Σ_{11},Σ_{22}, and Σ_{12} respectively.

Let $U=\frac{1}{\sqrt{T}}\left\{e^{\frac{2\pi ihk}{T}}\right\}$; $h,k=0,1,\cdots,T-1$

Then, we have asymptotically,[3]

$$\begin{bmatrix}U^* & 0\\ 0 & U^*\end{bmatrix}\begin{bmatrix}a\cdot I & c\cdot I\\ c\cdot I & b\cdot I\end{bmatrix}\begin{bmatrix}U & 0\\ 0 & U\end{bmatrix}\begin{bmatrix}U^* & 0\\ 0 & U^*\end{bmatrix}\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\begin{bmatrix}U & 0\\ 0 & U\end{bmatrix}$$
$$=\begin{bmatrix}a\cdot I & c\cdot I\\ c\cdot I & b\cdot I\end{bmatrix}\begin{bmatrix}F_{11} & F_{12}\\ F_{21} & F_{22}\end{bmatrix}$$
$$=\begin{bmatrix}a\cdot F_{11}+c\cdot F_{21} & a\cdot F_{12}+c\cdot F_{22}\\ c\cdot F_{11}+b\cdot F_{21} & c\cdot F_{12}+b\cdot F_{22}\end{bmatrix} \qquad (14)$$

where F_{11}, F_{22}, F_{12}, and F_{21} are diagonal matrices with the diagonal elements $f_{11}(\omega)$, $f_{22}(\omega)$, $f_{12}(\omega)$, and $f_{21}(\omega)$ respectively for $\omega=\frac{\pi}{T}j$, $j=0,1,\cdots\cdots,T-1$. $f_{11}(\omega)$ and, $f_{22}(\omega)$ are the spectral densities of u_1, and u_2 respectively, and $f_{12}(\omega)$ and $f_{21}(\omega)$ are the cross spectral densities between u_1, and u_2. They are defined by

$$\left.\begin{aligned}f_{11}(\omega)&=\sum_{n=-\infty}^{+\infty}(Eu_{1,t+n}u_{1,t})e^{-in\omega}\\ f_{22}(\omega)&=\sum_{-\infty}^{\infty}(Eu_{2,t+n}u_{2,t})e^{-in\omega}\\ f_{12}(\omega)&=\sum_{-\infty}^{\infty}(Eu_{1,1+n}u_{2,t})e^{-in\omega}\\ f_{21}(\omega)&=\sum_{-\infty}^{\infty}(Eu_{2,t+n}u_{1,t})e^{-in\omega}\end{aligned}\right\} \qquad (15)$$

But the first term of (14) has the same characteristic roots as $C_2^{-1}\Sigma$, for $UU^*=U^*U=I$. Calculating the characteristic roots of the last term of (14) directly, we have

$$\lambda_2(\omega)=\frac{1}{2(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}\cdot\{\sigma_{22}\cdot f_{11}(\omega)-2\sigma_{12}\cdot Ref_{12}(\omega)$$
$$+\sigma_{11}\cdot f_{22}(\omega)$$
$$\pm\sqrt{(\sigma_{22}f_{11}(\omega)-2\sigma_{12}\cdot Ref_{12}(\omega)+\sigma_{11}f_{22}(\omega))^2-4(\sigma_{11}\sigma_{22}-\sigma_{12}^2)(f_{11}(\omega)f_{22}(\omega)-|f_{12}(\omega)|^2)}\}$$
$$\omega=\frac{\pi}{T}j,\quad j=0,1,\cdots\cdots,T-1. \qquad (16)$$

Because of our specification (6), f's are related to r's as follows.

