Title	BEST LINEAR UNBIASED PREDICTOR AND THE PREDICTION ERROR
Sub Title	
Author	蓑谷, 千凰彦(MINOTANI, CHIOHIKO)
Publisher	Keio Economic Society, Keio University
Publication year	1971
Jtitle	Keio economic studies Vol.8, No.1 (1971.) ,p.49- 66
JaLC DOI	
Abstract	
Notes	
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA00260492-19710001-0049

BEST LINEAR UNBIASED PREDICTOR AND THE PREDICTION ERROR

CHIOHIKO MINOTANI

This paper is concerned with the best linear unbiased predictor (BLUP) and the prediction error when the random vector $\mathbf{y}$ is predicted over several times.

As is well known the best linear predictor (BLP) for some prediction problem exists and is unique under some conditions. ([3], [11], [15])

In section I the BLUP of a single equation and the simultaneous equations model is derived, and this BLUP is proved to satisfy the necessary and sufficient condition of uniqueness as a predictor.

If the conditions that the ordinary least square's estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is BLUE hold over the prediction periods, then the predictor $\mathbf{X}_f\hat{\boldsymbol{\beta}}$ is BLUP but the predictor $\mathbf{X}_f\tilde{\boldsymbol{\beta}}$ is not BLUP even if the conditions that the generalized least square's estimator $\tilde{\boldsymbol{\beta}}$ is BLUE hold over the prediction periods, where $\mathbf{X}_f$ is a $p \times G$ matrix of independent variables in the prediction periods.

A single prediction from a reduced form equations is treated in [7] but in section II we don't treat a single prediction but prediction in several periods and prediction interval is derived.

This paper refers to an extension of [4] and [7].

I

1. The Prediction Problem and Predictor: A Review(1)

Let $\mathbf{y}_s$ $(-\infty < s < \infty)$ be G-variates random vector. The prediction problem is to know the probabilistic structure of the $G \times 1$ vector $\mathbf{y}_s$ and predict the $G \times 1$ vector of future values of $\mathbf{y}_\nu$ $(\nu \in T_f)$ from the $G \times 1$ vector of observed values $\mathbf{y}_t$ $(t \in T_0)$, where T_0 and T_f represent observation periods and prediction periods respectively. Then the problem is to find the predictor $\hat{\mathbf{y}}_\nu$ of $\mathbf{y}_\nu$ $(\nu \in T_f)$ according to some criterion (ordinary the criterion of minimum mean square error is adopted).

Now let H be the linear space spanned by random vector $\mathbf{y}_s$ for all s and it is assumed that $\mathbf{y}_s$ has finite second-order moment. The inner product of $\mathbf{u}$ and $\mathbf{v}$ is defined by the relation

$$(\mathbf{u}, \mathbf{v}) = E(\mathbf{u}'\mathbf{v}) \tag{1-1}$$

and the norm of $\mathbf{u}$ is

$$||\mathbf{u}|| = (\mathbf{u}, \mathbf{u})^{1/2} = [E(\mathbf{u}'\mathbf{u})]^{1/2}\,, \tag{1-2}$$

where $\mathbf{u}, \mathbf{v} \in H$ and $E(\mathbf{u}) = E(\mathbf{v}) = \mathbf{0}$.

(1) This review is mainly owing to [15].

Let H_1 be the linear subspace of H spanned by $\mathbf{y}_t$ $(t \in T_0)$. If $\sum_{r=0}^{n} \mathbf{A}_r^{(n)} \mathbf{y}_{T-r}$, which is the linear combination of the observed value $\mathbf{y}_t$ $(t \in T_0)$, exists and $\hat{\mathbf{y}}_\nu$ can be written as

$$\hat{\mathbf{y}}_\nu = \lim_{n\to\infty} \sum_{r=0}^{n} \mathbf{A}_r^{(n)} \mathbf{y}_{T-r} \tag{1-3}$$

and $\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu = \mathbf{z}_\nu$ is orthogonal to the space H_1; that is, the condition

$$E(\mathbf{z}_\nu' \mathbf{u}) = \mathbf{0} \qquad \text{for any} \quad \mathbf{u} \in H_1 \tag{1-4}$$

is satisfied, then $\hat{\mathbf{y}}_\nu$ is the projection of $\mathbf{y}_\nu$ on the space H_1, where $\mathbf{A}_r^{(n)}$ is the $G \times G$ matrix which element is constant and T represents the length of the observation period. Then $\hat{\mathbf{y}}_\nu$ can be uniquely represented as the sum of two vectors

$$\mathbf{y}_\nu = \hat{\mathbf{y}}_\nu + \mathbf{z}_\nu\,, \tag{1-5}$$

where $\hat{\mathbf{y}}$ is in H_1, while $\mathbf{z}_\nu$ is orthogonal to H_1 (Projection Theorem). Further, for any $\hat{\mathbf{y}}_\nu^* \in H_1$, that is, any linear function $\hat{\mathbf{y}}_\nu^*$ of the observed value $\mathbf{y}_t$ $(t \in T_0)$, the following inequality holds

$$E[(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu^*)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu^*)] \geqq E[(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)] \tag{1-6}$$

with equality holding when $\hat{\mathbf{y}}_\nu^* = \hat{\mathbf{y}}_\nu$.

The equation (1-6) can be easily shown. Since $(\hat{\mathbf{y}}_\nu^* - \hat{\mathbf{y}}_\nu) \in H_1$, it is orthogonal to $(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)$. Then

$$\begin{aligned} &E[(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu^*)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu^*)] \\ &\quad = E\{[\mathbf{y}_\nu - \hat{\mathbf{y}} - (\hat{\mathbf{y}}_\nu^* - \hat{\mathbf{y}}_\nu)]'[\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu - (\hat{\mathbf{y}}_\nu^* - \hat{\mathbf{y}}_\nu)]\} \\ &\quad = E[(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)] + E[(\hat{\mathbf{y}}_\nu^* - \hat{\mathbf{y}}_\nu)'(\hat{\mathbf{y}}_\nu^* - \hat{\mathbf{y}}_\nu)] \\ &\quad \geqq E[(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)]\,. \end{aligned}$$

And hence if we take the $\hat{\mathbf{y}}_\nu$ which satisfies the condition of (1-3) and (1-4) as the predictor of $\mathbf{y}_\nu$, then it minimizes the mean square of prediction errors, that is, it is the best linear predictor. Now, further $E(\hat{\mathbf{y}}_\nu - \mathbf{y}_\nu) = \mathbf{0}$, then the predictor $\hat{\mathbf{y}}_\nu$ is thc best linear unbiased predictor (BLUP).

We have shown that the predictor $\hat{\mathbf{y}}_\nu$ which satisfies two conditions (1-3) and (1-4) posses a smaller mean square forecast error than any linear predictors, the linearity property indicating here that the predictors are linear function of $\mathbf{y}_t$ $(t \in T_0)$.

