Statistica Sinica 9(1999), 247-261

AN ASYMPTOTICALLY HONEST PREDICTION SET FOR THE MULTIVARIATE REGRESSION MODEL WHEN THE RESPONSE IS MEASURED WITH ERROR

Steve Y. H. Huang and Longcheen Huwang

National Tsing Hua University

Abstract: This paper constructs an asymptotically honest prediction set for the response variable, measured with error, in the multivariate regression model. By asymptotic honesty we mean that the limit inferior of the infimum coverage probability over the parameter space converges to the nominal level as the sample size goes to ∞. In the univariate case a desirable property of the length of this asymptotically honest prediction interval is obtained. A small simulation study shows that the coverage probability of this prediction set is close to the nominal level in the finite sample as well. Finally, we show that in errors-in-variables models, calibration and prediction problems can be solved by treating the models as special cases of the aforementioned model when the errors and the calibrated variables are assumed to be normally distributed.

Key words and phrases: Coverage probability, prediction set, regression model.

1. Introduction

It often happens that variables of interest are difficult or expensive to obtain and we replace them by variables obtained by some quick or cheap methods, but measured with errors. There are a lot of literatures concerning measurement error models and most of them are interested in problems with error in regressor. See for example Kendall and Stuart (1979), Anderson (1984), Fuller (1987), Cheng and Van Ness (1994), and Carroll, Ruppert and Stefanski (1995). Errors in response variables have received less attention because the inference of the major parameters can be handled by standard methodology when the measurement error is additive (see Buonaccorsi (1996)). But, if the objective is the prediction set (or interval) for the true response variable, the problem can not be resolved by the standard approach since the intuitive estimator of the variance of the true response is not a reasonable one. As a result, the coverage probability of the usual prediction set is appreciably below the nominal level for some values of parameters, even for large sample sizes.

Consider the usual multivariate regression model

$$\mathbf{Y}_i = \boldsymbol{\beta}\boldsymbol{X_i} + \boldsymbol{\epsilon_i}, i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta}$ is the p by q unknown coefficient matrix, $\mathbf{X}_i$ are q-dimensional column vectors with first component 1, $\boldsymbol{\epsilon_i}$ are p-dimensional equation errors, and $n \geq p+q$. Here it is assumed that the matrix $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is of rank q if $\mathbf{X}_i$ are fixed vectors, and that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is of rank q almost surely if $\mathbf{X}_i$ are i.i.d. continuous vector random variables. However, in the above setting we can not observe $\mathbf{Y}_i$ exactly. Instead we only observe $\mathbf{U}_i$ which is the true response $\mathbf{Y}_i$ plus an additive error $\boldsymbol{\delta}_i$, i.e.,

$$\mathbf{U}_i = \mathbf{Y}_i + \boldsymbol{\delta}_i, \; i = 1, \ldots, n. \tag{1.2}$$

Model (1.1)-(1.2) is called a measurement error model in the response. It is assumed that $(\boldsymbol{\epsilon_i}, \boldsymbol{\delta_i})$ are i.i.d. $N[\mathbf{0}, \; diag(\boldsymbol{\Sigma}_{\epsilon\epsilon}, \boldsymbol{\Sigma}_{\delta\delta})]$ and $\boldsymbol{\Sigma}_{\delta\delta}$ is known, where $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ and $\boldsymbol{\Sigma}_{\delta\delta}$ are the covariance matrices of $\boldsymbol{\epsilon_i}$ and $\boldsymbol{\delta_i}$, respectively. Knowledge of the covariance matrix $\boldsymbol{\Sigma}_{\delta\delta}$ of the measurement error $\boldsymbol{\delta_i}$ can come either from previous experiments or from repeated measurements on $\mathbf{Y}_i$. It is also allowed that some components of $\mathbf{Y}_i$ are measured exactly and some are measured with error. Hence the covariance matrix $\boldsymbol{\Sigma}_{\delta\delta}$ is only assumed non-negative definite. The present work focuses on a prediction problem where an unobservable future true response $\mathbf{Y}_{n+1}$ is to be predicted:

$$\mathbf{Y}_{n+1} = \boldsymbol{\beta}\mathbf{X}_{n+1} + \boldsymbol{\epsilon}_{n+1}. \tag{1.3}$$

We are interested in constructing a prediction set for $\mathbf{Y}_{n+1}$ based on $\mathbf{X}_{n+1}$ and $(\mathbf{U}_i, \mathbf{X}_i)$, $1 \leq i \leq n$. For instance, suppose that we have two variables $\mathbf{X}$ and $\mathbf{Y}$ satisfying the relation (1.1) where $\mathbf{X}$ can be observed exactly but $\mathbf{Y}$ cannot. In experiments, $\mathbf{X}$ is usually the value pre-assigned by the experimenters and $\mathbf{Y}$, which is difficult to obtain, represents the outcome of an expensive and time consuming method. Therefore a surrogate $\mathbf{U}$ of $\mathbf{Y}$, $\mathbf{U} = \mathbf{Y} + \boldsymbol{\delta}$, is observed instead by some other inexpensive and quick method. Based on a new $\mathbf{X}$ and the training data, we want to construct a prediction set for the true response $\mathbf{Y}$ associated with $\mathbf{X}$.

Throughout the paper, a $1-\alpha$ $(0 < \alpha < 1)$ prediction set R for $\mathbf{Y}_{n+1}$ is said to be asymptotically honest if

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} P(\mathbf{Y}_{n+1} \in R) = 1 - \alpha, \tag{1.4}$$

where $\boldsymbol{\theta}$ is the vector consisting of all parameters and Ω is the corresponding parameter space.

In the rest of this section, we demonstrate the deficiency of the coverage probability of the traditional prediction set. The discussion will focus on the univariate model ($p = 1$ and $q = 2$), the general case will be treated in Section 2. For now, model (1.1)-(1.3) reduces to

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \; U_i = Y_i + \delta_i, \quad i = 1, \ldots, n+1, \tag{1.5}$$

where $Y_i, X_i, \epsilon_i, U_i$, and δ_i are all one-dimensional. In the situation where there is no measurement error δ_i in (1.5), a $1-\alpha$ prediction interval for Y_{n+1} is given by

$$\Big\{Y_{n+1} : \frac{(Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1})^2}{[1 + \frac{1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_1^n (X_i - \bar{X})^2}]\hat{\sigma}_\epsilon^2} \le F_{n-2}^1(\alpha)\Big\}, \tag{1.6}$$

where $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$, $\hat{\beta}_1 = \sum_1^n (X_i - \bar{X})Y_i / \sum_1^n (X_i - \bar{X})^2$, $\hat{\sigma}_\epsilon^2 = \sum_1^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2/(n-2)$, and $F_{n-2}^1(\alpha)$ is the $1-\alpha$ quantile of the F distribution with 1 and $n-2$ degrees of freedom. The prediction interval (1.6) has coverage probability $1-\alpha$. When the response variable Y_i is measured by some instrument and there exists a measurement error δ_i, the approach to the estimation of β_0 and β_1 is the same (cf. Huwang (1996)) as if there is no measurement error in U_i. Now, however, the prediction interval for Y_{n+1} needs to be modified in order to satisfy (1.4).