$$\begin{bmatrix}f_{11}(\omega) & f_{12}(\omega)\\ f_{21}(\omega) & f_{22}(\omega)\end{bmatrix}=(I-Re^{i\omega})^{-1}(I-R'e^{-i\omega})^{-1} \qquad (17)$$

where

$$R=\begin{pmatrix}r_{11} & r_{12}\\ r_{21} & r_{22}\end{pmatrix}.$$

Therefore we have

$$\left.\begin{aligned}f_{11}(\omega)&=D^{-1}\cdot(1+r_{22}^2-2r_{22}\cos\omega+r_{12}^2)\\ f_{22}(\omega)&=D^{-1}\cdot(1+r_{11}^2-2r_{11}\cos\omega+r_{21}^2)\\ Ref_{12}(\omega)&=D^{-1}\cdot[(r_{12}+r_{21})\cos\omega\\ &\qquad-(r_{11}r_{12}+r_{22}r_{21})]\end{aligned}\right\} \qquad (18)$$

2) Professor Takeuchi pointed out to me that LBE (12) is equal to

$$\min_X\min_v\frac{v'(X'B^{-1}X)^{-1}v}{v'(X'A^{-1}X)^{-1}X'A^{-1}BA^{-1}X(X'A^{-1}X)^{-1}v}$$

3) For the asymptotic diagonalization of a covariance matrix, see Grenander and Rosenblatt [2, p. 103 f].

$$\left.\begin{aligned}|f_{12}(\omega)|^2=D^{-2}\cdot[4r_{12}r_{21}\cos^2\omega-2(r_{12}+r_{21})\\ (r_{11}r_{12}+r_{22}r_{21})\cos\omega+(r_{11}r_{12}\\ +r_{22}r_{21})^2+(r_{12}-r_{21})^2]\end{aligned}\right\}$$

where $D=(1+r_{11}^2)(1+r_{22}^2)+r_{12}r_{21}(r_{12}r_{21}-2r_{11}r_{22}+2)-2(r_{11}+r_{22})(r_{11}r_{22}-r_{12}r_{21}+1)\cos\omega+4(r_{11}r_{22}-r_{12}r_{21})\cos^2\omega.$

σ's are approximately related to f's by

$$\left.\begin{aligned}\sigma_{11}&=\frac{1}{T}\sum_{j=0}^{T-1}f_{11}\left(\frac{\pi}{T}j\right)\\ \sigma_{22}&=\frac{1}{T}\sum_{j=0}^{T-1}f_{22}\left(\frac{\pi}{T}j\right)\\ \sigma_{12}&=\frac{1}{T}\sum_{j=0}^{T-1}Ref_{12}\left(\frac{\pi}{T}j\right)\end{aligned}\right\}\tag{19}$$

(III) *The least squares method after a quasi first-difference transformation*

This method means premultiplying equation (5) by the transformation operator H to be defined below and then applying the least squares method to regression equation

$$Hy=HX\beta+Hu \tag{20}$$

where

$$H=\begin{pmatrix}H_1 & 0\\ 0 & H_2\end{pmatrix}$$

$$H_i=\begin{bmatrix}1 & 0 & \cdots & \cdots & 0\\ -\rho_i & 1 & 0 & & \vdots\\ 0 & -\rho_i & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \ddots & \vdots\\ 0 & & & -\rho_i & 1\end{bmatrix},\quad i=1,2.$$

and

$$\left.\begin{aligned}\rho_1&=p\lim\frac{\hat{u}_1'\hat{u}_{1,-1}}{\hat{u}'_{1,-1}\hat{u}_{1,-1}}=r_{11}+r_{12}\cdot\frac{\sigma_{12}}{\sigma_{11}}\\ \rho_2&=p\lim\frac{\hat{u}_2'\hat{u}_{2,-1}}{\hat{u}_2{}',_{-1}\hat{u}_{2,-1}}=r_{22}+r_{21}\cdot\frac{\sigma_{12}}{\sigma_{22}}\end{aligned}\right\}\tag{21}$$

where $\hat{u}$'s are the calculated residuals.

