Title	A note on approximation of prediction error of a proposed predictor with an estimated autoregressive parameter in a linear model
Sub Title	
Author	Toyooka, Yasuyuki(Tanimoto, Shin) 谷本, 信
Publisher	慶應義塾大学理工学部
Publication year	1983
Jtitle	Keio Science and Technology Reports Vol.36, No.6 (1983. 12) ,p.57- 64
JaLC DOI	
Abstract	We propose a predictor for one step ahead observation of the statistical linear model with an error of the first-order stationary autoregressive process from the viewpoint of the generalized least squares estimation. The predictor is constructed by making use of an estimator of the autoregressive parameter based on the past observations of the process. We give an expansion up to O(1/T) for the prediction mean squared error of our predictor and show that this expansion coincides with that of the predictor with an estimator of autoregressive parameter which is independent of the past observations.
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO50001004-00360006-0057

KEIO SCIENCE AND TECHNOLOGY REPORTS
VOL. 36, NO. 6, pp. 57-64, 1983

A NOTE ON APPROXIMATION OF PREDICTION ERROR OF A PROPOSED PREDICTOR WITH AN ESTIMATED AUTOREGRESSIVE PARAMETER IN A LINEAR MODEL

Yasuyuki TOYOOKA*

Department of Applied Mathematics
Faculty of Engineering Science
Osaka University
Toyonaka-shi 560, Osaka, Japan

Shin TANIMOTO

Department of Administration Engineering
Faculty of Science and Technology
Keio University
Yokonama-shi 223, Kanagawa, Japan

(*Received 11. 16. 1983*)

SUMMARY

We propose a predictor for one step ahead observation of the statistical linear model with an error of the first-order stationary autoregressive process from the viewpoint of the generalized least squares estimation. The predictor is constructed by making use of an estimator of the autoregressive parameter based on the past observations of the process. We give an expansion up to $O(1/T)$ for the prediction mean squared error of our predictor and show that this expansion coincides with that of the predictor with an estimator of autoregressive parameter which is independent of the past observations.

§ 1. Introduction

Consider a model,

* This research for Y. Toyooka is supported in part by the Ministry of Education of Japan under "Shorei" A57740120.
AMS 1970 subject classifications: Primary 62F10 secondary 62J10.
Key words and phrases: Expansion of prediction error, GLSE, Linear model with an autoregressive process.

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t \qquad t=1,2,\cdots$$
$$u_t = \theta u_{t-1} + \varepsilon_t \qquad t=\cdots,-1,0,1,\cdots, \tag{1.1}$$

where $\{y_t\}$ is a sequence of observable random variables, $\{\boldsymbol{x}_t\}$ a sequence of p-dimensional fixed variate explanatory vectors, $\boldsymbol{\beta}$ a p-dimensional vector of unknown parameters, $\{\varepsilon_t\}$ a sequence of independently identically normally distributed random variables with mean zeros and finite variances σ^2 for all t and θ an autoregressive parameter with $|\theta|<1$. In this model $\{u_t\}$ is a stationary ergodic first-order autoregressive process.

The problem is to predict one step ahead observation y_{T+1} when $y_1, y_2, \cdots, y_T$ are observed. For this purpose we rewrite the model (1.1) in the matrix notation

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}, \tag{1.2}$$

where $\boldsymbol{y}=(y_1, y_2, \cdots, y_T)'$, $X=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T)'$ and $\boldsymbol{u}=(u_1, u_2, \cdots, u_T)'$. Further we define the variance-covariance matrix of $\boldsymbol{u}$ and serial covariance vector of $\boldsymbol{u}$ and u_{T+1} as follows,

$$V(\theta) = E(\boldsymbol{u}\boldsymbol{u}')$$

and

$$\boldsymbol{v}(\theta) = E(u_{T+1}\boldsymbol{u}).$$

For simplicity, we assume that X is of full rank for all T with $T \geqq p$. Furthermore we assume that each element of $\boldsymbol{x}_t$ is bounded in t.