It is easy to show that the conditional distribution of $Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1}$ given $X_1, \ldots, X_{n+1}$ is

$$N(0, \sigma_{Y|X}^2 + H\sigma_{U|X}^2), \tag{1.7}$$

where $\sigma_{Y|X}^2$ and $\sigma_{U|X}^2$ are respectively the conditional variances of Y_i and U_i given X_i, and $H = n^{-1} + (X_{n+1} - \bar{X})^2 / \sum_1^n (X_i - \bar{X})^2$. Intuitively, one should use

$$\hat{\sigma}_{Y|X}^2 = \hat{\sigma}_{U|X}^2 - \sigma_\delta^2 \tag{1.8}$$

to estimate $\sigma_{Y|X}^2$ if (1.8) is positive, where $\hat{\sigma}_{U|X}^2 = \sum_1^n (U_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2/(n-2)$ is the unbiased estimator of $\sigma_{U|X}^2$ and σ_δ^2 is the variance of δ_i. Consequently, when n is large,

$$F = \frac{(Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1})^2}{(1+H)\hat{\sigma}_{U|X}^2 - \sigma_\delta^2} \tag{1.9}$$

has an approximate F_{n-2}^1 distribution and an approximate $1-\alpha$ prediction interval for Y_{n+1} is given by

$$\{Y_{n+1} : F \le F_{n-2}^1(\alpha)\}. \tag{1.10}$$

Note that unlike the prediction interval (1.6), (1.10) does not have an exact coverage probability $1-\alpha$ since F in (1.9) has only an approximate F_{n-2}^1 distribution. In fact, even for large n, the probability of $(1+H)\hat{\sigma}_{U|X}^2 - \sigma_\delta^2$ being negative can be close to 0.5 for some parameters. For example, in (1.5) assume that $X_i, i = 1, \ldots, n$, are fixed numbers satisfying $\bar{X} \to c_1$ and $\sum_1^n (X_i - \bar{X})^2/n \to c_2 > 0$ where c_1 and c_2 are constants. If $\sigma_\epsilon^2 = O(1/n)$, $(1+H)\hat{\sigma}_{U|X}^2 - \sigma_\delta^2$ has mean $(1+H)\sigma_{U|X}^2 - \sigma_\delta^2$ and variance $2(1+H)^2\sigma_{U|X}^4/(n-2)$, both of order $O(1/n)$.

Therefore, the mean goes to zero faster than the standard deviation. Consequently, $P[(1+H)\hat{\sigma}^2_{U|X} - \sigma^2_\delta < 0]$ can be close to 0.5 for σ^2_ϵ close to 0. In particular,

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} P[Y_{n+1} \text{ satisfies } (1.10)] \leq \frac{1}{2} \tag{1.11}$$

if we use $[(1+H)\hat{\sigma}^2_{U|X} - \sigma^2_\delta] \vee 0$ to estimate $(1+H)\sigma^2_{U|X} - \sigma^2_\delta$ and hence to replace $(1+H)\hat{\sigma}^2_{U|X} - \sigma^2_\delta$ in (1.9). Furthermore, even if we substitute $\mid (1+H)\hat{\sigma}^2_{U|X} - \sigma^2_\delta \mid$ for $(1+H)\hat{\sigma}^2_{U|X} - \sigma^2_\delta$ in (1.9), for the commonly adopted value of $1-\alpha$, the left hand side of inequality (1.11) is still less than the nominal level (this is due to the fact that $\sigma^2_\epsilon/(\sigma^2_\epsilon + \sigma^2_\delta)$ goes to zero at a rate equal to $n^{-\frac{1}{2}}$, see Remark 3 after Theorem 4 for an explanation). In summary, even for large n, there are values of parameters for which the coverage probability of the traditional prediction interval (1.10) is appreciably below $1-\alpha$. To construct an honest prediction interval for Y_{n+1}, we will have to modify the usual approach.

The rest of this article is organized as follows. In Section 2 we construct an asymptotically honest prediction set for model (1.1)-(1.3). A small simulation of coverage probabilities showing that the proposed prediction set is superior to the traditional one is provided as well. We investigate the desirable property of the expected length of the asymptotically honest prediction interval in Section 3. Section 4 discusses the related calibration and prediction problems in the errors-in-variables model. Proofs are presented in the Appendix.

2. The Asymptotically Honest Prediction Set

In this section we consider the general model (1.1)-(1.3). To construct an asymptotically honest prediction set for $\mathbf{Y}_{n+1}$, we first observe that

$$\mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1} \mid \mathbf{X}_1, \ldots, \mathbf{X}_{n+1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{YY|X} + H\boldsymbol{\Sigma}_{UU|X}), \tag{2.1}$$

where $\hat{\boldsymbol{\beta}} = \sum_1^n \mathbf{U}_i\mathbf{X}_i'(\sum_1^n \mathbf{X}_i\mathbf{X}_i')^{-1}$ is the m.l.e. of $\boldsymbol{\beta}$, $H = \mathbf{X}_{n+1}'(\sum_1^n \mathbf{X}_i\mathbf{X}_i')^{-1}\mathbf{X}_{n+1}$, $\boldsymbol{\Sigma}_{YY|X}$ and $\boldsymbol{\Sigma}_{UU|X}$ are respectively the covariance matrices of $\mathbf{Y}_i$ and $\mathbf{U}_i$ given $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_{YY|X} = \boldsymbol{\Sigma}_{UU|X} - \boldsymbol{\Sigma}_{\delta\delta}$. Traditionally one uses

$$\hat{\boldsymbol{\Sigma}}_{YY|X} = \hat{\boldsymbol{\Sigma}}_{UU|X} - \boldsymbol{\Sigma}_{\delta\delta} \tag{2.2}$$

to estimate $\boldsymbol{\Sigma}_{YY|X}$ if (2.2) is positive definite, where

$$\hat{\boldsymbol{\Sigma}}_{UU|X} = \frac{1}{n-q}\sum_1^n (\mathbf{U}_i - \hat{\boldsymbol{\beta}}\mathbf{X}_i)(\mathbf{U}_i - \hat{\boldsymbol{\beta}}\mathbf{X}_i)' \tag{2.3}$$

