Confidence Sets for High-dimensional Empirical Linear Prediction (HELP) with Dependent Error Structure

A. Adam Ding, Department of Mathematics, Northeastern University
and J.T. Gene Hwang, Department of Mathematics, Cornell University

Summary

In this paper, we provide asymptotic theory and use it to construct confidence sets for High-dimensional Empirical Linear Prediction (HELP). HELP is a statistical prediction technique based on a model similar to a factor analysis model. The aim of HELP model however is to predict future observations, and is different from the usual factor analysis. We generalize HELP to the cases of non-iid errors, thus going beyond the standard factor analysis model. We studied an unusual asymptotic theory where the size of the covariance matrix goes to infinity. This kind of asymptotic theory is important for practical applications and leads to confidence sets with much better coverage probabilities than the ones derived from the typical asymptotic theory where the size of the covariance matrix remains fixed. Although the covariance matrix may not be estimated consistently, it is shown that the point estimates of HELP are consistent. Also, valid asymptotic confidences sets are constructed based on the inconsistently estimated covariance matrix.

AMS 1991 subject classifications. Primary 62H25; secondary 62E20, 62F25, 62N99, 62F12.

KEY WORDS: Prediction; Asymptotic theory; Factor analysis; Principal component analysis.

1 Introduction.

Recently, a technique called High-dimensional Empirical Linear Prediction (HELP) is developed (Souders and Stenbakken, 1987, 1991, 1994; Stenbakken, 1996; Hwang and Liu, 1995a, b). The basic idea is to use an empirical linear model to predict a partially observed high-dimensional vector. In its simplest form (called purely empirical linear model), the model is very similar to a factor analysis model. Most of the successful applications in fact involve only purely empirical linear model, which we focus on in the present paper.

To describe the problem, let y be the measurement of an m-dimensional vector μ, which represents m characteristics of a device that the manufacturer is interested in. It is assumed that only a partial subvector y_1 of y is observed and the rest of y is denoted as y_2. The aim of HELP is to use y_1, together with some training data, to estimate μ or to predict y_2.

Is it possible to predict μ (therefore y_2) well by a y_1 that has a much smaller dimension? HELP technique was applied to electronic A/D converters testing in the papers cited above. In that example, each 13-bit converter has measurements of $m = 8192$ characteristics, called transition levels. In applying HELP, they used training data corresponding to exhaustively measuring 88 similar converters to estimate the design matrix. Then they use only 64 out of the 8192 measurements on a future converter to predict μ (or y_2) well. This reduces the measuring of y to less than one percent; the saving in time and money is tremendous.

Specifically, the purely empirical linear model assumes that the training data on the n converters satisfy:

$$y^i = \mu^i + \varepsilon^i = \chi\beta^i + \varepsilon^i, \qquad i = 1, 2, ..., n \tag{1}$$

where χ is a $m \times k$ matrix, $\mu^i = \chi\beta^i$ corresponds to the m characteristics of the ith device, and ε^i is a zero mean vector representing measurement errors. This model is similar to a factor analysis model since the quantities including χ on the right hand side of the above displayed equation are assumed to be unknown whereas y^i's are observed. In the above cited papers, it is assumed that ε^i has a zero mean vector and covariance matrix $\sigma^2 I$, where

I is an identity matrix. This assumption slightly differs from a standard factor analysis model where the covariance matrix of ε^i is assumed to be diagonal with possible different diagonal elements. The HELP would then estimate the design matrix χ through singular value decomposition in a spirit similar to a principal component analysis.

The vector of interest y satisfies

$$y = \mu + \varepsilon = \chi\beta + \varepsilon \tag{2}$$

However, only a t-dimensional subvector y_1 of y is observed. The estimated design matrix $\hat{\chi}$ and y_1 were then combined to yield an estimate for μ (and y_2).

An important problem is to asses the uncertainty of prediction. To solve this problem, Hwang and Liu (1995a,b) and Ding and Hwang (1999) construct statistical intervals; those for μ are called *confidence intervals* and those for y_2 are called *prediction intervals*. In this paper, we shall focus on confidence sets for μ since the theory for y_2 is very similar. Hwang and Liu's confidence intervals are based on a standard asymptotic approach, in which m is fixed and n approaches infinity. In contrast, the approach of Ding and Hwang (1999) is based on asymptotic theory where both m and n approach infinity, resulting in confidence intervals with better coverage probability. In applications, m is often larger than n. For example, in the aforementioned application to electronic A/D converters, $m = 8192$ and $n = 88$. Hence it seems intuitively more reasonable to allow both m and n approach infinity. However, the problem is much more difficult as $m \to \infty$, since one has to deal with covariance matrix with sizes increasing to infinity as in Ding and Hwang (1999). All the statistical intervals, however, are based on the assumption that the elements in ε^i are independently identically distributed (i.i.d).

The objective of this paper is two-folded. First, we construct statistical intervals for the cases where the elements in ε^i are not i.i.d. so that they can have wider applications in practice. The variances of elements in ε^i may not be equal. It is also quite likely in practice that there will be some correlation between the elements in ε^i so that their covariance matrix Σ is not necessarily diagonal. Hence we would like to consider a more general form for covariance matrix Σ. Therefore, we consider a HELP model that is more general and is no longer a special case of a factor analysis model. To gain some information on the covariance matrix, we assume that there are repeated measurements on some training devices.

Second, this paper provides a technical account, albeit in a more general model setup, for Ding and Hwang (1999) which states the theorems but omits most of the proofs. Since the asymptotic theory of letting both m and n increasing to infinity is highly unusual in literature, readers can better appreciate the technique through a more detailed account of the proofs.

Section 2 provides detailed description of the model, motivations, assumptions and the prediction procedure HELP. Section 3 provides some asymptotic theory which is used in Section 4 to construct confidence sets. Simulations were conducted to check the validity of the proposed confidence sets, and the results are reported in Section 5. The more involved technical details are given in the Appendix.