Next we shall show that if we consider any measurable functions of $\mathbf{y}_t$ by relaxing the restrictions of the linearity, then $E(\mathbf{y}_\nu \mid \mathbf{y}_t)$, that is, the conditional expectation of $\mathbf{y}_\nu$ $(\nu \in T_f)$ given $\mathbf{y}_t$ $(t \in T_0)$ or regression function is the best predictor of the class of any measurable functions, where the existence of $E(\mathbf{y}_\nu \mid \mathbf{y}_t)$ is assumed. Let any measurable function be

$$\boldsymbol{\phi}_\nu = \boldsymbol{\phi}_\nu(\mathbf{y}_t,\ t \in T_0)$$

and the conditional expectation of $\mathbf{y}_\nu$ given $\mathbf{y}_t$ be

$$\boldsymbol{\phi}_\nu = E(\mathbf{y}_\nu \mid \mathbf{y}_t,\ t \in T_0,\ \nu \in T_f)\,, \tag{1-7}$$

where

$$\boldsymbol{\psi}_\nu = \begin{bmatrix} \psi_{1\nu}(\mathbf{y}_t) \\ \vdots \\ \psi_{G\nu}(\mathbf{y}_t) \end{bmatrix}, \qquad t \in T_0\,, \quad \nu \in T_f$$

and

$$\mathbf{y}_\nu = \begin{bmatrix} y_{1\nu} \\ \vdots \\ y_{G\nu} \end{bmatrix}, \qquad \nu \in T_f$$

then $\boldsymbol{\phi}_\nu = E(\mathbf{y}_\nu \mid \mathbf{y}_t)$ stands for

$$\boldsymbol{\phi}_\nu = \begin{bmatrix} \phi_{1\nu} \\ \vdots \\ \phi_{G\nu} \end{bmatrix} = \begin{bmatrix} E(y_{1\nu} \mid \mathbf{y}_t) \\ \vdots \\ E(y_{G\nu} \mid \mathbf{y}_t) \end{bmatrix}.$$

[*Proof*]

Let E_1 be the expectation of $\mathbf{y}_t$ $(t \in T_0)$. Since

$$\begin{aligned} E[(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)'(\boldsymbol{\psi}_\nu - \boldsymbol{\phi}_\nu)] &= E_1\{E[(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)'(\boldsymbol{\psi}_\nu - \boldsymbol{\phi}_\nu) \mid \mathbf{y}_t,\ t \in T_0, \nu \in T_f]\} \\ &= \{E[(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu) \mid \mathbf{y}_t,\ t \in T_0, \nu \in T_f]\}'\{E_1(\boldsymbol{\psi}_\nu - \boldsymbol{\phi}_\nu)\} \\ &= 0\,, \end{aligned}$$

then

$$\begin{aligned} E[(\mathbf{y}_\nu - \boldsymbol{\psi}_\nu)'(\mathbf{y}_\nu - \boldsymbol{\psi}_\nu)] &= E[(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)'(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)] + E[(\boldsymbol{\psi}_\nu - \boldsymbol{\phi}_\nu)'(\boldsymbol{\psi}_\nu - \boldsymbol{\phi}_\nu)] \\ &\geqq [(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)'(\mathbf{y}_\nu - \boldsymbol{\phi}_\nu)] \end{aligned} \tag{1-8}$$

and hence the mean square prediction error becomes minimum when $\boldsymbol{\psi}_\nu = \boldsymbol{\phi}_\nu$.

We didn't specify the probability distribution of $\mathbf{y}_s$ until now, but if $\mathbf{y}_s$ is distributed according to G-variates normal distribution, the best linear predictor which satisfies two conditions (1-3) and (1-4) is the best predictor among the class of any preditors. For $\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu$ is not correlated to any vector $\mathbf{u} \in H_1$ and under the assumption of normality the lack of correlation is equivalent to the statistical independency, so any measurable functions of the observed value $\mathbf{y}_t$, that is, $\boldsymbol{\psi}_\nu = \boldsymbol{\psi}_\nu(\mathbf{y}_t)$, are indepent of $\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu$. Therefore the relation

$$E[(\mathbf{y}_\nu - \boldsymbol{\psi}_\nu)'(\mathbf{y}_\nu - \boldsymbol{\psi}_\nu)] \geqq E(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)'(\mathbf{y}_\nu - \hat{\mathbf{y}}_\nu)] \tag{1-9}$$

is easily verified. Eventually if $\mathbf{y}_s$ is distributed according to normal distribution, then the linear predictor of $\mathbf{y}_t$ $(t \in T_0)$ becomes the best one of the class of any other predictors represented as the function of $\mathbf{y}_t$ $(t \in T_0)$.

2. Linear Regression Function as a Predictor

We know that the regression function $\boldsymbol{\phi}$ is the best predictor and the linear regression function is the best predictor if $\mathbf{y}_s$ is distributed according to normal distribution. This linear regression function has the following properties as well as the minimum mean square error predictor.

(1) Unbiased predictor

$$E(\boldsymbol{\phi}_\nu - \mathbf{y}_\nu) = \mathbf{0}\,, \qquad \nu \in T_f\,.$$

This is obvious from the definition of the regression function $\boldsymbol{\phi}$.

(2) Predictor having maximum correlation

Let $\psi_{i\nu}$ be any predictor for $y_{i\nu}$ and $\phi_{i\nu}$ be a regression function. Then

$$|\rho(y_{i\nu}, \psi_{i\nu})| \leqq \rho(y_{i\nu}, \phi_{i\nu}) \qquad i = 1, \ldots, G\,, \tag{1-10}$$

where $\rho(u, v)$ stands for the simple correlation coefficient between u and v.(2)

3. BLUP of a Single Equation

We shall now get the best linear unbiased predictor (BLUP) in the case of the single equation, that is, $G = 1$.

The single equation regression model may be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\,, \tag{1-11}$$

where $\mathbf{y}$ is column vector of T observations on the dependent variable, $\mathbf{X}$ a $T \times k$ matrix of nonstochastic or fixed values taken by the k independent variables ($k \leqq T$), $\boldsymbol{\beta}$ a $k \times 1$ column vector of k unknown regression coefficients, and $\mathbf{u}$ a $T \times 1$ vector of disturbances. We shall assume

$$\begin{aligned} &E(\mathbf{u}) = \mathbf{0}\,, \\ &E(\mathbf{uu}') = \boldsymbol{\Omega}\,, \qquad \operatorname{rank} \boldsymbol{\Omega} = T\,, \\ &\operatorname{rank} \mathbf{X} = k\,. \end{aligned} \tag{1-12}$$

We further assume that the forecast is made over the p prediction periods and we have a set of values available for the independent variables for the prediction periods, say $\mathbf{X}_f$, (where $\mathbf{X}_f$ is a $p \times k$ matrix) then the true value of $\mathbf{y}$, say $\mathbf{y}_f$ (where $\mathbf{y}_f$ is a $p \times 1$ vector), in this prediction periods, can be represented as

$$\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \mathbf{u}_f\,, \tag{1-13}$$

where $\mathbf{u}_f$ is a $p \times 1$ vector of disturbances in the prediction periods. We shall assume

$$\begin{aligned} &E(\mathbf{u}_f) = \mathbf{0}\,, \\ &E(\mathbf{u}_f\mathbf{u}') = \mathbf{B}\,, \\ &E(\mathbf{u}_f\mathbf{u}_f') = \boldsymbol{\Omega}_f\,. \end{aligned} \tag{1-14}$$

---

(2) [12].

Let the linear unbiased predictor of $\mathbf{y}_f$ be a $\hat{\mathbf{y}}_f$, that is,

$$\hat{\mathbf{y}}_f = \mathbf{Ay} \tag{1-15}$$

where $\mathbf{A}$ is a $p \times T$ matrix of constants. If $\hat{\mathbf{y}}_f$ is to be an unbiased predictor of $\mathbf{y}_f$, that is, $E(\hat{\mathbf{y}}_f - \mathbf{y}_f) = \mathbf{0}$, we have

$$\begin{aligned} E(\hat{\mathbf{y}}_f) &= E(\mathbf{Ay}) \\ &= \mathbf{AX}\boldsymbol{\beta} \\ &= \mathbf{X}_f\boldsymbol{\beta} \\ &= E(\mathbf{y}_f) \end{aligned}$$

if and only if

$$\mathbf{AX} = \mathbf{X}_f\,. \tag{1-16}$$

Let

$$\mathbf{z}_f = \mathbf{y}_f - \hat{\mathbf{y}}_f = (\mathbf{X}_f - \mathbf{AX})\boldsymbol{\beta} + \mathbf{u}_f - \mathbf{Au}\,, \tag{1-17}$$

represent a $p \times 1$ column vector of the prediction errors. Then,

$$\mathrm{Var}\,(\mathbf{z}_f) = \boldsymbol{\Omega}_f - \mathbf{BA}' - \mathbf{AB}' + \mathbf{A}\boldsymbol{\Omega}\mathbf{A}'\,. \tag{1-18}$$