In order to construct a predictor of y_{T+1}, first, we consider the best linear unbiased predictor (BLUP) $\hat{y}^*_{T+1}$ of y_{T+1} in rather restrictive situation where θ is known and only $\boldsymbol{\beta}$ is unknown, that is,

$$\hat{y}^*_{T+1} = \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_w + \boldsymbol{v}(\theta)' V(\theta)^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_w),$$

where $\hat{\boldsymbol{\beta}}_w = \{X' V(\theta)^{-1} X\}^{-1} X' V(\theta)^{-1}\boldsymbol{y}$. However we are interested in more realistic situation where both θ and $\boldsymbol{\beta}$ are unknown. Therefore we construct a predictor $\hat{y}_{T+1}$ which is calculated by the following procedure. At the first stage, we estimate θ from the simple least squares residuals,

$$\hat{\theta} = \sum_{t=1}^{T-1} \hat{u}_t \hat{u}_{t+1} \Big/ \sum_{t=1}^{T-1} \hat{u}_t^2, \tag{1.3}$$

where $\hat{u}_t = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}$ $(t=1,2,\cdots,T)$ and $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$. Then, by making use of $\hat{\theta}$, we construct a predictor which is analogous to the BLUP, that is,

$$\hat{y}_{T+1} = \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_{\hat{w}} + \boldsymbol{v}(\hat{\theta})' V(\hat{\theta})^{-1}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_{\hat{w}}),$$

where $\hat{\boldsymbol{\beta}}_{\hat{w}} = \{X' V(\hat{\theta})^{-1} X\}^{-1} X' V(\hat{\theta})^{-1}\boldsymbol{y}$. Here we note that $\boldsymbol{y}$ and $\hat{\theta}$ are not independent. The purpose of this note is to give an expansion up to $O(1/T)$ of prediction mean squared error (we call this a prediction error) for $\hat{y}_{T+1}$ and to investigate the properties of $\hat{y}_{T+1}$.

Baillie (1979) obtained the expansion of the prediction error for the predictor with estimated parameters when an error process is a realization of the finite order autoregressive process from the viewpoint of the maximum likelihood estimation. On the other hand, Toyooka (1982b) also obtained it when an error process has more general structure, that is, its variance-covariance matrix is described with finite structural parameters from the point of the generalized least squares estimation. But the latter assumed that the estimators of parameters involved in the error process were obtained by another observation and were independent of $\boldsymbol{y}$. In this note, we treat the first-order autoregressive process as the error process, but we construct the estimator for the autoregressive parameter from the past observation $\boldsymbol{y}$ by (1.3). Therefore, $\hat{\theta}$ is not independent of $\boldsymbol{y}$ and there is a certain correlation between the two. This correlation may affect the prediction error for our predictor. But it is shown in this paper that the expansion of the prediction error coincides with those given by Baillie (1979) and Toyooka (1982b) up to $O(1/T)$ and Baillie conjectured this proposition in his paper published in 1979. So this paper can be viewed as a mathematical refinement of it. Moreover Baillie (1981) treated the more general model such as the dynamic simultaneous equation model with vector autoregressive errors.

In § 2, we prepare some lemmas due to Toyooka (1982b), (1982c), and give a stochastic expansion of $\hat{y}^*_{T+1}-\hat{y}_{T+1}$. In § 3, we evaluate the expansion of the prediction error for our predictor and discuss a statistical implication of the expansion.

§ 2 Stochastic Expansion of the Predictor

We decompose the prediction error of our predictor into three terms in order to expand it up to $O(1/T)$.

$$E(y_{T+1}-\hat{y}_{T+1})^2=E(y_{T+1}-\hat{y}^*_{T+1})^2+2E(y_{T+1}-\hat{y}^*_{T+1})(\hat{y}^*_{T+1}-\hat{y}_{T+1})$$
$$+E(\hat{y}^*_{T+1}-\hat{y}_{T+1})^2 . \tag{2.1}$$

The difference between y_{T+1} and $\hat{y}^*_{T+1}$ can be written as

$$y_{T+1}-\hat{y}^*_{T+1}=\{u_{T+1}-\boldsymbol{v}(\theta)' V(\theta)^{-1}\boldsymbol{u}\}$$
$$-\{\boldsymbol{x}'_{T+1}-\boldsymbol{v}(\theta)' V(\theta)^{-1}X\}(\hat{\boldsymbol{\beta}}_w-\boldsymbol{\beta}) . \tag{2.2}$$

The first term is of $O_p(1)$ and the second term is of $O_p(1/\sqrt{T})$. Then it is sufficient to evaluate $\hat{y}^*_{T+1}-\hat{y}_{T+1}$ only up to $O_p(1/T)$ to obtain an expansion of the prediction error of $\hat{y}_{T+1}$.

First we get the stochastic expansion of $\hat{y}^*_{T+1}-\hat{y}_{T+1}$ by using the Taylor expansion around θ when we regard $\hat{y}_{T+1}$ as a function of an estimator $\hat{\theta}$.