2 Model Assumptions and Prediction Procedures

2.1 Model description

We shall rewrite model (1) and (2) in a matrix form. Combining μ^i into a matrix and applying singular value decomposition, we arrive at

$$\text{(3)} \qquad \begin{cases} \mathcal{U} & = (\mu^1, \mu^2, ..., \mu^n) & = \sqrt{mn}SDV \\ \mu & & = \sqrt{m}S\eta \end{cases}$$

Here the unknown matrices S, D, V and η are of sizes $m \times k$, $k \times k$, $k \times n$ and $k \times 1$; $S'S = VV' = I_k$, the identity matrix of size $k \times k$; D is a diagonal matrix. The scaling constants $\sqrt{mn}$ and $\sqrt{m}$ are chosen for the convenience of presentation of asymptotic theory and they do not affect the end results.

In the present paper, each m-dimensional vector μ^i is measured repeatedly r_i times,

$$\text{(4)} \qquad y^i_{(j)} = \mu^i + \varepsilon^i_{(j)} = \chi\beta^i + \varepsilon^i_{(j)}, \qquad i = 1, ..., n \text{ and } j = 1, ..., r_i$$

The total number of measurements on the training vectors is therefore $r = r_1 + r_2 + ... + r_n$. In Ding and Hwang (1999), $r_i = 1$, which reduces (4) to (1). We shall assume that $\varepsilon^i_{(j)}$ has an m-dimensional normal distribution with mean zero and covariance matrix Σ. Here Σ is not necessarily diagonal, whereas in Ding and Hwang, $\Sigma = \sigma^2 I_m$.

The inference under an arbitrary covariance matrix Σ requires a consistent estimate of Σ with $\frac{m(m+1)}{2}$ free parameters, which requires that r to be much larger than m. However, in practice, that would be too costly. (In the A/D converters example, that means exhaustively measuring converters more than 8192 times comparing with 88 times now.) Hence we would need to make assumptions about Σ so that HELP still works for a sample size comparable to the previous applications (e.g., we can measure half of the training vectors twice resulting in $r = 1.5n$). On the other hand, these assumptions should be general enough to hold in likely applications.

Let $\sigma_1^2 \geq \sigma_2^2 \geq ... \geq \sigma_m^2 \geq 0$ denote the eigenvalues of matrix Σ, and $\bar{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma_i^2 = \frac{1}{m} trace(\Sigma^2)$. Then $\bar{\sigma}^2$ represents the average variation in the m elements of measurement errors ε. We expect that, as m increases, $\bar{\sigma}^2$ will remain at a constant order (i.e., not blow up to infinity or diminishes to zero). Also, σ_1^2 represents the largest single source of variation in the measurement errors. We expect that the ratio of σ_1^2 to the average variation $\bar{\sigma}^2$ would not blow up to infinity as m increases. Otherwise, it would mean that the measurement errors essentially comes from a single source with infinite variance. Mathematically, we can write the above intuition as

$$\liminf_{m\to\infty} \bar{\sigma} > 0, \qquad \limsup_{m\to\infty} \sigma_1 < \infty \tag{5}$$

This condition is the fundamental reason that HELP works well in practice. This condition is not very restrictive and is satisfied in many important cases. In the following Theorem, we show that condition (5) is satisfied if the measurement errors come from a causal auto regressive-moving average (ARMA) process. For the definition of a causal ARMA process, see Brockwell and Davis (1987). The condition of being causal is probably necessary for an ARMA process to satisfy (5).

Theorem 2.1 *Assume that e_i's follow a causal ARMA(p,q) process*

$$e_i = \phi_1 e_{i-1} + \cdots + \phi_p e_{i-p} + \theta_0 z_i + \theta_1 z_{i-1} + \cdots + \theta_q z_{i-q},$$

where z_i, $i = -\infty, ..., \infty$ are i.i.d. standard normal random variables. Then for any fixed values of p and q, as $m \to \infty$, the covariance matrix of $\varepsilon = (e_1, e_2, ..., e_m)'$ satisfies condition (5).

Proof is included in the Appendix.

HELP reduces the measurement on a future vector y to a t-dimensional subvector y_1 by identifying the underlining variance structure (the design matrix S) from the training data $\{y^i_{(j)}\}$. With the design matrix S known, the subvector y_1 can then be used in a linear regression model to provide inference on μ. Hence the reliable inference on μ requires a good estimate of S from the training data and also a good estimate of the variation in the resulting linear predictor. Next, we define the notations and motivate the assumptions. Then we provide a detailed description of how HELP works, and in the next section we provide theoretical justification for why HELP works.

The basic motivation for HELP is that excluding the measurement errors, the true variation of y comes only from a subspace with a limited number of dimension k respresented by the design matrix S of dimensions $m \times k$. When n is relatively larger than k, we can identify S from the training data reasonably well. When t is larger than k, we can then estimate the projection of y to this k-dimensional subspace well and hence the whole vector μ also. To assess the uncertainty of the estimator for μ, we also need to estimate the covariance matrix Σ, which requires a large number $(r-n)$ of repeated measurements. Therefore we shall provide theoretical foundation for the HELP techniques in the following asymptotic situation

(6) $$m \to \infty,\ t \to \infty,\ n \to \infty,\ (r-n) \to \infty \text{ and } \tfrac{m}{n} \to c \in [0, \infty].$$

Notice that m and n are of different orders when $c = 0$ or $c = \infty$.

As m and n change, we expect that the singular values d_i's remain constants. We also expect that the t-dimentional subvector would capture about $\frac{t}{m}$ portion of the total variation in the m-dimentional vector. Under (6), the above intuitions is represented mathematically by the following conditions:

$$\lim d_i = \bar{d}_i < \infty \tag{7}$$

$$\bar{d}_i\text{'s are distinct and positive if } i \leq k. \tag{8}$$

$$\lambda_1/\lambda_k = O(1), \qquad \lambda_1 = O(\sqrt{\frac{t}{m}}) \text{ and } (\lambda_k^*)^{-1} = O(\sqrt{\frac{t}{m}}), \tag{9}$$

where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k > 0$ and $\lambda_1^* \geq \lambda_2^* \geq ... \geq \lambda_k^* > 0$ denote respectively the singular values of S_1 and $(S_1'S_1)^{-1}S_1'M_1$. Here and below, M denotes the "square root" of Σ, i.e. $M^2 = \Sigma$ and M is symmetric and semi-positive definite. Further, S_1 and M_1 consist of the first t rows of S and M. More detailed explanation of the motivations behind these conditions can be found in Ding and Hwang (1999).