The problem to seek the BLUP of $\mathbf{y}_f$ becomes to find $\mathbf{A}$ such that $\mathrm{tr}\,[\mathrm{Var}\,(\mathbf{z}_f)]$ is minimized subject to condition $\mathbf{AX} = \mathbf{X}_f$, where $\mathrm{tr}\,[\mathrm{Var}\,(z_f)]$ represents the trace of $\mathrm{Var}\,(\mathbf{z}_f)$. The minimization problem can be solved by the Lagrangian method, that is, the problem may be described as the minimization of

$$S = \mathrm{tr}\,[\mathrm{Var}\,(\mathbf{z}_f)] - \mathrm{tr}\,[\boldsymbol{\Lambda}(\mathbf{AX} - \mathbf{X}_f)]\,, \tag{1-19}$$

where $\boldsymbol{\Lambda}$ is a $k \times p$ matrix of Lagrangian multipliers. Differentiating (1-19) with respect to $\mathbf{A}$ and $\boldsymbol{\Lambda}$ and setting equal to zero we obtain

$$\mathbf{A} = \mathbf{B}\boldsymbol{\Omega}^{-1} + \boldsymbol{\Lambda}'\mathbf{X}'\boldsymbol{\Omega}^{-1}\,. \tag{1-20}$$

If we postmultiply (1-20) by $\mathbf{X}$ we find

$$\boldsymbol{\Lambda}' = \mathbf{X}_f(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\,. \tag{1-21}$$

Substituting (1-21) in (1-20), we have

$$\mathbf{A} = \mathbf{B}\boldsymbol{\Omega}^{-1} + \mathbf{X}_f(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1} - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\,. \tag{1-22}$$

Thus the BLUP is

$$\begin{aligned} \hat{\mathbf{y}}_f &= \mathbf{Ay} \\ &= \mathbf{X}_f(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} + \mathbf{B}\boldsymbol{\Omega}^{-1}[\mathbf{y} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}] \end{aligned} \tag{1-23}$$

Now the expression $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} = \tilde{\boldsymbol{\beta}}$, say, is an Aitken's generalized least squares estimator of $\boldsymbol{\beta}$, we may therefore rewrite (1-23) as

$$\hat{\mathbf{y}}_f = \mathbf{X}_f\tilde{\boldsymbol{\beta}} + \mathbf{B}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\,. \tag{1-24}$$

We shall give some examples of the BLUP (1-24).

Case 1.

If

$$E(u_t u_{t+s}) = 0 \qquad \text{for all} \quad t \in T_0 \cup T_f \text{ and } s \neq 0 \in T_0 \cup T_f ,$$

that is, $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$, $\mathbf{B} = \mathbf{0}$, then

$$\hat{\mathbf{y}}_f = \mathbf{X}_f\tilde{\boldsymbol{\beta}}$$

is the BLUP of $\mathbf{y}_f$, where $\tilde{\boldsymbol{\beta}}$ is the ordinary least equares estimator of $\boldsymbol{\beta}$.

Case 2.

If

$$u_t = \rho u_{t-1} + \varepsilon_t , \qquad |\rho| < 1 ,$$
$$E(\varepsilon_t) = 0 ,$$
$$E(\varepsilon_t \varepsilon_{t+s}) = \begin{cases} \sigma_\varepsilon^2 & (s = 0) \\ 0 & (s \neq 0) \end{cases},$$

then

$$\boldsymbol{\Omega}^{-1} = \frac{1}{(1-\rho^2)\sigma^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix},$$

$$\mathbf{B} = \sigma^2\begin{bmatrix} \rho^T & \rho^{T-1} & \cdots & \rho \\ \rho^{T+1} & \rho^T & \cdots & \rho^2 \\ \vdots & \vdots & & \vdots \\ \rho^{T+p-1} & \rho^{T+p-2} & \cdots & \rho^p \end{bmatrix},$$

so

$$\mathbf{B}\boldsymbol{\Omega}^{-1} = \begin{bmatrix} 0 & \cdots & 0 & \rho \\ 0 & \cdots & 0 & \rho^2 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \rho^p \end{bmatrix},$$

therefore

$$\hat{\mathbf{y}}_f = \mathbf{X}_f\tilde{\boldsymbol{\beta}} + (\rho \quad \rho^2 \cdots \rho^p)'\hat{u}_T$$

becomes the BLUP of $\mathbf{y}_f$, where $\hat{u}_T$ is the residual of the last observational period.

We shall now show that the BLUP $\hat{\mathbf{y}}_f$ given by (1-24) satisfies the conditions (1-3) and (1-4).

The condition (1-3) is obviously satisfied since $\hat{\mathbf{y}}_f = \mathbf{A}\mathbf{y}$. We have now

$$\mathbf{y}_f - \hat{\mathbf{y}}_f = \mathbf{X}_f(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{u}_f - \mathbf{B}\Omega^{-1}\hat{\mathbf{u}}$$

and

$$\mathbf{X}_f(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = -\mathbf{X}_f(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{u}\,,$$

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{M}\mathbf{u}\,,$$

where $\mathbf{M} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}]$, hence

$$E(\hat{\mathbf{u}}\hat{\mathbf{u}}') = \Omega\mathbf{M}'\,,$$

$$E(\mathbf{u}_f\hat{\mathbf{u}}') = \mathbf{B}\mathbf{M}'\,,$$

$$E(\mathbf{u}\hat{\mathbf{u}}') = \Omega\mathbf{M}'\,.$$

Thus

$$\begin{aligned} E[(\mathbf{y}_f - \hat{\mathbf{y}}_f)\hat{\mathbf{u}}'] &= -\mathbf{X}_f(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega\Omega^{-1}\mathbf{M}' + \mathbf{B}\mathbf{M}' - \mathbf{B}\Omega^{-1}\Omega\mathbf{M} \\ &= -\mathbf{X}_f(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}' \\ &= \mathbf{0}\,, \end{aligned}$$

since $\mathbf{M}\mathbf{X} = \mathbf{0}$. Therefore the relation

$$y_\nu - \hat{y}_\nu \perp \hat{u}_1, \hat{u}_2, \ldots, \hat{u}_T$$

holds for all $\nu \in T_f$ ($\nu = 1, 2, \ldots, p$). Thus we have shown that the condition (1-4) is also satisfied.

So far we have made no assumptions about the probability distribution of the disturbance term. If the disturbance is distributed according to normal distribution, then the BLUP (1-24) is the best predictor of the class of any measurable functions of $\mathbf{y}_t$, that is, the minimum variance unbiased predictor, as we have shown in section II-1. This property can also be shown by the following way.