Thus we have

$$C_3=\begin{bmatrix}Q_1 & 0\\ 0 & Q_2\end{bmatrix}$$

where

$$Q_i=(H_i'H_i)^{-1}=\frac{1}{1-\rho_i^2}\begin{bmatrix}1 & \rho_i & \cdots & \rho_i^{T-1}\\ \rho_i & 1 & & \\ \vdots & & \ddots & \\ \rho_i^{T-1} & & & 1\end{bmatrix},\quad i=1,2.$$

The characteristic roots of $C_3^{-1}\Sigma$ are the same as those of

$$\begin{bmatrix}U^* & 0\\ 0 & U^*\end{bmatrix}\begin{bmatrix}Q_1 & 0\\ 0 & Q_2\end{bmatrix}^{-1}\begin{bmatrix}U & 0\\ 0 & U\end{bmatrix}\begin{bmatrix}U^* & 0\\ 0 & U^*\end{bmatrix}\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\begin{bmatrix}U & 0\\ 0 & U\end{bmatrix}$$

$$=\begin{bmatrix}G_1 & 0\\ 0 & G_2\end{bmatrix}\begin{bmatrix}F_{11} & F_{12}\\ F_{21} & F_{22}\end{bmatrix}=\begin{bmatrix}G_1F_{11} & G_1F_{12}\\ G_2F_{21} & G_2F_{22}\end{bmatrix}\tag{22}$$

where G_i is a diagonal matrix with the diagonal elements $1-2\rho_i\cos\left(\frac{\pi}{T}j\right)+\rho_i^2$, $j=0,1,\cdots\cdots,T-1$ for $i=1,2$.

Evaluating the characteristic roots of the last term of (22), we have

$$\lambda_3(\omega)=\frac{1}{2}\{g_1(\omega)f_{11}(\omega)+g_2(\omega)f_{22}(\omega)\pm\sqrt{[g_1(\omega)f_{11}(\omega)+g_2(\omega)f_{22}(\omega)]^2-4g_1(\omega)g_2(\omega)[f_{11}(\omega)f_{22}(\omega)-|f_{12}(\omega)|^2]}\}$$

$$\omega=\frac{\pi}{T}j,\ j=0,1,\cdots\cdots,T-1\tag{23}$$

where $g_i(\omega)=1-2\rho_i\cos\omega+\rho_i^2$, $i=1,2$.

(IV) *Zellner's method after a quasi first-difference transformation*

To evaluate C_4, we must first evaluate the probability limit of the variance and covariance of the calculated residuals of Hu_1 and Hu_2. They are

$$\left.\begin{aligned}v_{11}&=p\lim\frac{1}{T}(\hat{u}_1-\rho_1\hat{u}_{1,-1})'(\hat{u}_1-\rho_1\hat{u}_{1,-1})\\ &=1+r_{12}^2\left(\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}\right)\\ v_{22}&=p\lim\frac{1}{T}(\hat{u}_2-\rho_2\hat{u}_{2,-1})'(\hat{u}_2-\rho_2\hat{u}_{2,-1})\\ &=1+r_{21}^2\left(\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}\right)\\ v_{12}&=p\lim\frac{1}{T}(\hat{u}_1-\rho_1\hat{u}_{1,-1})'(\hat{u}_2-\rho_2\hat{u}_{2,-1})\\ &=\sigma_{12}r_{12}r_{21}\left(\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}-1\right)\end{aligned}\right\}\tag{24}$$

This method means premultiplying equation (20) by the operator J to be defined below and then applying the least squares method to

$$JHy=JHX\beta+JHu \tag{25}$$

where J is the matrix such that

$$J'J=\begin{bmatrix}v_{11}\cdot I & v_{12}\cdot I\\ v_{12}\cdot I & v_{22}\cdot I\end{bmatrix}^{-1}$$

Thus $C_4^{-1}\Sigma=H'J'JH\Sigma$, the characteristic roots of which we now set out to evaluate. They are the same as the characteristic roots of

$$W^*H'WW^*J'JWW^*HWW^*\Sigma W$$

where $W=\begin{bmatrix}U & 0\\ 0 & U\end{bmatrix}$.