We shall use $\|A\|$ to denote the L_2-norm of a matrix A, that is, the largest singular value of A. Also, we shall use $O_p(\cdot)$ and $o_p(\cdot)$ to denote the probability big "O" and little "o". They may be scalars or matrices. For a matrix A, $O_p(A)$ and $o_p(A)$ denote $O_p(\|A\|)$ and $o_p(\|A\|)$.

Remark 1: In practice, the coordinates, called test points, of the observable subvector y_1 should be chosen by an algorithm (Ding and Hwang 1999). For ease of presentation, we assume that we have switched them around so that y_1 consists of the first t elements of y.

2.2 Estimating μ: A generalized HELP.

Here we describe the point estimator for μ in the setup of previous subsection generalizing HELP to the case of non-diagonal and unknown Σ.

1. Estimation of the covariance matrix Σ.

With the repeated measurements, a natural estimator is the moment estimator

$$\hat{\Sigma} = \frac{1}{r-n} \sum_{i=1}^{n} \sum_{j=1}^{r_i} (y^i_{(j)} - \bar{y}^i)(y^i_{(j)} - \bar{y}^i)' \tag{10}$$

where $\bar{y}^i = \sum_{j=1}^{r_i} y^i_{(j)}/r_i$ is the mean measurement of the ith vector μ^i.

2. Estimation of the true dimension k.

We shall estimate k by

$$\hat{k} = \text{ number of singular values of } \bar{Y} \text{ that are greater than } \tilde{\sigma}(\sqrt{m} + \sqrt{n})(1 + \delta \cdot ln(min(m,n))) \tag{11}$$

where $\bar{Y} = (\bar{y}^1, \bar{y}^2, ..., \bar{y}^n)$, $\tilde{\sigma}^2 = \frac{1}{m} trace(\hat{\Sigma})$ and $\delta > 0$. Here $ln(min(m,n))$ can be replace by any function $l(m,n)$ such that as m, $n \to \infty$, $l(m,n) \to \infty$ and $l(m,n) \cdot (\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}) \to 0$.

3. Estimation of the design matrix S by a singular value decomposition.

That is, to decompose

$$\frac{\bar{Y}}{\sqrt{mn}} = (\hat{S}, \widehat{S_0}) \begin{pmatrix} \widehat{D} & 0 \\ 0 & \widehat{D_0} \end{pmatrix} \begin{pmatrix} \hat{V} \\ \hat{V_0} \end{pmatrix} \tag{12}$$

Here the right-hand side is the product of three matrices; the first matrix (with size $m \times m$) and the third matrix (with size $n \times n$) are orthogonal matrices and the second matrix (with size $m \times n$) is diagonal with decreasing nonnegative diagonal elements $\hat{d}_i$, called the singular values. Here and elsewhere in this paper, we use 0 to denote a zero matrix with appropriate sizes. Note that we partitioned the three matrices on the right hand side, so that $\hat{S}$, $\hat{D}$ and $\hat{V}$ has the same size as that of S, D and V ($m \times k$, $k \times k$ and $k \times n$ respectively). In practice, since k is also estimated, the sizes of $\hat{S}$, $\hat{D}$ and $\hat{V}$ are $m \times \hat{k}$, $\hat{k} \times \hat{k}$ and $\hat{k} \times n$ respectively.

It is easier to see why $\hat{S}$, $\hat{D}$ and $\hat{V}$ are natural estimators of S, D and V if we combine the first equation in (3) with (4) in the following form.

$$\frac{\bar{Y}}{\sqrt{mn}} = (S, S_0) \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V \\ V_0 \end{pmatrix} + \frac{M\mathcal{E}R}{\sqrt{mn}} \tag{13}$$

where R is a diagonal matrix with entries $1/\sqrt{r_1}, ..., 1/\sqrt{r_n}$, and $\mathcal{E}$ is an m by n matrix with independent entries that follows standard normal distribution. Here S_0 and V_0 are chosen so that (S, S_0) and (V', V_0') are orthogonal matrices with size $m \times m$ and $n \times n$.

4. Inference of μ.

Here and below we decompose

$$\widehat{S} = \begin{pmatrix} \widehat{S}_1 \\ \widehat{S}_2 \end{pmatrix} \qquad M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \tag{14}$$

where $\widehat{S}_1$ and M_1 consist the first t rows of the corresponding matrices $\widehat{S}$ and M.

We estimate η by

$$\widehat{\eta} = (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1' y_1/\sqrt{m}. \tag{15}$$

Hence $\mu = \sqrt{m}S\eta$ can be estimated by

$$\hat{\mu} = \sqrt{m}\widehat{S}\widehat{\eta} = \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1' y_1. \tag{16}$$

Then the uncertainty of $\hat{\mu}$ is assessed using information on $\hat{\Sigma}$, and we provide an asymptotically valid confidence set in Section 4.

For the prediction of y_2, we will use the corresponding coordinates from $\hat{\mu}$, $\hat{y}_2 = \sqrt{m}\widehat{S_2}\widehat{\eta}$. The prediction sets can be constructed similarly to the confidence sets. The theory is repetition for those of the estimation of μ. Hence we do not report results on the prediction of y_2 in this paper to keep it simple.