is the unbiased estimator of $\boldsymbol{\Sigma}_{UU|X}$. It follows that when n is large,

$$F = (\mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1})' \frac{n-p-q+1}{(n-q)p} [(1+H)\hat{\boldsymbol{\Sigma}}_{UU|X} - \boldsymbol{\Sigma}_{\delta\delta}]^{-1} (\mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1}) \tag{2.4}$$

has an approximate $F^p_{n-p-q+1}$ distribution. Consequently, an approximate $1-\alpha$ prediction set for $\mathbf{Y}_{n+1}$ is given by

$$\{\mathbf{Y}_{n+1} : F \leq F^p_{n-p-q+1}(\alpha)\}, \tag{2.5}$$

where $F^p_{n-p-q+1}(\alpha)$ is the $1-\alpha$ quantile of the F distribution with p and $n-p-q+1$ degrees of freedom. Prediction set (2.5), of course, does not have an exact coverage probability $1-\alpha$ because F is only an approximate $F^p_{n-p-q+1}$ distribution. Arguing as in Section 1, it is easy to show that the probability that $(1+H)\hat{\mathbf{\Sigma}}_{UU|X} - \mathbf{\Sigma}_{\delta\delta}$ is positive definite will not converge to 1 uniformly over the parameter space as $n \to \infty$. Due to this, and the coverage probability of (2.5) based on some simulation results not reported here, we conjecture that the prediction set (2.5) is not asymptotically honest. To overcome the difficulty, a matrix $a_n\mathbf{\Sigma}_{\delta\delta}$ is added to $(1+H)\hat{\mathbf{\Sigma}}_{UU|X} - \mathbf{\Sigma}_{\delta\delta}$, where a_n satisfies certain conditions imposed below. The probability that the resultant matrix is positive definite converges to 1 uniformly in all parameters as $n \to \infty$.

Lemma 1. *Let a_n be a positive sequence satisfying $a_n n^{1/2} \to \infty$ and $a_n \to 0$. Then as $n \to \infty$,*

$$\inf_{\boldsymbol{\theta}\in\Omega} P[(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta} \textit{ is positive definite}] \to 1.$$

Lemma 1 indicates that the modified estimator $(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta}$ of $\mathbf{\Sigma}_{YY|X} + H\mathbf{\Sigma}_{UU|X}$ is a reasonable one in the sense of the asymptotic uniformity of probability when n is large. Theorem 2 below then constructs a bounded asymptotically honest prediction set for $\mathbf{Y}_{n+1}$.

Theorem 2. *Assume model* (1.1) – (1.3) *holds and a_n satisfies the conditions in Lemma* 1. *Let*

$$\mathbf{G} = \begin{cases} (1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta} & \textit{if the defined matrix is positive definite,} \\ \{[(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta}]^2\}^{\frac{1}{2}} & \textit{otherwise.} \end{cases}$$

The approximate $1-\alpha$ prediction set for Y_{n+1}

$$R_h = \left\{\mathbf{Y}_{n+1} : (\mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1})' \frac{n-p-q+1}{(n-q)p} \mathbf{G}^{-1} (\mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1}) \leq F^p_{n-p-q+1}(\alpha)\right\} \tag{2.6}$$

is asymptotically honest.

Remark 1. Note that $\mathbf{G}$ is also positive definite even if $(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta}$ is not, and this guarantees that (2.6) is always a bounded set. Moreover, the definition of $\mathbf{G}$ in the case where $(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n-1)\mathbf{\Sigma}_{\delta\delta}$ is not positive definite does not affect R_h being asymptotically honest, Lemma 1.

Remark 2. The asymptotically honest prediction set (2.6) reduces to the ordinary confidence ellipsoid in the absence of measurement error in the response.

The major difference between (2.6) and (2.5) is the presence of $a_n \mathbf{\Sigma}_{\delta\delta}$. With fixed values of parameters, this term is negligible in estimating $\mathbf{\Sigma}_{YY|X} + H\mathbf{\Sigma}_{UU|X}$ when n is large. Thus for large samples, the additional term $a_n \mathbf{\Sigma}_{\delta\delta}$ does not increase the volume of (2.6) much and helps achieve asymptotic honesty.

In Tables 1 and 2 we compare the simulation coverage probabilities of (2.6) with $a_n = 0$ and $a_n = \log(\log n)/(n)^{\frac{1}{2}}$. The case $a_n = 0$ corresponds to the traditional prediction set, whereas $a_n = \log(\log n)/(n)^{\frac{1}{2}}$ corresponds to an asymptotically honest one. (Based on some simulation results, we find that the value $\log(\log n)/(n)^{\frac{1}{2}}$ is an adequate choice.)

For definiteness, assume model (1.1)-(1.3) with $p = q = 2$,

$$\boldsymbol{\beta} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} 1 \\ X_i \end{pmatrix}, \; X_i \sim N(0,1), \mathbf{\Sigma}_{\epsilon\epsilon} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \mathbf{\Sigma}_{\delta\delta} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} d,$$

and $d = 0.5, 1, 1.5, 2, 5$. For each sample size n, we simulated 5,000 samples and calculated the coverage probability. In each cell of the tables, the number in the parenthesis is the coverage probability of (2.6) with $a_n = \log(\log n)/n^{1/2}$ and the other one with $a_n = 0$. From the tables we see that all coverage probabilities of the traditional prediction set are appreciably below the nominal levels. For fixed n, it seems that the shortage of the coverage probability becomes more serious when d is large. On the other hand, for fixed d, the coverage probabilities of the asymptotically honest prediction set improve as n increases. Finally, the asymptotically honest prediction set, except for the case $n \leq 50$, gives values close to the nominal levels.