Lemma 1

Suppose that we get a consistent estimator $\tilde{\theta}$ *of* θ *such that* $\tilde{\theta}-\theta=O_p(1/\sqrt{T})$. *Under the model* (1.1), *a stochastic expansion of* $\hat{y}^*_{T+1}-\hat{y}_{T+1}$ *is obtained as follows,*

$$\hat{y}^*_{T+1}-\hat{y}_{T+1}=-(\tilde{\theta}-\theta)\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)' V(\theta)\}^{-1}\boldsymbol{u}$$
$$-\{\boldsymbol{x}'_{T+1}-\boldsymbol{v}(\theta)' V(\theta)^{-1}X\}(\hat{\boldsymbol{\beta}}_{\tilde{w}}-\hat{\boldsymbol{\beta}}_w)$$

$$+(\tilde{\theta}-\theta)\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}X(\hat{\boldsymbol{\beta}}_w-\beta)$$

$$-\frac{1}{2}(\tilde{\theta}-\theta)^2\frac{\partial^2}{\partial\theta^2}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}\boldsymbol{u}$$

$$+o_p(1/T) \qquad as\ T\to\infty\,.$$

Here we note that the stochastic expansion of $\hat{\boldsymbol{\beta}}_{\hat{w}}-\hat{\boldsymbol{\beta}}_w$ given by Toyooka (1982a) is

$$\hat{\boldsymbol{\beta}}_{\hat{w}}-\hat{\boldsymbol{\beta}}_w=-(\tilde{\theta}-\theta)\{X'V(\theta)^{-1}X\}^{-1}X'\frac{\partial}{\partial\theta}V(\theta)^{-1}X(\hat{\boldsymbol{\beta}}_w-\boldsymbol{\beta})$$

$$+(\tilde{\theta}-\theta)\{X'V(\theta)^{-1}X\}^{-1}X'\frac{\partial}{\partial\theta}V(\theta)^{-1}\boldsymbol{u}$$

$$+o_p(1/T) \qquad \text{as } T\to\infty\,.$$

Secondly we consider the stochastic expansion of $\hat{\theta}-\theta$. Let $\hat{\theta}_{LS}$ be

$$\hat{\theta}_{LS}=\sum_{t=1}^{T-1}u_tu_{t+1}\Big/\sum_{t=1}^{T-1}u_t^2\,.$$

We decompose $\hat{\theta}-\theta$ into two terms, that is,

$$\hat{\theta}-\theta=(\hat{\theta}-\hat{\theta}_{LS})+(\hat{\theta}_{LS}-\theta)\,. \tag{2.3}$$

Two terms of (2.3) are evaluated by the following two lemmas due to Toyooka (1982b), (1982c).

Lemma 2 (Toyooka (1982c))

Let $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ be,

$$\boldsymbol{Z}_1=\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\,,$$

$$\boldsymbol{Z}_2=(X'X)^{-1}\sum_{t=1}^{T-1}\boldsymbol{x}_tu_{t+1}$$

and

$$\boldsymbol{Z}_3=(X'X)^{-1}\sum_{t=1}^{T-1}\boldsymbol{x}_{t+1}u_t\,.$$

Assume that there exist two matrices $C(0)$ and $C(1)$ such that

$$\frac{1}{T}\sum_{t=1}^{T-1}\boldsymbol{x}_t\boldsymbol{x}_t'=C(0)+o(1)$$

and

$$\frac{1}{T}\sum_{t=1}^{T-1}\boldsymbol{x}_{t+1}\boldsymbol{x}_t'=C(1)+o(1) \qquad as\ T\to\infty\,.$$

Under the model(1.1), *we get the following stochastic expansion,*

$$\hat{\theta}-\hat{\theta}_{LS}=-\frac{1-\theta^2}{\sigma^2}\boldsymbol{Z}_1'[\{\theta C(0)+C(1)\}\boldsymbol{Z}_1-C(0)\boldsymbol{Z}_2-C(0)\boldsymbol{Z}_3]+o_p(1/T) \quad \text{as } T\to\infty .$$

Lemma 3 (Toyooka (1982c))

Under the model (1.1), *we get the following stoschastic expansion*

$$\hat{\theta}_{LS}-\theta=\frac{1}{\sqrt{T}}\frac{1-\theta^2}{\sigma^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}u_t\varepsilon_{t+1}$$
$$-\frac{1}{T}\left(\frac{1-\theta^2}{\sigma^2}\right)^2\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}u_t\varepsilon_{t+1}\right)\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-1}u_t^2-\frac{\sigma^2}{1-\theta^2}\right)$$
$$+o_p(1/T) \quad \text{as } T\to\infty .$$