<u>Remark 2</u>: Often in practice, $(r - n) << m$. In such cases, the estimate $\hat{\Sigma}$ in Step 1 is not a consistent estimator for Σ (it is not even full-ranked). However, $\tilde{\sigma}^2 = \frac{1}{m}trace(\hat{\Sigma})$ is consistent for $\bar{\sigma}^2$ and hence its use in Step 2 is valid. The only other use of $\hat{\Sigma}$ is in Step 4 to assess the uncertainty of the estimate $\hat{\mu}$. As we will see in the following sections, we also do not need $\hat{\Sigma}$ to be consistent in the usual sense to assess the uncertainty of $\hat{\mu}$.

<u>Remark 3</u>: The point estimates in Step 3 and 4 are based only on $\bar{y}^i$'s, and they are the same estimates as before when Σ is assumed to be $\sigma^2 I$. We do this because $\bar{y}^i$'s are the sufficient statistics for μ. As in a principal component analysis, the singular value decomposition results in a consistent estimate $\widehat{S}$ since on the k-dimensional subspace S the true variation builds up to the order $O_p(mn)$ while the zero-mean measurement errors only cumulate to a smaller order on this subspace. The estimate $\hat{\mu}$ is consistent because the

consistency of $\widehat{S}$. Since, the estimate $\hat{\Sigma}$ is often inconsistent, it is not used in the estimation for μ. However, for constructing a valid confidence set for $\hat{\mu}$, we do need the estimate $\hat{\Sigma}$.

In the next section, we shall show that $\hat{k}$ and $\widehat{S}$ are consistent estimators for k and S. We then provide analysis of the asymptotic error of $\hat{\mu}$. In Section 4, we provide an asymptotically valid confidence set.

3 Asymptotic analysis for empirical linear model

First we shall derive some properties relating to the distribution of $\hat{\Sigma}$.

Lemma 3.1 *Under model (4), one may write*

$$\hat{\Sigma} = M \frac{\sum_{i=1}^{r-n} e_i e_i'}{r-n} M \tag{17}$$

where e_i's are i.i.d. m-dimensional standard normal vectors. Furthermore, (5) implies

$$\tilde{\sigma}^2 = \frac{1}{m} trace(\widehat{\Sigma}) = \bar{\sigma}^2 + o_p(1) \tag{18}$$

Also, $\widehat{\Sigma}$ is statistically independent of $\bar{Y}$.

Proof of Lemma 3.1: From (4), it is obvious that

$$\hat{\Sigma} = \frac{1}{r-n} M[\sum_{i=1}^{n} \sum_{j=1}^{r_i} (\varepsilon_{(j)}^i - \bar{\varepsilon}^i)(\varepsilon_{(j)}^i - \bar{\varepsilon}^i)']M$$

where $\varepsilon_{(j)}^i$ are standard m-dimensional normal random vectors, and $\bar{\varepsilon}^i = \sum_{j=1}^{r_i} \varepsilon_{(j)}^i / r_i$.

Using standard normal theory, $\sum_{j=1}^{r_i} (\varepsilon_{(j)}^i - \bar{\varepsilon}^i)(\varepsilon_{(j)}^i - \bar{\varepsilon}^i)'$ can be expressed as the sum of $r_i - 1$ terms of $e_i e_i'$, and it is independent of $\bar{\varepsilon}^i$. Hence (17) holds and $\widehat{\Sigma}$ is independent of $\bar{Y}$.

Therefore

$$\begin{aligned}
\tilde{\sigma}^2 &= \tfrac{1}{m} trace(\widehat{\Sigma}) \\
&= \tfrac{1}{m} \sum_{i=1}^{r-n} trace(\tfrac{M e_i e_i' M}{r-n}) \\
&= \tfrac{1}{m} \sum_{i=1}^{r-n} \tfrac{e_i' M^2 e_i}{r-n} \\
&= \tfrac{1}{m} E[e_1' \Sigma e_1] + O_p(\tfrac{1}{m\sqrt{r-n}} \sqrt{Var[e_1' \Sigma e_1]})
\end{aligned}$$

To compute the $E[e_1'\Sigma e_1]$ and $Var[e_1'\Sigma e_1]$, we shall assume without loss of generality that

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{pmatrix}.$$

Otherwise,

$$\Sigma = P' \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{pmatrix} P.$$

Replacing e_1 by Pe_1 leaves the problem unchanged because Pe_1 is also a standard m-dimensional normal random vector.

Hence

$$E[e_1'\Sigma e_1] = E[\sum_{i=1}^m \sigma_i^2 e_{1i}^2] = \sum_{i=1}^m \sigma_i^2 = m\bar{\sigma}^2,$$

and

$$\begin{aligned} Var[e_1'\Sigma e_1] &= E[(\sum_{i=1}^m \sigma_i^2 e_{1i}^2)^2] - (\sum_{i=1}^m \sigma_i^2)^2 \\ &= 2\sum_{i=1}^m \sigma_i^4 \\ &\leq 2\sigma_1^2 \sum_{i=1}^m \sigma_i^2. \end{aligned}$$

Hence $\frac{1}{m\sqrt{r-n}}\sqrt{Var[e_1'\Sigma e_1]} \leq \frac{\sqrt{2}\sigma_1\bar{\sigma}}{\sqrt{m(r-n)}} = o_p(1)$. Consequently $\tilde{\sigma}^2 = \bar{\sigma}^2 + o_p(1)$. □

The above lemma shows that the estimator $\hat{\Sigma}$ is independent of $\bar{Y}$ the sufficient statistic that is used to estimate S. Although the lemma does not provide the consistency of $\hat{\Sigma}$, it does show $\tilde{\sigma}^2 = \frac{1}{m}trace(\widehat{\Sigma})$ to be a consistent estimate for $\bar{\sigma}^2$. Next, we shall use this fact to show that $\hat{k}$ is a consistent estimator for the true dimension k. The reason is that the estimated singular values $\widehat{d}_i$ are consistent estimators for true singular values d_i.