Table 1. Coverage probabilities of (2.6) with $a_n = 0$ and $a_n = \log(\log n)/n^{1/2}$ (the values in the parentheses), nominal level= 0.9, replication= 5,000

sample size n	d 0.5	1	1.5	2	5
20	0.853 (0.900)	0.793 (0.871)	0.762 (0.860)	0.749 (0.843)	0.813 (0.846)
30	0.873 (0.903)	0.818 (0.890)	0.771 (0.867)	0.760 (0.872)	0.780 (0.851)
50	0.885 (0.907)	0.849 (0.903)	0.811 (0.902)	0.775 (0.890)	0.756 (0.864)
100	0.888 (0.904)	0.881 (0.916)	0.851 (0.916)	0.822 (0.909)	0.738 (0.899)
200	0.887 (0.902)	0.889 (0.914)	0.879 (0.916)	0.854 (0.917)	0.765 (0.918)

Table 2. Coverage probabilities of (2.6) with $a_n = 0$ and $a_n = \log(\log n)/n^{1/2}$ (the values in the parentheses), nominal level= 0.95, replication= 5,000

sample size n	d 0.5	1	1.5	2	5
20	0.903 (0.934)	0.854 (0.920)	0.824 (0.908)	0.809 (0.887)	0.859 (0.883)
30	0.924 (0.950)	0.877 (0.938)	0.825 (0.911)	0.812 (0.912)	0.834 (0.889)
50	0.935 (0.952)	0.905 (0.947)	0.865 (0.942)	0.832 (0.934)	0.809 (0.901)
100	0.940 (0.953)	0.932 (0.957)	0.911 (0.953)	0.877 (0.946)	0.795 (0.931)
200	0.943 (0.952)	0.939 (0.957)	0.929 (0.959)	0.914 (0.959)	0.822 (0.945)

3. The Univariate Case

In the univariate case the prediction set R_h for Y_{n+1} in (2.6) reduces to the interval

$$I_h = \left\{ Y_{n+1} : \frac{(Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1})^2}{\mid (1 + \frac{1}{n} + \frac{(X_{n+1}-\bar{X})^2}{\sum_1^n (X_i - \bar{X})^2})\hat{\sigma}^2_{U|X} + (a_n - 1)\sigma^2_\delta \mid} \le F^1_{n-2}(\alpha) \right\}, \tag{3.1}$$

where the dependence of I_h on n has been suppressed,

$$\hat{\beta}_1 = \frac{\sum_1^n (X_i - \bar{X})U_i}{\sum_1^n (X_i - \bar{X})^2}, \hat{\beta}_0 = \bar{U} - \hat{\beta}_1 \bar{X}, \text{ and } \hat{\sigma}^2_{U|X} = \frac{\sum_1^n (U_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n-2}.$$

Since the prediction interval (3.1) is asymptotically honest, it is expected to have a longer length than the traditional asymptotic prediction interval. Generally, an asymptotic prediction interval for Y_{n+1} is defined to be any sequence of intervals whose coverage probability converges to the nominal level as $n \to \infty$. The interval I_h has the following desirable property.

Theorem 3. *Let I be any $1-\alpha$ asymptotic prediction interval for Y_{n+1} with length having finite first moment. Assume that $\sum_1^n X_i/n \to c_1$ and $\sum_1^n (X_i - \bar{X})^2/n \to c_2 > 0$, c_1 and c_2 constants, if X_i are fixed numbers; assume $(X_{n+1} - \bar{X})^2/\sum_1^n (X_i - \bar{X})^2$ converges to 0 in L^1 if X_i are i.i.d. continuous random variables. Then*

$$\liminf_{n\to\infty} \frac{EL(I)}{EL(I_h)} \ge 1, \forall\, \boldsymbol{\theta},$$

where $EL(I)$ and $EL(I_h)$ denote expected lengths of I and I_h, respectively.

Next we proceed to a formula which will facilitate the computation of the coverage probability of I_h.

Theorem 4. *Under the assumptions of Theorem* 3 *the coverage probability of* I_h *in* (3.1) *depends on the parameters only through* $c = \sigma_\delta^2/(\sigma_\delta^2+\sigma_\epsilon^2)$, *and*

$$P(Y_{n+1} \in I_h) = 1 - 2E\{\Phi[-t_{n-2}(\alpha/2) \mid \frac{m-c+a_n c}{m-c}(1+\sqrt{\frac{2}{n-2}}\frac{m\chi}{m-c+a_n c}) \mid^{\frac{1}{2}}]\}, \tag{3.2}$$

where χ *is a standardized* χ^2_{n-2} *random variable,* $t_{n-2}(\alpha/2)$ *is the* $1-\alpha/2$ *quantile of* t_{n-2}, Φ *is the c.d.f. of* $N(0,1)$, *and* $m = 1+1/n+(X_{n+1}-\bar{X})^2/\sum_1^n(X_i-\bar{X})^2$. *Furthermore, if* $X_1,\ldots,X_{n+1}$ *are i.i.d. normal random variables, then* $(X_{n+1}-\bar{X})^2/\sum_1^n(X_i-\bar{X})^2$ *has* $(n+1)[n(n-1)]^{-1}F^1_{n-1}$ *distribution and is independent of* χ.

Remark 3. In Theorem 4 assume that the X_i are fixed numbers (the case of random X_i can be dealt with similarly). If $a_n = 0$, we choose $\sigma_\epsilon^2 = \{1/[m(1-(2/(n-2))^{\frac{1}{2}}]-1\}\sigma_\delta^2$ or equivalently $m(2/(n-2))^{\frac{1}{2}}/(m-c) = 1$ (i.e., in this case, $\sigma_\epsilon^2/(\sigma_\epsilon^2+\sigma_\delta^2)$ goes to 0 at the rate $n^{-\frac{1}{2}}$). Then since $\chi \xrightarrow{L} N(0,1)$ and $t_{n-2}(\alpha/2) \to z_{\alpha/2}$, the $1-\alpha/2$ quantile of $N(0,1)$, (3.2) converges to $1-2E\{\Phi[-z_{\alpha/2} \mid 1+N(0,1) \mid^{\frac{1}{2}}]\}$ which can be computed numerically (it is approximately 0.877, 0.831, and 0.747 if $\alpha = 0.05$, 0.1, and 0.2, respectively) and is less than $1-\alpha$ for the usual nominal levels. This means that if we drop the term $a_n\sigma_\delta^2$ from (3.1), the resultant interval will no longer be an asymptotically honest prediction interval.

By Theorem 4 we see that the coverage probability of I_h is a function of c for a given a_n, and hence the infimum of this coverage probability over the parameter space can be approximated by the minimum simulation coverage probability for $0 < c < 1$. For a given a_n and each $c = 10^{-2}i,\ 1 \le i \le 100$, we generate 5,000 realizations of m and χ, where m and χ are independent random variables with distributions described in Theorem 4 and $X_1,\ldots,X_{n+1}$ are i.i.d. $N(0,1)$. Then the approximate coverage probability can be computed by taking the average of these 5,000 values of Φ as the expectation of the Φ function. Among the values of the approximate coverage probabilities corresponding to different $c's$, the minimum value is taken as the approximate infimum of the coverage probability. Table 3 gives the results for $a_n = n^{-\frac{1}{2}}\log(\log n)$ and $a_n = 0$, $n = 10, 20, 30, 50, 100, 200$, and nominal level= 0.9, 0.95. From the table we see that the approximate infimum coverage probabilities of the asymptotically honest prediction interval differ from the nominal levels by less than 3% when $n \ge 30$, and are very close (within 1%) to the nominal levels when $n \ge 50$. On the other hand, those of the usual interval have a shortage greater than 7% for all sample sizes.