Let the conditional experctation of $\mathbf{y}_f$ ($f \in T_f$) given $\mathbf{y}_t$ ($t \in T_0$) be $R(\mathbf{y}_f, \Theta)$, that is,

$$R(\mathbf{y}_f, \Theta) = E(\mathbf{y}_f \,|\, \mathbf{y})\,, \tag{1-25}$$

where Θ is the vector of parameters and $\mathbf{y} = \mathbf{y}_t$ ($t \in T_0$). If $R(\mathbf{y}_f, \Theta)$ can be decomposed as

$$R(\mathbf{y}_f, \Theta) = T(\mathbf{y}) + \xi(\Theta)\,, \tag{1-26}$$

where $T(\mathbf{y})$ is the function depending solely on the vector $\mathbf{y}$ and not on Θ and $\xi(\Theta)$ is the function of Θ, then

$$\phi^*(\mathbf{y}) = T(\mathbf{y}) + H(\mathbf{y}) \tag{1-27}$$

is the minimum variance unbiased predictor of $\mathbf{y}_f$, where $H(\mathbf{y})$ is the minimum variance unbiased estimator of $\xi(\Theta)$.[3] By assumption, since $\mathbf{y}$ is normally distributed

$$E(\mathbf{y}_f \,|\, \mathbf{y}) = E(\mathbf{y}_f) + \operatorname{Cov}(\mathbf{y}_f, \mathbf{y}) \cdot [\operatorname{Var}(\mathbf{y})]^{-1}[\mathbf{y} - E(\mathbf{y})]\,,$$

(3) [14].

where $\text{cov}(\mathbf{y}_f, \mathbf{y})$ represents the covariance between $\mathbf{y}_f$ and $\mathbf{y}$. Now we substitute

$$\begin{aligned} &E(\mathbf{y}_f) = \mathbf{X}_f\boldsymbol{\beta} , \\ &\text{Cov}(\mathbf{y}_f, \mathbf{y}) = E(\mathbf{u}_f\mathbf{u}') = \mathbf{B} , \\ &\text{Var}(\mathbf{y}) = \boldsymbol{\Omega} , \\ &\mathbf{y} - E(\mathbf{y}) = \mathbf{u} , \end{aligned}$$

in the above equation we have

$$\begin{aligned} R(\mathbf{y}_f, \Theta) &= \mathbf{X}_f\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{u} \\ &= \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{y} + (\mathbf{X}_f - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X})\boldsymbol{\beta} \\ &= T(\mathbf{y}) + \xi(\Theta) , \end{aligned}$$

where $T(\mathbf{y}) = \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{y}$ and $\xi(\Theta) = (\mathbf{X}_f - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X})\boldsymbol{\beta}$ since $\mathbf{B}$ and $\boldsymbol{\Omega}$ are assumed to be known and only $\boldsymbol{\beta}$ is the parameters. Let the minimum variance unbiased estimator of $\boldsymbol{\beta}$ be $\tilde{\boldsymbol{\beta}}$. Now $\tilde{\boldsymbol{\beta}}$ is given by the Aitken's generalized least squares method, then

$$\begin{aligned} \phi^*(\mathbf{y}) &= \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{y} + (\mathbf{X}_f - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X})\tilde{\boldsymbol{\beta}} \\ &= \mathbf{X}_f\tilde{\boldsymbol{\beta}} + \mathbf{B}\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}_f\tilde{\boldsymbol{\beta}}) . \end{aligned}$$

Thus $\phi^*(\mathbf{y})$ coincide with the relation (1–24).

4. The BLUP of the Simultaneous Equations Systems

The BLUP of the single equation model which we have obtained can be easily extended to the simultaneous equations systems. The simultaneous equations systems can be written as

$$\mathbf{Y} = \boldsymbol{\Pi}\mathbf{X} + \mathbf{V} , \tag{1–28}$$

where $\mathbf{Y}$ represents a $G \times T$ matrix of jointly dependent variables, $\boldsymbol{\Pi}$ a $G \times K$ matrix of unknown coefficients, $\mathbf{X}$ a $K \times T$ nonstochastic matrix of independent variables and rank $\mathbf{X} = K \leqq T$, and $\mathbf{V}$ a $G \times T$ matrix of disturbance terms.

It is assumed that a nonstochastic matrix of the independent variables in the prediction period, say, $\mathbf{X}_f$ (where $\mathbf{X}_f$ is a $K \times P$ matrix) is known and the relation represented by (1–28) is also satisfied in the prediction period between $\mathbf{X}_f$ and the true value of $\mathbf{Y}$, say $\mathbf{Y}_f$ (where $\mathbf{Y}_f$ is a $G \times P$ matrix), where P represents the length of the prediction periods. Then the true value of $\mathbf{Y}$ can be written as

$$\mathbf{Y}_f = \boldsymbol{\Pi}\mathbf{X}_f + \mathbf{V}_f , \tag{1–29}$$

where $\mathbf{V}_f$ is a $G \times P$ matrix of the disturbance terms in the prediction period. It is further assumed that

$$\begin{aligned} &E(\mathbf{V}) = \mathbf{0} , \\ &E(\mathbf{V}'\mathbf{V}) = \boldsymbol{\Omega} , \qquad \text{rank}\,\boldsymbol{\Omega} = T , \\ &E(\mathbf{V}'\mathbf{V}_f) = \mathbf{B} . \end{aligned} \tag{1–30}$$

Let the BLUP of $\mathbf{Y}_f$ be $\hat{\mathbf{Y}}_f$. Then

$$E(\hat{\mathbf{Y}}_f - \mathbf{Y}_f) = \mathbf{0} \Leftrightarrow \mathbf{XA} = \mathbf{X}_f \tag{1-31}$$

and our problem is to minimize

$$S = \mathrm{tr}\,[\mathrm{Var}\,(\mathbf{Z}_f)] - \mathrm{tr}\,[\boldsymbol{\Lambda}(\mathbf{XA} - \mathbf{X}_f)]\,, \tag{1-32}$$

where $\mathbf{Z}_f$ is a $G \times P$ matrix of the prediction errors and $\boldsymbol{\Lambda}$ is a $P \times K$ matrix of Lagrangian multiplies. Solving

$$\begin{cases} \dfrac{\partial S}{\partial \mathbf{A}} = -2\mathbf{B} + 2\boldsymbol{\Omega}\mathbf{A} - \mathbf{X}'\boldsymbol{\Lambda}' = \mathbf{0}\,, \\ \dfrac{\partial S}{\partial \boldsymbol{\Lambda}} = (\mathbf{XA} - \mathbf{X}_f)' = \mathbf{0}\,, \end{cases}$$

we obtain

$$\mathbf{A} = \boldsymbol{\Omega}^{-1}\mathbf{B} + \boldsymbol{\Omega}^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X}')^{-1}(\mathbf{X}_f - \mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{B})\,. \tag{1-33}$$

Then the BLUP of $\mathbf{Y}_f$ is

$$\begin{aligned} \hat{\mathbf{Y}}_f &= \mathbf{Y}\boldsymbol{\Omega}^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X}')^{-1}\mathbf{X}_f + (\mathbf{Y} - \mathbf{Y}\boldsymbol{\Omega}^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X}')^{-1}\mathbf{X})\boldsymbol{\Omega}^{-1}\mathbf{B} \\ &= \tilde{\boldsymbol{\Pi}}\mathbf{X}_f + (\mathbf{Y} - \tilde{\boldsymbol{\Pi}}\mathbf{X})\boldsymbol{\Omega}^{-1}\mathbf{B}\,, \end{aligned} \tag{1-34}$$

where $\tilde{\boldsymbol{\Pi}} = \mathbf{Y}\boldsymbol{\Omega}^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\Omega}^{-1}\mathbf{X}')^{-1}$.

The relation

$$E[(\mathbf{Y} - \tilde{\boldsymbol{\Pi}}\mathbf{X})'(\mathbf{Y}_f - \hat{\mathbf{Y}}_f)] = \mathbf{0}$$

can be easily demonstrated, then we find that

$$\mathbf{y}_{fi} - \hat{\mathbf{y}}_{fi} \perp \tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_T \qquad \text{for all} \quad i = 1, \ldots, P\,,$$

where $\mathbf{y}_{fi}$, $\hat{\mathbf{y}}_{fi}$ and $\tilde{\mathbf{v}}_i$ are the i-th column vector ($G \times 1$) in $\mathbf{Y}_f$, $\hat{\mathbf{Y}}_f$ and $\mathbf{Y} - \tilde{\boldsymbol{\Pi}}\mathbf{X}$ respectively.