Lemma 3.2 *Assume model (13) and condition (5), then as* $m \to \infty$, $n \to \infty$ *and* $\frac{m}{n} \to c \in [0, \infty]$,

(19) $$|\widehat{d}_i - d_i| \;\leq\; \sigma_1(\sqrt{\tfrac{1}{m}} + \sqrt{\tfrac{1}{n}})(1 + o_p(1)) \qquad \textit{for all } \; i,$$

where $d_i = 0$ *for* $i > k$.

Proof of Lemma 3.2: From (12), we know that the singular values of $\bar{Y}/\sqrt{mn}$ are $\widehat{d}_i$'s, $1 \leq i \leq min(m,n)$. On the other hand, (13) implies

$$\frac{\bar{Y}}{\sqrt{mn}} = SDV + \frac{M\mathcal{E}R}{\sqrt{mn}}$$

By Corollary 8.3.2 in Golub and Van Loan (1989) page 428,

$$\begin{array}{rl} & \max_{i=1,..,min(m,n)} |\widehat{d}_i - d_i| \\ \leq & \|\frac{1}{\sqrt{mn}}\bar{Y} - SDV\| \\ \leq & \frac{1}{\sqrt{mn}}\|M\|\|\mathcal{E}\|\|R\| \\ \leq & \sigma_1(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}})[1 + o_p(1)] \end{array} \tag{20}$$

The last line used the result $\|\mathcal{E}'\mathcal{E}\| = m(1+\sqrt{\frac{n}{m}})^2[1+o(1)]$, which is established in Yin, Bai and Krishnaiah (1988). Hence (19) is proved. □

Lemma 3.2 can be used to prove that $\hat{k}$ is a consistent estimator for k as shown below.

Corollary 3.3

$$P(\widehat{k} = k) \rightarrow 1.$$

Proof of Corollary 3.3:

$$\begin{array}{rcl} P[\widehat{k} > k] & = & P[\widehat{d}_{k+1} > \tilde{\sigma}(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + \delta \cdot ln(min(m,n)))] \\ & \leq & P[\sigma_1(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + o_p(1)) > \bar{\sigma}(1 + o_p(1))(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + \delta \cdot ln(min(m,n)))] \\ & = & P[\sigma_1 > \bar{\sigma}(1 + o_p(1))(1 + \delta \cdot ln(min(m,n)))] \\ & \rightarrow & 0 \end{array}$$

according to condition (5).

$$\begin{array}{rcl} P[\widehat{k} < k] & = & P[\widehat{d}_k \leq \tilde{\sigma}(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + \delta \cdot ln(min(m,n)))] \\ & \leq & P[d_k - \sigma_1(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + o_p(1)) \leq \bar{\sigma}(1 + o_p(1))(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})(1 + \delta \cdot ln(min(m,n)))] \\ & \rightarrow & P[d_k \leq 0] \\ & = & 0. \end{array}$$

□

Due to its consistency, $\widehat{k}$ may be and will be replaced by k without affecting the asymptotic assertions below.

Next, we show that $\widehat{S}$ is a consistent estimator of S, which is the fundamental reason that our procedure works asymptotically.

Lemma 3.4 *Together with the assumptions in Lemma 3.2, conditions (7) and (8) imply that*

$$\left\{ \begin{array}{rcl} V\widehat{V}' & = & I_k + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}), \\ \widehat{S}'S & = & I_k + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}), \end{array} \right. \tag{21}$$

where I_k denotes a $k \times k$ identity matrix.

Proof of Lemma 3.4:

As in (13), let us choose V_0 such that $\begin{pmatrix} V \\ V_0 \end{pmatrix}$ is a $n \times n$ orthogonal matrix. Then we may find a $k \times n$ matrix A such that

$$\widehat{V} = A \begin{pmatrix} V \\ V_0 \end{pmatrix}.$$

Since $\widehat{V}\widehat{V}' = I$, $AA' = I$. Also, since A is a partial matrix of an orthogonal matrix, the sum of squared elements of A in a given column is no greater than one. Let $\widehat{V}'_{(j)}$ denotes the jth row of $\widehat{V}$, i.e. $\widehat{V}'_{(j)} = e'_{(j)}V$, where $e_{(j)}$ is the jth coordinate column vector of dimension k. Note that $\frac{1}{\sqrt{mn}}Y\widehat{V}_{(j)} = d_j\widehat{S}_{(j)}$, where $\widehat{S}_{(j)}$ is the jth column of $\widehat{S}$. Hence

$$\begin{array}{rcl} \widehat{d}_1 & = & \|\frac{Y}{\sqrt{mn}}\widehat{V}_{(1)}\| \\ & = & \|SDV\widehat{V}_{(1)} + \frac{1}{\sqrt{mn}}M\mathcal{E}\widehat{V}_{(1)}\| \\ & \leq & \|SDV\widehat{V}_{(1)}\| + \frac{1}{\sqrt{mn}}\|M\|\|\mathcal{E}\|\|\widehat{V}_{(1)}\| \\ & = & \|SDAe_{(1)}\| + \frac{1}{\sqrt{mn}}\sigma_1\|\mathcal{E}\| \\ & = & \sqrt{a_{11}^2(d_1)^2 + ... + a_{1k}^2(d_k)^2} + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}). \end{array}$$

Since

$$\widehat{d}_1 = d_1 + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}),$$

we have

$$d_1 \leq \sqrt{a_{11}^2(d_1)^2 + ... + a_{1k}^2(d_k)^2} + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}).$$

Note that (8) implies that $\liminf_{m,n\to\infty} d_1 \geq \Delta$ and $\liminf_{m,n\to\infty}(d_i - d_{i+1}) \geq \Delta$. Using this and $a_{11}^2 + a_{12}^2 + ... + a_{1k}^2 \leq a_{11}^2 + a_{12}^2 + ... + a_{1n}^2 = 1$, we conclude

$$a_{11}^2 = 1 + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}),$$

and

$$a_{1i}^2 = O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}) \qquad \text{for } i = 2, ..., k.$$

Similarly we derive

$$d_2 \leq \sqrt{a_{21}^2(d_1)^2 + ... + a_{2k}^2(d_k)^2} + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}).$$