Table 3. Approximate infimum coverage probabilities of I_h

$a_n = n^{-\frac{1}{2}} \log(\log n)$

sample size n	nominal level 0.9	0.95	a_n
10	0.854 (0.52)	0.908 (0.66)	0.263
20	0.885 (0.69)	0.917 (0.68)	0.245
30	0.896 (0.78)	0.930 (0.71)	0.223
50	0.899 (0.05)	0.942 (0.82)	0.193
100	0.899 (0.01)	0.950 (0.10)	0.153
200	0.900 (0.03)	0.950 (0.01)	0.117

$a_n = 0$

sample size	nominal level 0.9	0.95
10	0.824 (0.61)	0.877 (0.64)
20	0.820 (0.75)	0.867 (0.71)
30	0.819 (0.77)	0.868 (0.80)
50	0.818 (0.80)	0.867 (0.82)
100	0.820 (0.87)	0.865 (0.86)
200	0.822 (0.88)	0.871 (0.91)

Number in each cell is the simulation value of the approximate

$$\inf_{0<c<1} \Big(1\text{–}2E\{\Phi[-t_{n-2}(\alpha/2) \mid \frac{m-c+a_n c}{m-c}(1+(\frac{2}{n-2})^{\frac{1}{2}} \frac{\chi^m}{m-c+a_n c}) \mid^{\frac{1}{2}}]\}\Big).$$

Number in the parenthesis is the value of c when coverage probability attains the minimum.

4. Calibration and Prediction in Errors-in-Variables Model

There are two problems in the errors-in-variables model related to model (1.1)-(1.3). One is the calibration problem and the other is the prediction problem.

Suppose that we have the following errors-in-variables model:

$$W_i = a + bZ_i + e_i, V_i = Z_i + \tau_i, 1 \le i \le n, \tag{4.1}$$

where $Z_i \sim N(\mu_z, \sigma_z^2), e_i \sim N(0, \sigma_e^2)$, and $\tau_i \sim N(0, \sigma_\tau^2)$ are three i.i.d. sequences of random variables. Here we can only observe $(W_i, V_i), 1 \le i \le n$. It is assumed that σ_τ^2, the variance of the measurement error τ_i, is known. In the calibration problem, based on the observed W_{n+1} and (W_i, V_i), $1 \le i \le n$, we would like to construct an asymptotically honest calibration interval for Z_{n+1}. In the prediction problem, if a true Z_{n+1} can be observed (for example, the instrument has been improved), we want to construct an asymptotically honest prediction interval for W_{n+1} on the basis of Z_{n+1} and (W_i, V_i), $1 \le i \le n$.

4.1. Calibration interval

Suppose that we are interested in interval estimation of Z_{n+1} based on W_{n+1} and $(W_i, V_i), 1 \le i \le n$. This problem can be treated as a prediction problem under (1.1)-(1.3). Since (W_i, V_i) has a bivariate normal distribution, we can inversely regress Z_i on W_i to obtain the model

$$Z_i = a^* + b^* W_i + e_i^*, V_i = Z_i + \tau_i, 1 \le i \le n+1, \tag{4.2}$$

where $b^* = b\sigma_z^2/(b^2\sigma_z^2 + \sigma_e^2)$, $a^* = (1 - bb^*)\mu_z - ab^*$, and $e_i^* = Z_i - a^* - b^* W_i$. Note that the error term e_i^* is uncorrelated with W_i. Obviously (4.2) is a special case of (1.1)-(1.3) except that the variables and parameters have been renamed. Hence the calibration problem in model (4.1) becomes one of prediction in the model (1.1)-(1.3) and we can apply the results in Section 3 to obtain an asymptotically honest calibration interval for Z_{n+1}.

4.2. Prediction interval

In model (4.1) suppose that we can observe a true Z_{n+1} in the future. It is not difficult to construct an asymptotic prediction interval for W_{n+1} which has finite length almost surely on the basis of Z_{n+1} and (W_i, V_i), $1 \le i \le n$. But, according to result of Hwang (1992), Theorem 2.2, such an interval can not be asymptotically honest. (This is due to the fact that the variance σ_z^2 of Z_i in the parameter space can be arbitrarily close to 0 and consequently, this reduces the information toward estimation of the parameter b.) In other words, any asymptotically honest prediction interval for W_{n+1} based on Z_{n+1} and (W_i, Z_i), $1 \le i \le n$, must have a positive probability of having an infinite length. We can use Theorem 2 to derive such a prediction interval.

As described in Section 4.1, we have model (4.2). To adopt the notations in model (1.1)-(1.3), we let $X_i = W_i$, $Y_i = Z_i$, $U_i = V_i$, $\epsilon_i = e_i^*$, $\delta_i = \tau_i$, $\beta_0 = a^*$, and $\beta_1 = b^*$. Then we can find a prediction interval for X_{n+1} ($= W_{n+1}$) by solving the inequality in (3.1) for X_{n+1} instead for Y_{n+1}. Although this prediction interval for X_{n+1} could be unbounded, it is asymptotically honest due to Theorem 2.

Appendix

Let $\mathbf{A}$ and $\mathbf{B}$ be any two matrices of the same dimension. From now on, the notation $\mathbf{A} < \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is a positive definite matrix.

Proof of Lemma 1. We only consider the case where $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$ are i.i.d. continuous random vectors. The case where $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$ are fixed vectors can be dealt with similarly. In this case the matrix $\hat{\mathbf{\Sigma}}_{UU|X}$ is positive definite almost surely. From model (1.1)-(1.3), we know that $\mathbf{\Sigma}_{UU|X} = \mathbf{\Sigma}_{\epsilon\epsilon} + \mathbf{\Sigma}_{\delta\delta}$ and

$$(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{\delta\delta} > \hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{UU|X}.$$

Consequently,

$$P[(1+H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{\delta\delta} \text{ is p.d.}] > P(\hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{UU|X} \text{ is p.d.}). \tag{A.1}$$

The probability on the right of inequality (A.1) is equivalent to

$$P[\mathbf{k}'(\hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{UU|X})\mathbf{k} > 0 \ \forall \ \mathbf{k} \ \text{ with } \| \ \mathbf{k} \ \|= 1]. \tag{A.2}$$