But we have

$$W^*H'W=\begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

where $P_i(i=1,2)$ is the diagonal matrix with the diagonal elements

$$1-\rho_i e^{\frac{2\pi i}{T}j}, \qquad j=0,1,\cdots\cdots,T-1.$$

and W^*HW is the complex conjugate of the above. Thus the lengthy matrix product above can be simplified as

$$\frac{1}{v_{11}v_{22}-v_{12}^2}\cdot\begin{bmatrix} L_{11}F_{11}+L_{12}F_{21} & L_{11}F_{12}+L_{12}F_{22} \\ L_{21}F_{11}+L_{22}F_{21} & L_{21}F_{12}+L_{22}F_{22} \end{bmatrix}$$

where L_{11} is a diagonal matrix with the diagonal elements $v_{22}\cdot|1-\rho_1 e^{\frac{2\pi i}{T}j}|^2$, $j=0,1,\cdots\cdots,T-1$, L_{22} with $v_{11}\cdot|1-\rho_2 e^{\frac{2\pi i}{T}j}|^2$, L_{12} with $-v_{12}\cdot(1-\rho_1 e^{\frac{2\pi i}{T}j})(1-\rho_2 e^{-\frac{2\pi i}{T}j})$, and L_{21} with the complex conjugate of the elements of L_{12}. F's are defined after equation (14).

Evaluating the characteristic roots of above directly, we have

$$\lambda_4(\omega)=\frac{1}{2(v_{11}v_{22}-v_{12}^2)}\cdot\{S\pm\sqrt{S^2-4(|f_{12}(\omega)|^2-f_{11}(\omega)f_{22}(\omega))(v_{12}^2-v_{11}v_{22})(1+\rho_1^2-2\rho_1\cos\omega)(1+\rho_2^2-2\rho_2\cos\omega)}\}$$

$$\omega=\frac{\pi}{T}j,\ j=0,1,\cdots\cdots,T-1 \qquad (26)$$

where $S=v_{22}(1-2\rho_1\cos\omega+\rho_1^2)f_{11}(\omega)+v_{11}(1-2\rho_2\cos\omega+\rho_2^2)f_{22}(\omega)+\frac{2v_{12}}{D}\{2(\rho_1 r_{21}+\rho_2 r_{12})\cos^2\omega-[(r_{21}+r_{12})(1+\rho_1\rho_2)+(\rho_1+\rho_2)(r_{11}r_{12}+r_{22}r_{21})]\cos\omega+(1+\rho_1\rho_2)(r_{11}r_{12}+r_{22}r_{21})+(\rho_2-\rho_1)(r_{21}-r_{12})\}$.

4. Computation

In section 3 we have obtained the formulas for the characteristic roots required to calculate the lower bound of efficiency (8) for each of the four estimators to be compared. From (13), (16), (17), (19), (21), (23), (24), and (26), the characteristic roots $\lambda_i(\omega)$, $i=1,2,3,4$, can be expressed as functions of the four parameters r_{11}, r_{12}, r_{21}, and r_{22} as well as T.

In computing the characteristic roots we fix T to be 20. That means, there are 40 roots to compute for each estimation method, from which the largest and smallest are to be chosen. For the values of r's, the following 108 combinations are chosen:

r_{11}: 0.2, 0.5, 0.7

r_{22}: 0.2, 0.5, 0.7

r_{21}: 0.2, 0.5

r_{12}: −0.7, −0.5, −0.2, 0.2, 0.5, 0.7

For r_{11} and r_{22}, only positive values are chosen for simplicity, for positive autoregressive coefficients are more likely in economic data. Once we assume r_{11} and r_{22} to be positive, there is no loss of generality in assuming r_{21} to be positive, for fixing r_{11} and r_{22} and changing signs of r_{12} and r_{21} simultaneously keep the values of $\lambda_i(\omega)$ invariant.