The equations (1–28) can be interpreted as the reduced form equations in which $\mathbf{X}$ consists of exogeneous variables only and include no lagged endogeneous variables since $\mathbf{X}$ is assumed to be nonstochastic matrix. Thus the BLUP of $\mathbf{Y}_f$ in the simultaneous equatios systems of which the reduced form equations are (1–28) is given by (1–34).

We shall examine the following special case. A $G \times T$ matrix of disturbance terms is

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_T)\,,$$

where $\mathbf{v}_i$ a $G \times 1$ column vector and it is assumed that

$$\begin{aligned} &E(\mathbf{v}_i) = \mathbf{0}\,, \\ &E(\mathbf{v}_i\mathbf{v}_j') = \begin{cases} \mathbf{0} & i \neq j \\ \boldsymbol{\Sigma} & i = j \end{cases}, \qquad \text{for all} \quad i, j = 1, \ldots, T\,. \\ &\boldsymbol{\Sigma} = \{\sigma_{uv}\}\,, \end{aligned} \tag{1-35}$$

This assumption allows for contemporaneous correlation among the disturbances but not for non-contemporaneous correlation among the disturbances. Then we have

$$E(\mathbf{V}'\mathbf{V}) = \sigma^2\mathbf{I}\,,$$
$$E(\mathbf{V}'\mathbf{V}_f) = \mathbf{0}\,,$$

where $\sigma^2 = \sum_{i=1}^{G} \sigma_{ii}$, and hence

$$\tilde{\boldsymbol{\Pi}} = \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1} = \hat{\boldsymbol{\Pi}}\,,$$

therefore

$$\hat{\mathbf{Y}}_f = \hat{\boldsymbol{\Pi}}\mathbf{X}_f = \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}_f$$

is the BLUP of $\mathbf{Y}_f$.

II. THE DISTRIBUTION OF THE PREDICTION ERROR

1. A Single Equation

We shall begin by analyzing the prediction error of the single equation. The true value of $\mathbf{y}$ for the prediction periods can be written as

$$\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \mathbf{u}_f$$

and we have shown in (1–24) that the BLUP of $\mathbf{y}_f$ is given by

$$\hat{\mathbf{y}}_f = \mathbf{X}_f\tilde{\boldsymbol{\beta}} + \mathbf{B}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{u}}\,,$$

where $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ and the assumptions (1–12) and (1–14) are satisfied. Let a $p \times 1$ vector of the prediction errors be $\mathbf{z}_f$. Then $\mathbf{z}_f$ can be written as

$$\mathbf{z}_f = \mathbf{y}_f - \hat{\mathbf{y}}_f = \mathbf{X}_f(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \mathbf{u}_f - \mathbf{B}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{u}}\,, \tag{2–1}$$

so the cause of the prediction error is

(1) the estimation error of the parameters vector $\boldsymbol{\beta}$
(2) the value of the disturbance term $\mathbf{u}_f$

since $\mathbf{B}$ and $\boldsymbol{\Omega}$ are assumed to be known. It is further assumed that $\mathbf{X}_f$, a $p \times k$ matrix of the independent variables, is available. Let the covariance matrix of $\mathbf{z}_f$ be $\boldsymbol{\Sigma}_f$ then we have

$$\begin{aligned}\boldsymbol{\Sigma}_f &= E(\mathbf{z}_f\mathbf{z}_f') \\ &= \mathbf{X}_f\mathbf{S}^{-1}\mathbf{X}_f' - \mathbf{X}_f\mathbf{S}^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{B}' - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{X}\mathbf{S}^{-1}\mathbf{X}_f' + \boldsymbol{\Omega}_f - \mathbf{B}\boldsymbol{\Omega}^{-1}\mathbf{M}\mathbf{B}'\,,\end{aligned} \tag{2–2}$$

where $\mathbf{S} = \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}$, $\mathbf{M} = \mathbf{I} - \mathbf{X}\mathbf{S}^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}$.

We shall give the distribution of the prediction error and the prediction interval in the simple case where the following assumptions for the disturbance are made,

$$\boldsymbol{\Omega} = \sigma^2\mathbf{I}_T\,, \qquad \boldsymbol{\Omega}_f = \sigma^2\mathbf{I}_p\,, \qquad \mathbf{B} = \mathbf{0}\,, \tag{2–3}$$

that is,

$$E(u_t u_{t-s}) = \begin{cases} \sigma^2 & s = 0 \\ 0 & s \neq 0 \end{cases}$$

for all $t, s \in T_0 \cup T_f$. Then the BLUP $\hat{\mathbf{y}}_f$ becomes

$$\hat{\mathbf{y}}_f = \mathbf{X}_f \hat{\boldsymbol{\beta}}$$

and the prediction error $\mathbf{z}_f$ becomes

$$\mathbf{z}_f = \mathbf{X}_f(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{u}_f\,,$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, so we obtain

$$\boldsymbol{\Sigma}_f = \sigma^2[\mathbf{I}_p + \mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_f'] = \sigma^2\mathbf{G}\,. \tag{2-4}$$

We know from (2-4) that $\boldsymbol{\Sigma}_f$ is the sum of the covariance matrix of $\hat{\mathbf{y}}_f$, $\sigma^2\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_f'$, and the covariance matrix of $\mathbf{u}_f$, $\sigma^2\mathbf{I}_p$, where the suffix T or p of $\mathbf{I}_T$ or $\mathbf{I}_p$ represents the order of the identity matrix $\mathbf{I}$. Let the unbiased estimator of σ^2 be $\hat{\sigma}^2$, then

$$\hat{\boldsymbol{\Sigma}}_f = \hat{\sigma}^2\mathbf{G} \tag{2-5}$$

becomes the unbiased estimator of $\boldsymbol{\Sigma}_f$. If $\mathbf{u}$ is normally distributed, then

$$\mathbf{z}_f \sim N(\mathbf{0}, \sigma^2\mathbf{G}) \tag{2-6}$$

and for any non-zero $p \times 1$ vector $\mathbf{a}$

$$\mathbf{a}'\mathbf{z}_f \sim N(0, \sigma^2\mathbf{a}'\mathbf{G}\mathbf{a})\,, \tag{2-7}$$

hence

$$\mathbf{a}'\mathbf{z}_f(\sigma^2\mathbf{a}'\mathbf{G}\mathbf{a})^{-1/2} \sim N(0, 1)\,. \tag{2-8}$$

We know further that

(1) $(T - k)\hat{\sigma}^2/\sigma^2 \sim \chi^2(T - k)$,

(2) $\hat{\boldsymbol{\beta}}$ is independent of $\hat{\sigma}^2$,

(3) $\mathbf{u}_f$ is independent of $\mathbf{u}$ and hence $\mathbf{u}_f$ is independent of $\hat{\sigma}^2$, therefore $\mathbf{z}_f$ is independent of $\hat{\boldsymbol{\Sigma}}_f$. Thus we obtain

$$\mathbf{a}'\mathbf{z}_f[\hat{\sigma}^2\mathbf{a}'\mathbf{G}\mathbf{a}]^{-1/2} \sim t(T - k)\,. \tag{2-9}$$

Let t_α be the critecal point of the t distribution with $T - k$ degrees of freedom, where α is the level of the significance. We have then the probability statement

$$P\{|\mathbf{a}'\mathbf{z}_f[\hat{\sigma}^2\mathbf{a}'\mathbf{G}\mathbf{a}]^{-1/2}| \leqq t_\alpha\} \geqq 1 - \alpha\,, \tag{2-10}$$

hence from (2-10) the joint confidence interval of $\mathbf{a}'\mathbf{y}_f$ for all $\mathbf{a}$ at the α level of significance can be obtained as

$$P\{\mathbf{a}'\mathbf{y}_f \in [\mathbf{a}'\mathbf{y}_f \pm \hat{\sigma}(\mathbf{a}'\mathbf{G}\mathbf{a})^{1/2}t_\alpha]\} \geqq 1 - \alpha\,. \tag{2-11}$$

If, in paticular, $\mathbf{a}$ is a $p \times 1$ vector with unit in the i-th element and zeros everywhere else, then from (2–11) the confidence interval for y_{fi} can be obtained as

$$P\{y_{fi} \in [\hat{y}_{fi} \pm \hat{\sigma}(1 + \chi_{fi}(\mathbf{X}'\mathbf{X})^{-1}\chi'_{fi})^{1/2}t_\alpha]\} \geqq 1 - \alpha\,, \qquad (2\text{–}12)$$

where y_{fi} and $\hat{y}_{fi}$ represent the i-th element in the $p \times 1$ vector $\mathbf{y}_f$ and $\hat{\mathbf{y}}_f$ respectively and χ_{fi} the i-th row vector $(1 \times k)$ in the $p \times k$ matrix $\mathbf{X}_f$.