Let $\mathbf{T}$ be a nonsingular matrix such that $\mathbf{T}'\mathbf{\Sigma}_{UU|X}\mathbf{T} = \mathbf{I}_p$, the p by p identity matrix. Then (A.2) is $P[\mathbf{k}'(\hat{\mathbf{I}}_p - \mathbf{I}_p + a_n\mathbf{I}_p)\mathbf{k} > 0, \ \forall \ \ \| \ \mathbf{k} \ \|= 1]$, where $\hat{\mathbf{I}}_p = \mathbf{T}'\hat{\mathbf{\Sigma}}_{UU|X}\mathbf{T}$. It is clear that $\hat{\mathbf{I}}_p$ has a Wishart$(\mathbf{I}_p, n-q)$ distribution. Consequently, the elements of $\hat{\mathbf{I}}_p - \mathbf{I}_p$ have order $O_p(n^{-\frac{1}{2}})$ and their distributions are independent of all parameters. This implies that $\mathbf{k}'(\hat{\mathbf{I}}_p - \mathbf{I}_p)\mathbf{k} = O_p(n^{-\frac{1}{2}})$ independent of all parameters for any $\| \ \mathbf{k} \ \|= 1$. As a consequence, (A.2) is $P[O_p(n^{-\frac{1}{2}}) + a_n > 0, \ \forall \ \| \ \mathbf{k} \ \|= 1]$, independent of all parameters, and hence converges to 1 as $n \to \infty$. Combining this with (A.1), Lemma 1 is proved.

In order to prove Theorem 2, we need the following lemma.

Lemma 5. *Let $\{\mathbf{M}_n\}$ be a sequence of symmetric r by r random matrices and let $\{\lambda_n\}$ be the maximum absolute latent roots of $\{\mathbf{M}_n\}$. Then, as $n \to \infty$, every element of $\{\mathbf{M_n}\}$ converges to 0 in probability if λ_n does the same.*

Proof of Theorem 2 (Case 1). First we consider the case where $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$ are fixed vectors. For convenience, let $\mathbf{\Sigma}_{\cdot\cdot|X}$ represent the covariance matrix conditioned on $\mathbf{X}_i$ even if $\mathbf{X}_i$ are fixed vectors. Let $S = \{(\mathbf{t}_1, \ldots, \mathbf{t}_{n+1}) : \mathbf{t}_i \in R^q, \ 1 \le i \le n+1$ and the matrix $(\mathbf{t}_1, \ldots, \mathbf{t}_n)$ is of rank q $\}$, $\mathbf{X}^* = (\mathbf{X}_1, \ldots, \mathbf{X}_{n+1})$, $\mathbf{M} = (1 + H)\mathbf{\Sigma}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{\delta\delta}$, $\hat{\mathbf{M}} = (1 + H)\hat{\mathbf{\Sigma}}_{UU|X} + (a_n - 1)\mathbf{\Sigma}_{\delta\delta}$, and $\mathbf{V} = \mathbf{Y}_{n+1} - \hat{\boldsymbol{\beta}}\mathbf{X}_{n+1}$. We prove Theorem 2 by showing the following results:

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} P[\mathbf{V}'\mathbf{M}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)] = 1 - \alpha \tag{A.3}$$

and

$$| \ \mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} - \mathbf{V}'\mathbf{M}^{-1}\mathbf{V} \ | \le o_p(1), \tag{A.4}$$

where $o_p(1)$ converges to 0 in probability uniformly in all parameters.

Due to the fact that $\mathbf{V}'(\mathbf{M} - a_n\mathbf{\Sigma}_{\delta\delta})^{-1}\mathbf{V} \sim \chi^2_p$ and $\mathbf{M} \ge \mathbf{M} - a_n\mathbf{\Sigma}_{\delta\delta} > 0$, we have

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} P[\mathbf{V}'\mathbf{M}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)] \ge P[\mathbf{V}'(\mathbf{M}-a_n\mathbf{\Sigma}_{\delta\delta})^{-1}\mathbf{V} \le \chi^2_p(\alpha)] = 1-\alpha, \tag{A.5}$$

where $\chi^2_p(\alpha)$ is the $1 - \alpha$ quantile of χ^2_p. Since $a_n \to 0$ in (A.5), by Slutsky's Theorem it is obvious that

$$\lim_{n\to\infty} P[\mathbf{V}'\mathbf{M}^{-1}\mathbf{V} \le \chi^2_p(\alpha)] = 1 - \alpha.$$

Then with (A.5), (A.3) is established.

Remark 4. Since the result of (A.3) holds for all $\mathbf{X}^* \in S$, in fact we have proved that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} \inf_{\mathbf{X}^*\in S} P[\mathbf{V}^{'}\mathbf{M}^{-1}\mathbf{V} \le pF^{p}_{n-p-q+1}(\alpha)] = 1-\alpha.$$

To prove (A.4), let $\mathbf{C} = \hat{\mathbf{M}} - \mathbf{M} = (1+H)(\hat{\mathbf{\Sigma}}_{UU|X} - \mathbf{\Sigma}_{UU|X})$. By Muirhead (1982), Pg.592, there exists a nonsingular matrix $\mathbf{\Gamma}$ such that $\mathbf{\Sigma}^{-1}_{UU|X} = \mathbf{\Gamma}^{'}\mathbf{\Gamma}$ and $\mathbf{M}^{-1} = \mathbf{\Gamma}^{'}\mathbf{D}\mathbf{\Gamma}$, where $\mathbf{D} = \mathrm{diag}(d_1,\ldots,d_p)$ with $d_1,\ldots,d_p$ being latent roots of $\mathbf{\Sigma}_{UU|X}\mathbf{M}^{-1}$. Consequently,

$$\mathbf{\Gamma}\frac{C}{1+H}\mathbf{\Gamma}^{'} = \mathbf{\Gamma}\hat{\mathbf{\Sigma}}_{UU|X}\mathbf{\Gamma}^{'} - \mathbf{I}_p. \tag{A.6}$$

Since $(n-q)\hat{\mathbf{\Sigma}}_{UU|X}$ has a Wishart$(\mathbf{\Sigma}_{UU|X}, n-q)$ distribution, it follows that $(n-q)\mathbf{\Gamma}\hat{\mathbf{\Sigma}}_{UU|X}\mathbf{\Gamma}^{'}$ has a Wishart$(\mathbf{I}_p, n-q)$ distribution. By (A.6), we know that the elements of $\mathbf{\Gamma}\mathbf{C}\mathbf{\Gamma}^{'}/(1+H)$ have order $O_p(n^{-\frac{1}{2}})$ and their distributions are independent of all parameters.