Since $a_{21}^2 \leq 1 - a_{11}^2 = O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}})$ and $d_2 > ... > d_k$,

$$a_{22}^2 = 1 + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}),$$

and

$$a_{2i}^2 = O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}) \qquad \text{for } i = 1, 3, 4, ..., k.$$

By induction,

$$a_{ij}^2 = \delta_{ij} + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}) \qquad \text{for } i = 1, ..., k \text{ and } j = 1, ..., n.$$

where δ_{ij} is 1 if $i = j$ and 0 otherwise. Hence

$$V\widehat{V}' = (I_k\ 0)\, A' = I_k + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}}).$$

By symmetry, the second half of (21),

$$\widehat{S}'S = I_k + O_p(\sqrt{\frac{1}{m}} + \sqrt{\frac{1}{n}})$$

is proved too. □

Note that $A = S$ is the only matrix such that $A'S$ is identity matrix I_k. In this sense the above displayed equation implies that $\widehat{S}$ is a consistent estimator of S. The next lemma provides the asymptotic distribution of the estimate $\widehat{\mu}$.

Theorem 3.5 *We assume that the model (3) also satisfies conditions (7), (8) and (9). Then under the asymptotic (6) and conditions $m/(m-t) = O(1)$, $n = o(m^2)$, we have*

$$\hat{\mu} - \mu = T + o_p(\sqrt{\frac{m}{n}} + \sqrt{\frac{m}{t}}), \tag{22}$$

where

$$T = \frac{1}{\sqrt{n}} M\mathcal{E}RV'D^{-1}\eta + \widehat{S}[\widehat{S}_1'\widehat{S}_1]^{-1}\widehat{S}_1'M_1\varepsilon \tag{23}$$

Proof of Theorem 3.5: From (12), we have

$$\frac{1}{\sqrt{mn}}Y = \widehat{S}\widehat{D}\widehat{V} + \widehat{S}_0\widehat{D}_0\widehat{V}_0.$$

Post-multiplying both sides by $\widehat{V}'$, we have

$$\frac{1}{\sqrt{mn}}Y\widehat{V}' = \widehat{S}\widehat{D}. \tag{24}$$

This implies by (13)

$$SDV\widehat{V}' + \frac{1}{\sqrt{mn}}M\mathcal{E}R\widehat{V}' = \widehat{S}\widehat{D}.$$

Since $(V\widehat{V}')$ is a square matrix of size $k \times k$, we may write with probability one

$$S = (\widehat{S}\widehat{D} - \frac{1}{\sqrt{mn}}M\mathcal{E}R\widehat{V}')(V\widehat{V}')^{-1}D^{-1}. \tag{25}$$

Using (16) and, similar to y, partitioning $M = (M_1', M_2')'$, we may write

$$\begin{aligned} & \hat{\mu} - \mu \\ = \; & \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'(\sqrt{m}S_1\eta + M_1\varepsilon) - \sqrt{m}S\eta \\ = \; & \sqrt{m}[(\widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1' \;,\; 0) - I_m]S\eta + \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon, \end{aligned}$$

where I_m is the $m \times m$ identity matrix and 0 is a zero matrix of size $m \times (m-k)$. Substituting S in the above displayed equation by the expression in (25) and noting

$$[(\widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1' \;,\; 0) - I_m]\widehat{S} = \widehat{S} - \widehat{S} = 0,$$

we establish that $\hat{\mu} - \mu$ equals

$$\begin{aligned} & \sqrt{m}[(\widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1' \; 0) - I_m][-\tfrac{1}{\sqrt{mn}}M\mathcal{E}R\widehat{V}'][V\widehat{V}']^{-1}D^{-1}\eta + \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon \\ = \; & \tfrac{1}{\sqrt{n}}M\mathcal{E}R\widehat{V}'[V\widehat{V}']^{-1}D^{-1}\eta + \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon - \tfrac{1}{\sqrt{n}}\widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}'[V\widehat{V}']^{-1}D^{-1}\eta. \end{aligned} \tag{26}$$

Notice that this is almost (22). To finish the proof, we need to carefully analyze the order of the terms. We summarized some useful facts for the order analysis in Lemma 7.1, which is included and proven in the Appendix.

From (21), $[V\widehat{V}']^{-1} = I_k + o_p(1)$. Using this together with facts (42) and (47), the first term is $\frac{1}{\sqrt{n}}M\mathcal{E}RV'D^{-1}\eta + o_p(\sqrt{\frac{m}{n}})$, which is of order $O_p(\sqrt{\frac{m}{n}})$ by (41). The second term according to (43) is of order $O_p(\sqrt{\frac{m}{t}})$. Hence what is left to be proven is that the remainder term (the third term) is indeed $o_p(\sqrt{\frac{m}{n}} + \sqrt{\frac{m}{t}})$.

The order of the remainder term is

$$\begin{aligned} & \tfrac{1}{\sqrt{n}}O_p(\|\widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}'[V\widehat{V}']^{-1}D^{-1}\eta\|) \\ \leq\ & \tfrac{1}{\sqrt{n}}O_p(\|\widehat{S}\| \cdot \|(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}'\| \cdot \|[V\widehat{V}']^{-1}\| \cdot \|D^{-1}\eta\|) \\ =\ & \tfrac{1}{\sqrt{n}}O_p(\sqrt{\tfrac{m}{t}} + \sqrt{\tfrac{n}{t}} + \sqrt{\tfrac{m}{n}}), \end{aligned}$$

where the last equation follows from (45) and (21).

Multiply out the terms, the above order is

$$O_p(\sqrt{\frac{m}{t}}\frac{1}{\sqrt{n}} + \sqrt{\frac{m}{t}}\frac{1}{\sqrt{m}} + \sqrt{\frac{m}{n}}\frac{1}{\sqrt{n}}) = o_p(\sqrt{\frac{m}{t}} + \sqrt{\frac{m}{n}})$$

This finishes the proof. □

Notice that on the right-hand side of (23), the first term is of order $O(\sqrt{\frac{m}{n}})$ and the second term is of order $O(\sqrt{\frac{m}{t}})$. This is the optimal rate of estimation (Ding 1996, Ph.D thesis at Cornell University).