Therefore we get the following stochastic expansion of $\hat{\theta}-\theta$,

$$\hat{\theta}-\theta=\mathrm{a}/\sqrt{T}+b/T+o_p(1/T) \quad \text{as } T\to\infty ,$$

where

$$a=\frac{1-\theta^2}{\sigma^2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}u_t\varepsilon_{t+1}$$

and

$$b=\frac{1-\theta^2}{\sigma^2}T\boldsymbol{Z}_1'[\{\theta C(0)+C(1)\}\boldsymbol{Z}_1-C(0)\boldsymbol{Z}_2-C(0)\boldsymbol{Z}_3]$$
$$-\left(\frac{1-\theta^2}{\sigma^2}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T-1}u_t\varepsilon_{t+1}\right)\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-1}u_t^2-\frac{\sigma^2}{1-\theta^2}\right).$$

Consequently we get the stochastic expansion up to $O_p(1/T)$ of $\hat{y}^*_{T+1}-\hat{y}_{T+1}$ as follows by using the stochastic expansion of $\hat{\theta}-\theta$ and Lemma 1,

$$\hat{y}^*_{T+1}-\hat{y}_{T+1}=-\frac{a}{\sqrt{T}}\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}\boldsymbol{u}$$
$$-\frac{b}{T}\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}\boldsymbol{u}$$
$$+\frac{a}{\sqrt{T}}\{\boldsymbol{x}_{T+1}'-\boldsymbol{v}(\theta)'V(\theta)^{-1}X\}\{X'V(\theta)^{-1}X\}^{-1}\cdot$$
$$X'\frac{\partial}{\partial\theta}V(\theta)^{-1}X\{(\hat{\boldsymbol{\beta}}_w-\boldsymbol{\beta})-\boldsymbol{u}\}$$
$$+\frac{a}{\sqrt{T}}\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}X(\hat{\boldsymbol{\beta}}_w-\boldsymbol{\beta})$$
$$-\frac{a^2}{2T}\frac{\partial^2}{\partial\theta^2}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}\boldsymbol{u}$$
$$+o_p(1/T) \quad \text{as } T\to\infty . \tag{2.4}$$

The first term is of $O_p(1/\sqrt{T})$, the terms from second to fifth are of $O_p(1/T)$ and the sixth term vanishes in this situation.

§ 3. Expansion of Prediction Error

We evaluate the prediction error of our predictor $\hat{y}_{T+1}$. The first term of (2.1) is the prediction error of the BLUP, and is

$$E(y_{T+1}-\hat{y}^*_{T+1})^2=\sigma^2+\{\boldsymbol{x}'_{T+1}-\boldsymbol{v}(\theta)'V(\theta)^{-1}X\}Cov(\hat{\boldsymbol{\beta}}_w)\cdot$$
$$\{\boldsymbol{x}_{T+1}-X'V(\theta)^{-1}\boldsymbol{v}(\theta)\}$$
$$=\sigma^2+(\boldsymbol{x}'_{T+1}-\theta\boldsymbol{x}'_T)Cov(\hat{\boldsymbol{\beta}}_w)(\boldsymbol{x}_{T+1}-\theta\boldsymbol{x}_T)\,. \qquad (3.1)$$

In order to evaluate the second term of (2.1), we write the first term of (2.2) as follows,

$$u_{T+1}-\boldsymbol{v}(\theta)'V(\theta)^{-1}\boldsymbol{u}=u_{T+1}-\theta u_T$$
$$=\varepsilon_{T+1}\,.$$

Since ε_{T+1} and $\boldsymbol{u}$ are mutually independent and all random variables in (2.4) are functions of $\boldsymbol{u}$,

$$E\{(u_{T+1}-\boldsymbol{v}(\theta)'V(\theta)^{-1}\boldsymbol{u})(\hat{y}^*_{T+1}-\hat{y}_{T+1})\}=0\,.$$

So the term of b in $\hat{\theta}-\theta$ does not affect this expansion. We evaluate the prediction error up to $O(1/T)$. It is sufficient that we calculate the expectation of the product of the second term of (2.2) and the first term of (2.4).

$$E\left[\{\boldsymbol{x}'_{T+1}-\boldsymbol{v}(\theta)'V(\theta)^{-1}X\}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\frac{a}{\sqrt{T}}\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)^{-1}\}\boldsymbol{u}\right]=o(1/T)$$
$$\text{as } T\to\infty\,.$$

Generally this term is $O(1/T)$ and if $\hat{\theta}$ and $\boldsymbol{y}$ are mutually independent then this is trivially $o(1/T)$ since the independence of a with $\boldsymbol{u}$. But even if $\hat{\theta}$ is constructed by $\boldsymbol{u}$ this term is $o(1/T)$ because of the structure of $\boldsymbol{v}(\theta)'V(\theta)^{-1}$ in the autoregression. This very fact is essential in our discussion. Therefore the second term in (2.1) is of $o(1/T)$ and it does not contribute the evaluation of prediction error up to $O(1/T)$.