Let $\mathbf{z}_{fi}$ be the i-th element in the $p \times 1$ vector $\mathbf{z}_f$ and $\hat{\sigma}_f(i, i)$ be the i-th diagonal element in $\hat{\boldsymbol{\Sigma}}_f$. Then from (2–9) we obtain

$$\mathbf{z}_{fi}[\hat{\sigma}_f(i, i)]^{-1}\mathbf{z}'_{fi} \sim F(1, T - k)\,. \qquad (2\text{–}13)$$

Rewriting (2–9) as (2–13) we shall know in next section that the prediction interval for the simultaneous equations systems is the generalization of (2–13).

2. The Simultaneous Equations Systems

The result of the previous section can be extended to the simultaneous equations systems. The model can be written as

$$\mathbf{Y} = \boldsymbol{\Pi}\mathbf{X} + \mathbf{V}\,, \qquad (2\text{–}14)$$

where $\mathbf{Y}$ represents a $G \times T$ matrix of jointly dependent variable, $\mathbf{X}$ a $K \times T$ nonstochastic matric of independent variables and $\operatorname{rank}\mathbf{X} = K \leqq T$, $\boldsymbol{\Pi}$ a $G \times K$ matrix of unknown coefficients and $G \leqq T - K$, and $\mathbf{V}$ a $G \times T$ matrix of disturbance terms.

Let

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_T)$$

and the $G \times 1$ vector $\mathbf{y}_i$ is assumed normally distributed with a null expectation vector and covariance matrix $\boldsymbol{\Sigma}$. This assumption permits contemporanous correlation among the disturbances, but no autocorrelation, that is, the same as the assumption (1–35). We now write this assumption as

$$\mathbf{V} \sim N(\mathbf{0}, \boldsymbol{\Sigma})\,,$$

then the joint probability density function of $\mathbf{Y}$ can be written as

$$\begin{aligned} f(\mathbf{Y}) &= \frac{1}{(2\pi)^{GT/2}|\Sigma|^{T/2}} \exp\left\{-\frac{1}{2}\Sigma(\mathbf{y}_i - \boldsymbol{\Pi}\mathbf{X}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\Pi}\mathbf{X}_i)\right\} \\ &= \frac{1}{(2\pi)^{GT/2}|\Sigma|^{T/2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\Pi}\mathbf{X})(\mathbf{Y} - \boldsymbol{\Pi}\mathbf{X})'\right\}, \end{aligned} \qquad (2\text{–}15)$$

where $\mathbf{y}_i$ $(G \times 1)$ represents the i-th column vector of $\mathbf{Y}$ and $\mathbf{X}_i$ $(K \times 1)$ the i-th column vector of $\mathbf{X}$.

Under the assumption that the structure

$$S = S(\boldsymbol{\Pi}, \boldsymbol{\Sigma})$$

is unchanged for all $t \in T_0 \cup T_f$, the future value $\mathbf{Y}_f$ $(G \times P)$ of $\mathbf{Y}$ can be given as

$$\mathbf{Y}_f = \boldsymbol{\Pi}\mathbf{X}_f + \mathbf{V}_f\,, \tag{2-16}$$

where P, $\mathbf{X}_f$ $(G \times P)$ and $\mathbf{V}_f$ $(G \times P)$ are the same as in the relation (1–28). The assumption (1–35) is satisfied for each column vector of $\mathbf{V}_f$, that is,

$$\begin{aligned} E(\mathbf{v}_{fi}\mathbf{v}'_{fj}) &= \begin{cases} \mathbf{0} & i \neq j \\ \boldsymbol{\Sigma} & i = j \end{cases}, \quad i, j \in T_f\,, \quad t \in T_0\,, \\ E(\mathbf{v}_{fi}\mathbf{v}'_t) &= \mathbf{0}\,, \end{aligned} \tag{2-17}$$

hence, as proved in I–4, the BLUP of $\mathbf{Y}_f$ is given as

$$\hat{\mathbf{Y}}_f = \hat{\boldsymbol{\Pi}}\mathbf{X}_f\,, \tag{2-18}$$

where $\hat{\boldsymbol{\Pi}} = \mathbf{YX}'(\mathbf{XX}')^{-1}$.

Let the prediction error be a $G \times P$ matrix $\mathbf{Z}_f$, that is,

$$\mathbf{Z}_f = \mathbf{Y}_f - \hat{\mathbf{Y}}_f = (\boldsymbol{\Pi} - \hat{\boldsymbol{\Pi}})\mathbf{X}_f + \mathbf{V}_f = \mathbf{V}_f - \mathbf{VX}'(\mathbf{XX}')^{-1}\mathbf{X}_f\,, \tag{2-19}$$

then a $G \times 1$ column vector $\mathbf{z}_{fi}$ of $\mathbf{Z}_f$ can be written as

$$\mathbf{z}_{fi} = \mathbf{v}_{fi} - \mathbf{VX}'(\mathbf{XX}')^{-1}\mathbf{x}_{fi}\,, \qquad i = 1, \ldots, P\,, \tag{2-20}$$

where $\mathbf{v}_{fi}$ $(G \times 1)$ and $\mathbf{x}_{fi}$ $(K \times 1)$ denote the column vector of $\mathbf{V}_f$ and $\mathbf{X}_f$ respectively.

If we put

$$\mathbf{X}'(\mathbf{XX}')^{-1}\mathbf{x}_{fi} = \mathbf{A}_i\,, \qquad i = 1, \ldots, P\,,$$

then $\mathbf{A}_i$ is a $T \times 1$ column vector and we shall write $\mathbf{A}_i$ as

$$\mathbf{A}_i = \begin{bmatrix} a_1^i \\ \vdots \\ a_T^i \end{bmatrix}.$$

Thus $\mathbf{z}_{fi}$ can be rewritten as

$$\mathbf{z}_{fi} = \mathbf{v}_{fi} - \mathbf{VA}_i\,, \tag{2-21}$$

therefore

$$\begin{aligned} \operatorname{Cov}(\mathbf{z}_{fi}, \mathbf{z}_{fj}) &= E(\mathbf{z}_{fi}\mathbf{z}'_{fj}) \\ &= \left\{\left(\mathbf{v}_{fi} - \sum_{t=1}^{T} a_t^i \mathbf{v}_t\right)\left(\mathbf{v}_{fj} - \sum_{t=1}^{T} a_t^j \mathbf{v}_t\right)'\right\} \\ &= \delta_{ij}\boldsymbol{\Sigma} + \sum_{t=1}^{T} [a_t^i a_t^j E(\mathbf{v}_t\mathbf{v}'_t)] \\ &= \delta_{ij}\boldsymbol{\Sigma} + \mathbf{A}'_i\mathbf{A}_j\boldsymbol{\Sigma} \\ &= [\delta_{ij} + \mathbf{x}'_{fi}(\mathbf{XX}')^{-1}\mathbf{x}_{fj}]\boldsymbol{\Sigma} \\ &= \boldsymbol{\Sigma}_{f(i,j)} \end{aligned} \tag{2-22}$$