If the roots of dererminantal equation

$$\begin{vmatrix} r_{11}-\lambda & r_{12} \\ r_{21} & r_{22}-\lambda \end{vmatrix}=0$$

exceed unity in modulus, equations (6) become unstable. There are 15 combinations in which unstability occurs, and computation is made for the remaining 93 combinations. The results are shown in the appended table.[4)]

In the table, the values of $\sqrt{\rho_1\rho_2}$ and $\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ and their ratio are also shown. The former is the geometric mean of the correlation coefficients between $u_{1,t}$ and $u_{1,t-1}$ and between $u_{2,t}$ and $u_{2,t-1}$. The latter is the correlation coefficient between u_{1t} and u_{2t}. ρ_1 and ρ_2 are always positive in our cases.

5. Conclusion

It is clearly shown in the table that the lower bound of efficiency of estimator (I) is almost identical with that of estimator (II), that of (III) is almost identical with that of (IV), and that the latter two are considerably higher than the former two in every combination considered. In other words, when the residuals follow process (6), application of a quasi first-difference transformation increases the lower bound of efficiency considerably, whereas application of Zellner's method leaves it almost unaltered. Before extracting any general conclusion from this result, we must answer a few doubts that might be cast about this analysis.

Firstly, the use of the particular lower bound of efficiency may be questioned. We have already men-

4) Computation was carried out on Burroughs E101 in the computation office of the Institute of Economic Research, Histotsubashi University.

Table: Lower Bound of Efficiency of the Four Estimators and Correlation Coefficients (Values in Percentage)[5]

	r_{11}	r_{22}	r_{21}	r_{12}	$\frac{4\bar{\lambda}_1\underline{\lambda}_1}{(\bar{\lambda}_1+\underline{\lambda}_1)^2}$	$\frac{4\bar{\lambda}_2\underline{\lambda}_2}{(\bar{\lambda}_2+\underline{\lambda}_2)^2}$	$\frac{4\bar{\lambda}_3\underline{\lambda}_3}{(\bar{\lambda}_3+\underline{\lambda}_3)^2}$	$\frac{4\bar{\lambda}_4\underline{\lambda}_4}{(\bar{\lambda}_4+\underline{\lambda}_4)^2}$	$\sqrt{\rho_1\rho_2}$	$\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$	$\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\Big/\sqrt{\rho_1\rho_2}$
1	0.2	0.2	0.2	−0.7	39	37	43	45	21	−8	−36
2				−0.5	52	51	59	60	21	−5	−27
3				−0.2	73	73	84	84	20	0	0
4				0.2	53	53	78	78	22	11	50
5				0.5	30	30	48	47	27	20	74
6				0.7	20	20	32	31	30	25	84
7			0.5	−0.7	19	19	20	20	20	−2	−9
8				−0.5	31	31	33	33	20	0	0
9				−0.2	52	51	59	60	21	5	27
10				0.2	30	30	48	47	27	20	74
11				0.5	12	12	19	16	38	35	94
12				0.7	5	5	8	6	49	49	100
13		0.5	0.2	−0.7	26	22	31	34	40	−26	−66
14				−0.5	32	30	50	51	37	−21	−58
15				−0.2	39	39	82	82	33	−8	−25
16				0.2	22	22	65	65	36	20	56
17				0.5	10	11	26	24	49	42	85
18				0.7	6	6	13	11	57	53	93
19			0.5	−0.7	13	13	15	16	36	−18	−50
20				−0.5	21	21	28	29	35	−15	−43
21				−0.2	30	30	56	56	32	−3	−10
22				0.2	10	11	35	32	42	31	75
23				0.5	2	2	5	2	69	68	98
24				0.7	0	0	0	0	87	85	98
25		0.7	0.2	−0.7	14	13	18	22	51	−33	−65
26				−0.5	15	16	35	37	50	−36	−73
27				−0.2	14	15	78	78	41	−18	−43
28				0.2	6	7	44	43	47	32	69
29				0.5	2	2	7	7	70	67	95
30				0.7	0	0	0	0	78	74	95
31			0.5	−0.7	9	9	11	13	47	−31	−66
32				−0.5	13	13	23	25	45	−28	−61
33				−0.2	15	16	52	53	39	−12	−31
34				0.2	2	2	15	13	56	48	86
35	0.5	0.2	0.2	−0.7	22	22	40	41	32	1	4
36				−0.5	30	30	56	56	32	3	10
37				−0.2	39	39	82	82	33	8	25
38				0.2	22	22	66	65	36	20	56
39				0.5	10	11	35	32	42	31	75
40				0.7	6	6	21	17	47	39	84
41			0.5	−0.7	13	13	16	16	34	13	38
42				−0.5	21	21	28	29	35	15	43
43				−0.2	32	30	50	51	37	21	58
44				0.2	10	11	26	24	49	42	85
45				0.5	2	2	5	3	69	68	98
46				0.7	0	0	0	0	85	83	98
47		0.5	0.2	−0.7	13	12	25	28	51	−10	−19
48				−0.5	19	17	43	43	51	−13	−25
49				−0.2	31	31	76	76	50	0	0
50				0.2	12	12	52	51	57	33	58
51				0.5	3	3	16	13	69	58	84
52				0.7	0	0	0	0	74	63	85
53			0.5	−0.7	7	6	9	9	50	−3	−6
54				−0.5	11	11	19	19	50	0	0
55				−0.2	19	17	44	43	51	13	25
56				0.2	3	3	16	13	69	58	84
57		0.7	0.2	−0.7	8	6	13	15	64	−32	−50
58				−0.5	11	9	26	28	63	−29	−46
59				−0.2	14	14	68	69	60	−12	−19