From the fact $(H+a_n)\mathbf{\Sigma}_{UU|X} \le \mathbf{M} \le (1+H)\mathbf{\Sigma}_{UU|X}$, we have

$$\frac{\mathbf{\Sigma}^{-1}_{UU|X}}{1+H} \le \mathbf{M}^{-1} \le \frac{\mathbf{\Sigma}^{-1}_{UU|X}}{H+a_n}. \tag{A.7}$$

Let the latent vectors associated with the latent roots d_i of $\mathbf{\Sigma}_{UU|X}\mathbf{M}^{-1}$ be denoted by $\boldsymbol{\nu_i}$. Then $\boldsymbol{\nu}_i^{'}\mathbf{M}^{-1}\boldsymbol{\nu}_i - d_i\boldsymbol{\nu}_i^{'}\mathbf{\Sigma}^{-1}_{UU|X}\boldsymbol{\nu}_i = 0$. Consequently, by (A.7)

$$\frac{1}{1+H}\boldsymbol{\nu}_i^{'}\mathbf{\Sigma}^{-1}_{UU|X}\boldsymbol{\nu}_i \le d_i\boldsymbol{\nu}_i^{'}\mathbf{\Sigma}^{-1}_{UU|X}\boldsymbol{\nu}_i \le \frac{1}{H+a_n}\boldsymbol{\nu}^{'}\mathbf{\Sigma}^{-1}_{UU|X}\boldsymbol{\nu}_i$$

and hence

$$\frac{1}{1+H} \le d_i \le \frac{1}{H+a_n}. \tag{A.8}$$

By a straightforward matrix computation,

$$\hat{\mathbf{M}}^{-1} - \mathbf{M}^{-1} = (\mathbf{M}+\mathbf{C})^{-1} - \mathbf{M}^{-1} = -\mathbf{M}^{-1}\mathbf{C}(\mathbf{I}_p + \mathbf{M}^{-1}\mathbf{C})^{-1}\mathbf{M}^{-1}. \tag{A.9}$$

Moreover,

$$\mathbf{M}^{-1}\mathbf{C} = \mathbf{\Gamma}^{'}\mathbf{D}\mathbf{\Gamma}\mathbf{C} = \frac{1+H}{(H+a_n)\sqrt{n}}\mathbf{\Gamma}^{'}\mathbf{\Psi}\mathbf{\Gamma}^{-1^{'}}, \tag{A.10}$$

where $\mathbf{\Psi} = (n)^{\frac{1}{2}}(H+a_n)\mathbf{D}\mathbf{\Gamma}\mathbf{C}\mathbf{\Gamma}^{'}/(1+H)$. Here the elements of the matrix $\mathbf{\Psi}$ are $O_p(1)$ free of all parameters, by (A.6) and (A.8). Substituting (A.10) in (A.9) and simplifying the result, we have

$$\mathbf{\Gamma}^{-1^{'}}[\hat{\mathbf{M}}^{-1} - \mathbf{M}^{-1}]\mathbf{\Gamma}^{-1} = \mathbf{K}\mathbf{D}, \tag{A.11}$$

where

$$\mathbf{K} = \frac{-(1+H)}{(H+a_n)\sqrt{n}}\mathbf{\Psi}[\mathbf{I}_p + \frac{1+H}{(H+a_n)\sqrt{n}}\mathbf{\Psi}]^{-1}.$$

As $n \to \infty$ (and hence $a_n(n)^{\frac{1}{2}} \to \infty$),

$$\frac{1+H}{(n)^{\frac{1}{2}}(H+a_n)} \le \begin{cases} \frac{2}{n^{1/2}a_n} & \text{if } 0 < H \le 1, \\ \frac{2}{n^{1/2}} & \text{if } H > 1, \end{cases}$$

converges to 0. Consequently, $\mathbf{K}$ converges to the zero matrix in probability uniformly over the parameter space. From (A.11) it is obvious that $\mathbf{KD}$ is symmetric and hence

$$\mathbf{D}^{-\frac{1}{2}}\mathbf{KD}^{\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}K'\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{\frac{1}{2}}\mathbf{K}'\mathbf{D}^{-\frac{1}{2}}.$$

Thus, $\mathbf{D}^{-\frac{1}{2}}\mathbf{KD}^{\frac{1}{2}}$ is symmetric as well. Note that the latent roots of $\mathbf{K}$ converge to 0 in probability uniformly in all parameters since the matrix $\mathbf{K}$ does the same. As a result, $\mathbf{D}^{-\frac{1}{2}}\mathbf{KD}^{\frac{1}{2}}$, which has the same latent roots as $\mathbf{K}$, converges to the zero matrix in probability uniformly in all parameters by Lemma 5. Recall that $\mathbf{M}^{-1} = \mathbf{\Gamma}'\mathbf{D\Gamma}$ and $\mathbf{V}'\mathbf{M}^{-1}\mathbf{V} \le \mathbf{V}'(\mathbf{M} - a_n\mathbf{\Sigma}_{\delta\delta})^{-1}\mathbf{V} \sim \chi^2_p$. This implies that $\mathbf{V}'\mathbf{\Gamma}'\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\mathbf{\Gamma V} = O_p(1)$ or $\| \mathbf{D}^{\frac{1}{2}}\mathbf{\Gamma V} \| = O_p(1)$, where $O_p(1)$ is free of all parameters. By this, the uniformity result on $\mathbf{D}^{-\frac{1}{2}}\mathbf{KD}^{\frac{1}{2}}$, and (A.11) we have

$$\mathbf{V}'(\hat{\mathbf{M}}^{-1} - \mathbf{M}^{-1})\mathbf{V} = \mathbf{V}'\mathbf{\Gamma}'\mathbf{KD\Gamma V} = \mathbf{V}'\mathbf{\Gamma}'\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}\mathbf{KD}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\mathbf{\Gamma V} = o_p(1), \quad \text{(A.12)}$$

where $o_p(1)$ is free of all parameters.

Remark 5. In the previous proof, the norm of $\mathbf{D}^{\frac{1}{2}}\mathbf{\Gamma V}$ is $O_p(1)$, free of all parameters and the values of $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$. Also $\mathbf{K}$ converges to the zero matrix in probability uniformly over the parameter space and S. Consequently the supremum of the term on the left of (A.12) over S converges to 0 in probability uniformly in all parameters. Combining this and Remark 4, we obtain

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\Omega} \inf_{\mathbf{X}^*\in S} P[\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)] = 1 - \alpha,$$

which is stronger than Theorem 2.