where δ_{ij} denotes the Kronecker delta. Hence we know that a $G \times 1$ vector $\mathbf{z}_{fi}$ is normally distributed with an expectation zero and with a variance-cavariance matrix $\boldsymbol{\Sigma}_{f(i,i)}$, and further $\mathbf{z}_{fi}$ and $\mathbf{z}_{fj}$ $(i \neq j)$ are not independently distributed. From (2–22) the variance-covariance matrix of $\mathbf{Z}_f$ can be easily obtained, that is,

$$\begin{aligned}\operatorname{Var}(\mathbf{Z}_f) &= E(\mathbf{Z}_f\mathbf{Z}'_f)\\ &= \left(\sum_{i=1}^{P}\mathbf{z}_{fi}\mathbf{z}'_{fi}\right)\\ &= \sum_{i=1}^{P}[1 + \mathbf{x}'_{fi}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{x}_{fi}]\boldsymbol{\Sigma}\\ &= [P + \operatorname{tr}\mathbf{X}'_f(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}_f]\boldsymbol{\Sigma}\\ &= (P + q)\boldsymbol{\Sigma}\,, \end{aligned} \tag{2–23}$$

where $q = \operatorname{tr}\mathbf{X}'_f(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}_f$.

If $\hat{\boldsymbol{\Sigma}}$ is the unbiased estimator of $\boldsymbol{\Sigma}$, then $\hat{\boldsymbol{\Sigma}}$ is obtained as

$$\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \hat{\boldsymbol{\Pi}}\mathbf{X})(\mathbf{Y} - \hat{\boldsymbol{\Pi}}\mathbf{X})'/(T - K)\,, \tag{2–24}$$

therefore the unbiased estimator of Var $(\mathbf{Z}_f)$ is given by

$$(P + q)\hat{\boldsymbol{\Sigma}}\,. \tag{2–25}$$

3. The Prediction Interval for the Simultaneous Equations System

We shall now get the prediction interval. From (2–24) we have

$$(\mathrm{T} - K)\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \hat{\boldsymbol{\Pi}}\mathbf{X})(\mathbf{Y} - \hat{\boldsymbol{\Pi}}\mathbf{X})' = \mathbf{V}\mathbf{M}\mathbf{V}'\,, \tag{2–26}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}$ and $\mathbf{M}$ is an idempotent matrix of rank $T - K$. Thus there exists a $T \times T$ orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{M}\mathbf{P} = \begin{bmatrix}\mathbf{I}_{T-K} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}.$$

If $\mathbf{V}$ is transformed to $\mathbf{W}$ by

$$\mathbf{W} = \mathbf{V}\mathbf{P}\,, \tag{2–27}$$

then

$$(T - K)\hat{\boldsymbol{\Sigma}} = \mathbf{W}\mathbf{P}'\mathbf{M}\mathbf{P}\mathbf{W}' = \sum_{\alpha=1}^{T-K}\mathbf{w}_\alpha\mathbf{w}'_\alpha\,, \tag{2–28}$$

where $\mathbf{w}_\alpha$ $(G \times 1)$ is a column vector of $\mathbf{W}$ $(G \times T)$. If we write

$$\mathbf{P} = \{p_{ij}\}$$

and $\mathbf{p}_i$ is to be the i-th column vector of $\mathbf{P}$, then from (2–27) we have

$$\mathbf{w}_\alpha = \mathbf{V}\mathbf{P}_\alpha = \sum_{i=1}^{T}p_{i\alpha}\mathbf{V}_i\,. \tag{2–29}$$

Hence

$$E(\mathbf{w}_\alpha) = \mathbf{0}\,, \qquad \alpha = 1, \ldots, T-K\,,$$

$$\begin{aligned}
\operatorname{Cov}(\mathbf{w}_\alpha, \mathbf{w}_\beta) &= E(\mathbf{w}_\alpha \mathbf{w}'_\beta) \\
&= E\left[\left(\sum_{i=1}^{T} p_{i\alpha}\mathbf{V}_i\right)\left(\sum_{j=1}^{T} p_{j\beta}\mathbf{V}_j\right)'\right] \\
&= \sum_{i=1}^{T}\sum_{j=1}^{T} p_{i\alpha}p_{j\beta}E(\mathbf{v}_i\mathbf{v}'_j) \\
&= \sum_{i=1}^{T}\sum_{j=1}^{T} p_{i\alpha}p_{j\beta}\delta_{ij}\boldsymbol{\Sigma} \\
&= \sum_{i=1}^{T} p_{i\alpha}p_{i\beta}\boldsymbol{\Sigma} \\
&= \delta_{\alpha\beta}\boldsymbol{\Sigma}\,,
\end{aligned} \tag{2-30}$$

where $\delta_{\alpha\beta}$ denotes the Kronecker delta, therefore $\mathbf{w}_\alpha$ and $\mathbf{w}_\beta$ are independently distributed as a normal variables with an expectation zero and a varance-covariance matrix $\boldsymbol{\Sigma}$. Thus if we write

$$\mathbf{u}_\alpha = (1 + q_\alpha)^{1/2}\mathbf{w}_\alpha\,, \tag{2-31}$$

where $q_\alpha = \mathbf{X}'_{f\alpha}(\mathbf{XX}')^{-1}\mathbf{X}_{f\alpha}$, then a $G \times 1$ vector $\mathbf{u}_\alpha$ is distributed according to

$$N(\mathbf{0}, (1 + q_\alpha)\boldsymbol{\Sigma}) = N(\mathbf{0}, \boldsymbol{\Sigma}_{f(\alpha,\alpha)}) \tag{2-32}$$

and $\mathbf{u}_\alpha$ and $\mathbf{u}_\beta$ $(\alpha \neq \beta)$ are independent. We have further

$$(T-K)(1+q_\alpha)\hat{\boldsymbol{\Sigma}} = (T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)} = \sum_{\alpha=1}^{T-K} \mathbf{u}_\alpha\mathbf{u}'_\alpha \tag{2-33}$$

and hence $(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)}$ is distributed according to Wishart distribution. This is denoted by

$$(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)} \sim W(G, T-K, \boldsymbol{\Sigma}_{f(\alpha,\alpha)})\,. \tag{2-34}$$

We shall next show that $\mathbf{z}_{f\alpha}$ and $(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)}$ are independent. We have

$$\mathbf{z}_{f\alpha} = \mathbf{v}_{f\alpha} + (\boldsymbol{\Pi} - \hat{\boldsymbol{\Pi}})\mathbf{X}_{f\alpha} = \mathbf{v}_{f\alpha} - \mathbf{V}Q\mathbf{X}_{f\alpha}\,, \tag{2-35}$$

where $Q = \mathbf{X}'(\mathbf{XX}')^{-1}$ and

$$(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)} = (T-K)(1+q_\alpha)\mathbf{VMV}'\,.$$

Since $\mathbf{V}Q = \sum_{i=T-k+1}^{T} \mathbf{w}_i\mathbf{P}'_i$ and $\mathbf{w}_i$ is distributed according to $N(\mathbf{0}, \boldsymbol{\Sigma})$ independently of $\mathbf{w}_j$ $(i \neq j)$, then $\mathbf{V}Q$ and $\mathbf{VMV}'$ are independent. Further since $\mathbf{VMV}' = \sum_{\alpha=1}^{T-K} \mathbf{w}_\alpha\mathbf{w}'_\alpha$ and $\mathbf{w}_\alpha$ is normally distributed, then the relation

$$E(\mathbf{w}_\beta\mathbf{V}'_{f\alpha}) = \sum_{i=1}^{T} p_{i\beta}E(\mathbf{V}_i\mathbf{V}'_{f\alpha}) = \mathbf{0}\,, \tag{2-36}$$

means that $\mathbf{VMV}'$ and $\mathbf{V}_{f\alpha}$ are independent. Thus the independency between