	r_{11}	r_{22}	r_{21}	r_{12}	$\frac{4\bar{\lambda}_1\lambda_1}{(\bar{\lambda}_1+\lambda_1)^2}$	$\frac{4\bar{\lambda}_2\lambda_2}{(\bar{\lambda}_2+\lambda_2)^2}$	$\frac{4\bar{\lambda}_3\lambda_3}{(\bar{\lambda}_3+\lambda_3)^2}$	$\frac{4\bar{\lambda}_4\lambda_4}{(\bar{\lambda}_4+\lambda_4)^2}$	$\sqrt{\rho_1\rho_2}$	$\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$	$\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} / \sqrt{\rho_1\rho_2}$
60	0.5	0.7	0.2	0.2	3	4	32	30	70	48	70
61				0.5	0	0	0	0	86	78	90
62				0.7	0	0	0	0	97	96	98
63			0.5	−0.7	0	0	0	0	61	−16	−27
64				−0.5	6	6	11	12	60	−14	−24
65				−0.2	9	9	35	36	59	1	1
66				0.2	0	0	0	0	82	70	86
67	0.7	0.2	0.2	−0.7	12	13	36	39	39	11	28
68				−0.5	15	16	52	53	39	12	31
69				−0.2	14	15	78	78	41	18	43
70				0.2	6	7	44	43	47	32	69
71				0.5	2	2	15	13	56	48	86
72				0.7	0	0	0	0	59	50	84
73			0.5	−0.7	8	9	13	14	44	25	57
74				−0.5	13	13	23	25	45	28	61
75				−0.2	15	16	35	37	50	36	73
76				0.2	2	2	7	7	70	67	95
77		0.5	0.2	−0.7	0	0	0	0	59	−3	−6
78				−0.5	0	0	0	0	59	−1	−1
79				−0.2	14	14	68	68	60	12	19
80				0.2	3	4	32	30	70	48	70
81				0.5	0	0	0	0	82	70	86
82				0.7	0	0	0	0	95	93	97
83			0.5	−0.7	3	3	4	4	61	11	18
84				−0.5	6	6	11	12	60	14	24
85				−0.2	11	9	26	28	63	29	46
86				0.2	0	0	0	0	86	78	90
87		0.7	0.2	−0.7	3	3	8	10	71	−17	−24
88				−0.5	5	4	20	21	71	−16	−22
89				−0.2	10	10	56	56	70	0	0
90				0.2	0	0	0	0	82	60	73
91			0.5	−0.7	0	0	0	0	70	−2	−2
92				−0.5	0	0	0	0	70	0	0
93				−0.2	5	5	20	21	71	16	22
AVERAGE					13.3	13.2	28.8	28.8			
Average of cases where none of the parameters is ±0.7					24.6	24.4	44.8	44.0			