Proof of Theorem 2 (Case 2). Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$ are i.i.d. continuous random vectors with joint distribution function $G_{\boldsymbol{\theta}^*}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1})$, $\boldsymbol{\theta}^*$ the vector parameter involved in this distribution. Let $\boldsymbol{\eta}= (\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ be the entire vector parameter where $\boldsymbol{\theta}$ consists of all parameters not involved in $G_{\boldsymbol{\theta}^*}$ and let Δ be the corresponding parameter space. Since $G_{\boldsymbol{\theta}^*}$ is continuous, it is obvious that $P(\mathbf{X}^* \in S) = 1$, where $\mathbf{X}^*$ and S are defined the same as in the proof of Case 1. Consequently,

$$P_{\boldsymbol{\eta}}[\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)] = \int_S P_{\boldsymbol{\theta}}[\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha) \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}]dG_{\boldsymbol{\theta}^*}(\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}). \quad \text{(A.13)}$$

Since the integrand of the integral on the right side of equality (A.13) is larger than

$$\inf_{\mathbf{x}^*\in S} P_{\boldsymbol{\theta}}[\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha) \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{n+1} = \mathbf{x}_{n+1}],$$

so is the probability $P_{\boldsymbol{\eta}}[\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)]$. By Remark 5 in Case 1, it follows that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\eta}\in\Delta} P_{\boldsymbol{\eta}}(\mathbf{V}'\hat{\mathbf{M}}^{-1}\mathbf{V} \le pF^p_{n-p-q+1}(\alpha)) \ge 1-\alpha.$$

It is trivial to prove that the equality holds exactly.

Proof of Theorem 3. We only provide the proof where $X_1, \ldots, X_{n+1}$ are fixed numbers. The proof where $X_1, \ldots, X_{n+1}$ are i.i.d. continuous random variables can be dealt with similarly. By the assumptions,

$$EL(I_h) = E\Big[2t_{n-2}(\alpha/2)((1+\frac{1}{n}+\frac{(X_{n+1}-\bar{X})^2}{\sum_1^n (X_i-\bar{X})^2})\hat{\sigma}^2_{U|X}+(a_n-1)\sigma^2_\delta)^{\frac{1}{2}}\Big] \to 2z_{\alpha/2}\sigma_\epsilon,$$

where $t_{n-2}(\alpha/2)$ and $z_{\alpha/2}$ are respectively the $1-\alpha/2$ quantiles of t_{n-2} and $N(0,1)$. Suppose that there exists a θ_0 such that

$$\liminf_{n\to\infty} \frac{EL(I)}{EL(I_h)} = 1-\epsilon < 1,$$

where $0 < \epsilon < 1$. Due to the fact that $EL(I_h) \to 2z_{\alpha/2}\sigma_\epsilon$, there exists a subsequence n_k of n such that

$$\frac{EL(I)}{2z_{\alpha/2}\sigma_\epsilon} < 1-\frac{\epsilon}{2} \quad \text{for } n = n_k.$$

Moreover, since Y_{n+1} is normally distributed with mean EY_{n+1}, we have

$$\begin{aligned} P(Y_{n+1}\in I) &\le P\Big(\mid \frac{Y_{n+1}-EY_{n+1}}{L(I)} \mid\le \frac{1}{2}\Big) = P\Big(\mid \frac{Y_{n+1}-EY_{n+1}}{\sigma_\epsilon} \mid\le \frac{L(I)}{2\sigma_\epsilon}\Big) \\ &= 1-2E\Phi[-\frac{L(I)}{2\sigma_\epsilon}] < 1-2\Phi[-\frac{EL(I)}{2\sigma_\epsilon}] < 1-\alpha \quad \text{for } n = n_k. \end{aligned}$$

Strict inequality is valid by Jensen's inequality. This shows that I is not a $1-\alpha$ asymptotic prediction interval for Y_{n+1}, which contradicts the assumption.

Proof of Theorem 4. Given $X_1, \ldots, X_{n+1}$, the conditional distribution of $Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1}$ is $N[0, m\sigma^2_{U|X} - \sigma^2_\delta]$. Let

$$V_1 = m\sigma^2_{U|X} - \sigma^2_\delta,\ V_2 = m\sigma^2_{U|X} + (a_n-1)\sigma^2_\delta,\ \hat{V}_2 = m\hat{\sigma}^2_{U|X} + (a_n-1)\sigma^2_\delta.$$

Then

$$P(Y_{n+1} \in I_h) = P\Big[\frac{(Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1})^2}{| \, m\hat{\sigma}^2_{U|X} + (a_n - 1)\sigma^2_\delta \, |} \le F^1_{n-2}(\alpha)\Big]$$
$$= P\Big[\, | \, Z \, | \le t_{n-2}(\alpha/2)(\frac{V_2}{V_1}\frac{| \, \hat{V}_2 \, |}{V_2})^{\frac{1}{2}}\Big], \tag{A.14}$$

where $Z = (Y_{n+1} - \hat{\beta}_0 - \hat{\beta}_1 X_{n+1})/\sqrt{V_1}$ is $N(0,1)$. Note that the distribution of Z is independent of X_i and $\hat{\sigma}^2_{U|X}$ and hence Z is independent of $\hat{V}_2$. By a straightforward computation, we have

$$\frac{\hat{V}_2}{V_2} = 1 + \chi(\frac{2}{(n-2)})^{\frac{1}{2}}\frac{m}{m - c + a_n c} \quad \text{and} \quad \frac{V_2}{V_1} = \frac{m - c + a_n c}{m - c}, \tag{A.15}$$

where $\chi = [(n-2)\hat{\sigma}^2_{U|X}/\sigma^2_{U|X} - (n-2)]/(2(n-2))^{\frac{1}{2}}$. Since $\hat{\sigma}^2_{U|X}$ is independent of $\sum_1^n (X_i - \bar{X})^2$, $\bar{X}$ and X_{n+1}, it follows that m is independent of χ. Theorem 4 is an obvious consequence if we replace $\hat{V}_2/V_2$ and V_2/V_1 in (A.14) by those in (A.15), respectively.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley, New York.

Buonaccorsi, J. P. (1996). Measurement error in the response in the general linear model. *J. Amer. Statist. Assoc.* **91**, 633-642.

Cheng, C and Van, N. (1994). On estimating linear relationships when both variables are subject to errors. *J. Roy. Statist. Soc. Ser. B* **56**, 167-183.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley, New York.

Gleser, L. J. and Hwang, J. T. (1987). The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables. *Ann. Statist.* **15**, 1351-1362.

Hwang, J. T. G. (1992). Prediction and tolerance intervals for linear measurement error models with application in predicting the compressive strength of concrete. Technical report, Cornell University, Ithaca, New York.

Huwang, L. (1996). Asymptotically honest confidence sets for structural errors-in-variables models. *Ann. Statist.* **24**, 1536-1546.

Kendall, M. A. and Stuart, A. (1979). *The Advanced Theory of Statistics* 2, 4th ed. Hafner, New York.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Institute of Statistics, National Tsing Hua University, Hsinchu 30043, Taiwan.

E-mail: huwang@stat.nthu.edu.tw

(Received December 1996; accepted April 1998)