$\mathbf{z}_{f\alpha}$ and $(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)}$ has been proved. As above proved the following relations are satisfied.

$$\begin{aligned}&(1)\quad \mathbf{z}_{f\alpha} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{f(\alpha,\alpha)}),\\ &(2)\quad (T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)} \sim W(G, T-K, \boldsymbol{\Sigma}_{f(\alpha,\alpha)}),\\ &(3)\quad \mathbf{z}_{f\alpha} \text{ is independent of } \hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)}.\end{aligned} \tag{2-37}$$

From (2–37) we have

$$\begin{aligned}\eta &= \mathbf{z}'_{f\alpha}[(T-K)\hat{\boldsymbol{\Sigma}}_{f(\alpha,\alpha)}]^{-1}\mathbf{z}_{f\alpha}\cdot(T-K-G+1)/G\\ &= \mathbf{z}'_{f\alpha}\hat{\boldsymbol{\Sigma}}^{-1}_{f(\alpha,\alpha)}\mathbf{z}_{f\alpha}(T-K-G+1)/G(T-K)\\ &\sim F(G, T-K-G+1)\,, \qquad \alpha = 1,\ldots,P\,.\end{aligned} \tag{2-38}$$

If we put $G = 1$ in (2–38) we obtain (2–13).

4. Joint Confidence Interval for $\mathbf{a}'\mathbf{y}_{fi}$

Let $F_\alpha(G, T-K-G+1)$ be the upper significance point corresponding to the α significance level and with G and $T-K-G+1$ degrees of freedom. Wd shall denote this point by F_α. Then from (2–38) we have the probability statement

$$P\left\{\mathbf{z}'_{fi}\hat{\boldsymbol{\Sigma}}_{f(i,i)}\mathbf{z}_{fi}\frac{T-K-G+1}{G(T-K)} \leqq F_\alpha\right\} \geqq 1-\alpha\,. \tag{2-39}$$

This relation can be rewritten as

$$P\{\mathbf{z}'_{fi}\hat{\boldsymbol{\Sigma}}_{f(i,i)}\mathbf{z}_{fi} \leqq mF_\alpha\} \geqq F_\alpha\,, \tag{2-40}$$

where $m = G(T-K)/(T-K-G+1)$.

Let $\mathbf{a}$ be a $G\times 1$ any non-zero vector, then (2–40) is equivalent to

$$P\left\{\frac{\mathbf{a}'\mathbf{z}_{fi}\mathbf{z}'_{fi}\mathbf{a}}{\mathbf{a}'\hat{\boldsymbol{\Sigma}}_{f(i,i)}\mathbf{a}} \leqq mF_\alpha\right\} \geqq 1-\alpha\,. \tag{2-41}$$

Let $\mathbf{y}_{fi}$ and $\hat{\mathbf{y}}_{fi}$ be the i-th column vector in $\mathbf{Y}_f$ and $\hat{\mathbf{Y}}_f$ respectively. Then from (2–41) we have

$$P\left\{\frac{|\mathbf{a}'(\mathbf{y}_{fi}-\hat{\mathbf{y}}_{fi})|}{(\mathbf{a}'\hat{\boldsymbol{\Sigma}}_{f(i,i)}\mathbf{a})^{1/2}} \leqq (mF_\alpha)^{1/2}\right\} \geqq 1-\alpha$$

and hence the joint confidence interval for $\mathbf{a}'\mathbf{y}_{fi}$ can be obtained as

$$P\{\mathbf{a}'\mathbf{y}_{fi} \in \mathbf{a}'\mathbf{y}_{fi} \pm (mF_\alpha\cdot\mathbf{a}'\hat{\boldsymbol{\Sigma}}_{f(i,i)}\mathbf{a})^{1/2}\} \geqq 1-\alpha\,, \tag{2-42}$$

for all $\mathbf{a}$. Let $y^i_{f\beta}$ and $\hat{y}^i_{f\beta}$ be the i-th element in a $G\times 1$ vector $\mathbf{y}_{f\beta}$ and $\hat{\mathbf{y}}_{f\beta}$ respectively and

$$\mathbf{a}' = (0, 0, \ldots, \overset{i-1}{0}, \overset{i}{1}, \overset{i+1}{0}, \ldots, 0)\,,$$

then we have

$$\mathbf{a}'\hat{\boldsymbol{\Sigma}}_{f(\beta,\beta)}\mathbf{a} = \hat{\sigma}_{ii}(1+\mathbf{X}'_{f\beta}(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}_{f\beta})\,,$$

therefore from (2–42) the confidence interval for $y^i_{f\beta}$ can be obtained as

$$P\{y^i_{f\beta} \in \hat{y}^i_{f\beta} \pm [\hat{\sigma}_{ii}(1 + \mathbf{X}'_{f\beta}(\mathbf{XX}')^{-1}\mathbf{X}_{f\beta})mF_\alpha]^{1/2}\} \geqq 1 - \alpha ,$$

where $\hat{\sigma}_{ii}$ denotes the (i, i) element in $\hat{\Sigma}$. If $G = 1$ then $m = 1$ and (2–43) is equal to (2–12).

III. CONCLUSIONS

The matrix $\mathbf{X}$ of independent variables is assumed to be nonstochastic in our analysis. This assumption means that $\mathbf{X}$ does not include the lagged endogeneous variables, so it is very simple assumption as econometric model. Hence the important problem still remains that the analysis of the prediction errors have to be extended to the model which include the lagged endogeneous variables as the independent variables.

The variance and covariance matrix of the prediction errors in the case of deriving the reduced form parameters from consistently estimated structural parameters is given in [5] and using the asymptotic distribution of the prediction errors the prediction interval have been derived in [8]. This paper does not refer to such a case. The problem still remains in this respect.

Keio University

REFERENCES

[1] Anderson, T. W. (1958): *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc.

[2] Brown, T. M. (1954): "Standard Errors of Forecast of a Complete Econometric Model", Econometrica, Vol. 22, pp. 178–192.

[3] Cramer H. and M. R. Leadbetter (1967): *Stationary and Related Stochastic Processes*, John Wiley & Sons, Inc.

[4] Goldberger A. S. (1962): "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", *Journal of the American Statistical Association*, Vol. 57, pp. 369–375.

[5] Goldberger A. S., A. L. Nagar and H. S. Odeh (1961): "The Covariance Matrices of Reduced-Form Coefficients and of Forecasts for a Structural Econometric Model", *Econometrica*, Vol. 29, pp. 556–573.

[6] Graybill F. A. (1961): *An Introduction to Linear Statistical Models*, Volume I, McGraw-Hill Book Company, Inc.

[7] Hooper J. W. and A. Zellner (1961): "The Error of Forecast for Multivariate Regression Models", *Econometrica*, Vol. 29, pp. 544–555.

[8] Hymans S. H. (1968): "Simultaneous Confidence Intervals in Econometric Forecasting", *Econmetrica*, Vol. 36, pp. 18–30.

[9] Koerts J. and A. P. J. Abrahamse (1969): *On the Theory and Application of the General Linear Model*, Rotterdom the Univ. Press.

[10] Koichi Ito (1969): *Theory of Multivariate Analysis* (in Japanese), Baifu Kan.

[11] Masani P. (1966): "Recent Trends in Multivariate Prediction Theory", *Multivariate Analysis*, Academic Press, pp. 351–382.

[12] Rao C. R. (1965): *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Inc.
[13] Scheffé H. (1959): *Analysis of Variance*, John Wiley & Sons, Inc.
[14] Takamitsu Sawa (1970): *Foundations of Econometrics*, (in Japanese), Toyo Keizai.
[15] Toshio Kitagawa (1965): *Stochastic Process*, (in Japanese), Kyoritsu Shuppan.