tioned this point in section 2.

Secondly it may be thought that the result is peculiar to the first-order autoregressive model assumed in our study. But we think that this model does contain some essential characteristics common with a more general model and we can extend our conclusion to a more general model to a certain extent. It would be reasonable to guess that Zellner's method is not helpful in a more general model whenever the residuals are auto-correlated to some extent and that a quasi first-order transformation is helpful to the extent the residuals are dependent on their first lags. Regarding this former point, it should be noted that after the transformation the residuals no longer follow the first-order autoregression as in (6) and yet Zellner's method represented by (IV) doesn't do any better than (III).

Thirdly, it may be thought that the superiority of (III) and (IV) over (I) and (II) and the similarity of (I) with (II) and (III) with (IV) in the lower bound of efficiency are due to the fact that in the combinations of the parameters considered in this analysis the dependence of a residual on its own

5) For calculating σ_{12}, σ_{11} and σ_{22} in cases 24, 30, 46, 52, 61, 62, 66, 72, 81, 82, 86, 90, and 91 in the table, the following exact formula was used, rather than the approximate formula (23).

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{12}\\ \sigma_{22}\end{bmatrix}=\begin{bmatrix}1-r_{11}^2 & -2r_{11}r_{12} & -r_{12}^2\\ -r_{11}r_{21} & 1-r_{11}r_{22}-r_{12}r_{21} & -r_{12}r_{22}\\ -r_{21}^2 & -2r_{21}r_{22} & 1-r_{22}^2\end{bmatrix}^{-1}\begin{bmatrix}1\\ 0\\ 1\end{bmatrix}$$

The approximation is not good in these cases becasue the value of D becomes extremely small for some value of ω.

lagged value is always greater than its dependence on the other residual. Note that in the table $\sqrt{\rho_1\rho_2}$ is meant to be a measure of the dependence of a residual on its own lagged value and $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{12}}$ a measure of the dependence on the other residual. This observation, however, is only partially true. It is true that comparison of these two measures explain the superiority of (III) over (II) to some extent. But the correspondence is never consistent, and moreover the superiority of (III) over (II) is much greater than one could imagine from the largeness of $\sqrt{\rho_1\rho_2}$ over $\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$.

In conclusion, we believe that the result of this study indicates the following advices: (1) If the residuals are believed to be both autocorrelated and correlated among each other, and if we are to use either a quasi first-difference transformation or Zellner's method, it would be much more rewarding to use the former. (2) Zellner's method is worth trying, for, if the residuals are independent either to begin with or after some transformation, Zellner's method does increase efficiency as proved in Zellner's original paper, and, even if they are not, one can't be much worse off.

References

1. Diananda, P. H., "Some Probability Limit Theorems with Statistical Applications," *Proceedings of Cambridge Philosophical Society,* 49 (1953), 239–46.
2. Grenander, Ulf, and Murray Rosenblatt, *Statistical Analysis of Stationary Time Series,* New York: John Wiley, 1957.
3. Watson, G. S., "Serial Correlation in Regression Analysis I," *Biometrika,* 42 (1955), 327–41.
4. Watson, G. S., and E. J. Hannan, "Serial Correlation in Regression Analysis II," *Biometrika,* 43 (1956), 436–48.
5. Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association,* 57 (1962), 348–68.