4 Confidence Sets

Based on the asymptotic theory in the previous section, we can try to construct asymptotically valid confidence sets for μ.

Theorem 4.1 *Under the assumptions of Theorem 3.5, and assume also that $t \leq n$, then*

$$(\hat{\mu} - \mu)'(\hat{\mu} - \mu) - \frac{m}{n}\tilde{\sigma}^2\hat{\eta}'\widehat{D}^{-1}\widehat{V}R^2\widehat{V}'\widehat{D}^{-1}\hat{\eta} \leq c_\alpha^2 \tag{27}$$

is a confidence set for μ *with an asymptotic coverage probability* $(1-\alpha)$. *Here* c_α^2 *(obtained by numerical simulation) denotes the* $(1-\alpha)$ *quantiles of*

$$T_1 = \sum_{i=1}^{k} \widehat{\lambda}_i^2 z_i^2, \tag{28}$$

where $z_1, z_2, ..., z_k$ *are statistically independent standard normal random variables;* $\widehat{\lambda}_i^2$*'s,* $i = 1, ...k$*, are the eigenvalues of* $[\widehat{S}_1'\widehat{S}_1]^{-1}\widehat{S}_1'\widehat{\Sigma}_{11}\widehat{S}_1[\widehat{S}_1'\widehat{S}_1]^{-1}$*, and* $\widehat{\Sigma}_{11}$ *is the* $t \times t$ *sub-matrix of* $\widehat{\Sigma}$ *by removing the last* $(m-t)$ *rows and the last* $(m-t)$ *columns.*

Remark 4: Although ignoring the non-diagonal covariance matrix Σ can result in the same consistent point estimator $\hat{\mu}$, the valid confidence set for $\hat{\mu}$ involves the estimator $\widehat{\Sigma}$. The estimate $\widehat{\Sigma}$ may not be consistent in the standard sense, but it captures the main structure of the covariance matrix Σ, therefore resulting in an asymptotically valid confidence set for $\hat{\mu}$.

Remark 5: Inequality (27) gives a spherical confidence set. As in Ding and Hwang (1999), a more useful confidence set is the cubic confidence set constructed from (22), albeit its asymptotic validity is not established. To describe the cubic confidence set, let $\|a\|_\infty$ denote the L_∞-norm, i.e., $\max_{1\le i\le m} |a_i|$ for a vector $a = (a_1, ..., a_m)'$. A recommended $(1-\alpha)$ cubic confidence set is

$$\|\widehat{\mu} - \mu\|_\infty \le C_\alpha, \tag{29}$$

where C_α is the $(1-\alpha)$ quantile of $\|Z\|_\infty$. Here Z is a m-dimensional normal random vector with mean zero and covariance matrix

$$\frac{\|R\widehat{V}'\widehat{D}\widehat{\eta}\|^2}{n}\widehat{\Sigma} + \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'\widehat{\Sigma}_{11}\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}'.$$

This type of sets have the advantage of providing confidence intervals with the same length for each coordinate of μ.

Remark 6: We may also derive prediction sets for y_2 as in Ding and Hwang (1999).

Remark 7: The extension of the above results to the empirical linear model (Ding and Hwang 1999) is straightforward albeit tedious.

Proof of Theorem 4.1:

Let $q = \sum_{i=1}^{k} \widehat{\lambda}_i^{-1}$. The condition (9) together with facts (43) and (44) implies that q is at least of the same order as m/t. Because $t \leq n$, Theorem 3.5 states that

$$\widehat{\mu} - \mu = \frac{1}{\sqrt{n}} M\mathcal{E}RV'D^{-1}\eta + \widehat{S}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon + o_p(\sqrt{q}).$$

Hence

$$\begin{aligned} & (\widehat{\mu} - \mu)'(\widehat{\mu} - \mu) \\ = \; & \tfrac{1}{n}\eta'D^{-1}VR\mathcal{E}'M^2\mathcal{E}RV'D^{-1}\eta + \varepsilon'M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-2}\widehat{S}_1'M_1\varepsilon + 2\tfrac{1}{\sqrt{n}}\varepsilon'M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}'M\mathcal{E}RV'D^{-1}\eta + o_p(q). \end{aligned}$$

Notice that ε is independent of $\mathcal{E}$ and $\widehat{S}$, hence the order of the third term is

$$O_p(\frac{1}{\sqrt{n}}\|M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}'M\mathcal{E}RV'D^{-1}\eta\|),$$

which is $O_p(\frac{1}{\sqrt{n}}\|\widehat{S}_1\| \cdot \|(\widehat{S}_1'\widehat{S}_1)^{-1}\| \cdot \|\widehat{S}'M\mathcal{E}RV'\|)$. Using (9), (38), (40), (46) and $t \leq n$, this equals

$$O_p(\frac{1}{\sqrt{n}}\sqrt{\frac{t}{m}}\frac{m}{t}(\sqrt{\frac{m}{n}} + 1)) = O_p(\frac{m}{t}\sqrt{\frac{t}{n}}(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})) = o_p(q).$$

From (41), the first term is

$$\bar{\sigma}^2\frac{m}{n}\eta'D^{-1}VR^2V'D^{-1}\eta + o_p(q),$$

which equals

$$\tilde{\sigma}^2\frac{m}{n}\widehat{\eta}'\widehat{D}^{-1}\widehat{V}R^2\widehat{V}'\widehat{D}^{-1}\widehat{\eta} + o_p(q)$$

according to (18), (42) and (47).