Lastly we evaluate $E(\hat{y}^*_{T+1}-\hat{y}_{T+1})^2$. The term which contributes the evaluation of $E(\hat{y}^*_{T+1}-\hat{y}_{T+1})^2$ up to $O(1/T)$ is only the expectation of the square of the first term of (2.4).

$$E\left[-\frac{a}{\sqrt{T}}\frac{\partial}{\partial\theta}\{\boldsymbol{v}(\theta)'V(\theta)\}^{-1}\boldsymbol{u}\right]^2=\frac{1}{T}E(a^2u_T^2)$$
$$=\frac{1}{T}\sigma^2+O(1/T^2) \qquad \text{as } T\to\infty\,.$$

Hence we establish the following theorem.

Theorem

Under the model (1.1), *assume that there exist two matrices* $C(0)$ *and* $C(1)$ *such that as* $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T-1} \boldsymbol{x}_t \boldsymbol{x}_t' = C(0) + o(1)$$

and

$$\frac{1}{T}\sum_{t=1}^{T-1} \boldsymbol{x}_{t+1} \boldsymbol{x}_t' = C(1) + o(1) .$$

Then as $T \to \infty$,

$$E(y_{T+1} - \hat{y}_{T+1})^2 = \sigma^2 + (\boldsymbol{x}'_{T+1} - \theta \boldsymbol{x}'_T) Cov(\hat{\boldsymbol{\beta}}_w)(\boldsymbol{x}_{T+1} - \theta \boldsymbol{x}_T)$$
$$+ \frac{\sigma^2}{T} + o(1/T) . \qquad (3.2)$$

Remark

The second term of (3.2) is the effect of the estimation of $\boldsymbol{\beta}$ and the third one of (3.2) is that of θ. And this expansion is the same as that of Toyooka (1982b). From the proof of the theorem, the expansion (3.2) is valid as long as $\hat{\theta}$ is so-called first-order efficient estimator of θ. In fact $\hat{\theta}$ and the maximum likelihood estimator $\hat{\theta}_{MLE}$ are BAN estimators. Though there is a difference in the second order expansion between $\hat{\theta}$ and $\hat{\theta}_{MLE}$, there is no difference from the viewpoint of the prediction of y_{T+1}. But it seems that the higher order terms of the predictor $\hat{y}_{T+1}$, higher than the term up to $O(1/T)$, depend on the choice of the estimator for θ. On the other hand, in the estimation problem of $\boldsymbol{\beta}$ in (1.2) Toyooka (1982c) discussed the second order comparisons of the risk matrices of the simple least squares estimator with the generalized least squares estimator and with the maximum likelihood estimator for $\boldsymbol{\beta}$. It showed that the maximum likelihood estimator can be viewed as a generalized least squares estimator and that the risk matrix of the generalized least squares estimator does not depend on the choice of the estimators for θ up to $O(1/T^2)$ as long as the estimator is BAN estimator. In this sense our result is a prediction version of Toyooka (1982c).

Acknowledgement

The authors would like to thank Prof. YASUTOSHI WASHIO, Keio University for his many conscientious suggestions and encourangements. Thanks are also due to the referee for his constructive and helpful comments.

REFERENCES

[1] Baillie, R.T. (1979) The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors, *J. Amer. Statist. Assoc.*, **74**, 175–184.

[2] Baillie, R.T. (1981) Prediction from the dynamic simultaneous equation model with vector autoregressive errors, *Econometrica*, **49**, 1331-1337.

[3] Toyooka, Y. (1982a) Second-order expansion of mean squared error matrix of generalized least squares estimator with estimated parameters, *Biometrika*, **69**, 269-273.

[4] Toyooka, Y. (1982b) Prediction error in a linear model with estimated parameters, *Biometrika*, **69**, 453-459.

[5] Toyooka, Y. (1982c) Second-order risk comparison of SLSE with GLSE and MLE in a regression with serial correlation (to appear).