Therefore

$$(\widehat{\mu} - \mu)'(\widehat{\mu} - \mu) - \frac{m}{n}\tilde{\sigma}^2\widehat{\eta}'\widehat{D}^{-1}\widehat{V}R^2\widehat{V}'\widehat{D}^{-1}\widehat{\eta} = \varepsilon'M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-2}\widehat{S}_1'M_1\varepsilon + o_p(q).$$

Since ε is distributed as standard normal m-dimensional vector, the first term on the right hand side equals $\sum_{i=1}^{k} \tilde{\lambda}_i^2 e_i^2$ where e_1,..., e_k are i.i.d. standard normal variables and $\tilde{\lambda}_1$,...,$\tilde{\lambda}_k$ are the singular values of $(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1$. According to (44), we can replace $\tilde{\lambda}_i$'s by $\hat{\lambda}_i$'s.

Therefore,

$$[(\widehat{\mu}-\mu)'(\widehat{\mu}-\mu)-\frac{m}{n}\tilde{\sigma}^2\widehat{\eta}'\widehat{D}^{-1}\widehat{V}R^2\widehat{V}'\widehat{D}^{-1}\widehat{\eta}]/q=\sum_{i=1}^{k}a_iz_i^2+o_p(1) \tag{30}$$

where $a_i=\widehat{\lambda}_i^2/q$, $i=1,...,k$.

If we had assumed that λ_i^*'s in (9) have asymptotic limits, it would have been obvious that (30) implies the correct asymptotic coverage of (27). Since we do not make such assumptions, we have to use a more elaborate argument.

Let $T=\sum_{i=1}^{k}a_iz_i^2$, then by definition of c_α,

$$P(T<c_\alpha^2/q)=(1-\alpha). \tag{31}$$

Because of (30), to prove (27), we only need to prove that

$$P(T+o_p(1)<c_\alpha^2/q)=(1-\alpha)+o_1(1). \tag{32}$$

Let us denote $\underline{a}=(a_1,...,a_k)$ and $c=c_\alpha^2/q$. Obviously $a_1+a_2+...+a_k=1$. We write $c(\underline{a})$ to emphasize that c is a function of $a_1,...,a_k$.

To complete the proof, it suffices to show that the absolute difference between the left-hand side of (31) and the left-hand side of (32) approaches zero, i.e.,

$$|P(T+o_p(1)\leq c(\underline{a}))-P(T\leq c(\underline{a}))|\to 0. \tag{33}$$

Now the left-hand side of (33) is bounded above by

$$P(|T(\underline{a})-c(\underline{a})|\leq|o_p(1)|),$$

whose asymptotic limit in turn is bounded above, by the asymptotic limit of

$$P(|T(\underline{a})-c(\underline{a})|\leq\delta)$$

for any $\delta>0$.

Since a_i's depends on m,n and the data as well, we will need the compactness argument below. The asymptotic limit of the last displayed expression is obviously bounded above by

$$\sup_{\underline{a}\in K}P(|T(\underline{a})-c(\underline{a})|\leq\delta), \tag{34}$$

where $K = \{\underline{a} : a_1 + a_2 + ... + a_k = 1; a_i \geq 0, i = 1, .., k.\}$ is a compact set. Therefore, for $\delta_i = \frac{1}{i}$, there exists $\underline{a}^i$ such that the supremum is achieved at $\underline{a} = \underline{a}^i$. Again from the compactness of K, there exists a subsequence of $\{\underline{a}^i\}$ such that it converges to $\underline{a}^\infty \in K$. Without loss of generality, we may call the subsequence $\{\underline{a}^i\}$ itself. Hence the limit of (34) as $\delta \to 0$ equals

$$\lim_{i \to \infty} P(|T(\underline{a}^i) - c(\underline{a}^i)| \leq \frac{1}{i}),$$

by the monotonicity of (34) in δ. Since $\underline{a}^\infty \in K$, $T(\underline{a}^\infty)$ has a nondegenerate distribution. This and the continuity of $T(\underline{a})$ and $c(\underline{a})$ in $\underline{a}$ imply that the last expression equals

$$P(|T(\underline{a}^\infty) - c(\underline{a}^\infty)| \leq 0) = 0. \qquad \square$$

5 Simulation

To study the coverage probability of the proposed intervals, we did some numerical studies. In particular, we want to see if the proposed intervals do improve upon the intervals of Ding and Hwang (1999) when the measurement errors are not i.i.d.

First, we generate a covariance matrix Σ. To study the effect of unequal magnitude of measurement error sources, we generate a diagonal matrix D^* with first three elements as 400, 200, 100, and the rest of elements as ones. (Therefore, there are three sources contributing measurement errors about 10-20 times larger than the rest.) To study the effect of correlated errors, we than randomly generate an orthogonal matrix Q of size $m \times m$. The final covariance matrix is $\Sigma = QD^*Q'$.

Samples were generated from the model (3), where Σ is fixed as above and $k = 3$. Each vector is assumed to be measured twice, i.e., $r_1 = r_2 = ... = r_n = 2$. The HELP procedure is then applied on the sample. The proposed intervals and intervals of Ding and Hwang (1999) are calculated and their coverage recorded. Then the process was repeated by generating 400 such samples, and the empirical coverage probabilities were calculated from the 400 samples. We simulated for various sizes of m, n and t. The simulation in Ding and Hwang (1999) shows that generally the asymptotic works well for m and n both bigger than 40. Therefore, we chose to simulate values of $m = 50, 100, 500$, $n = 50, 100, 500$. We did not simulate for m and n values greater than 500 because that would take too much running time. The simulation was coded in Matlab running on a computer with 525MHz EV6 Compaq Alpha processors. Table 1 reports the summarized empirical coverage probabilities for various values of m and n, and $t = 30$. The results are similar for different t values so that we only reported for the case of $t = 30$.

Table 1: Coverage probabilities for the proposed new confidence sets and previous confidence region for i.i.d. errors (Ding & Hwang 1999) at nominal level 95% and 90% (in the parenthesis).

		Proposed		Ding & Hwang (1999)	
m	n	Spherical	Cube-shaped	Spherical	Cube-shaped
50	50	91.50% (85.50%)	93.25% (85.75%)	20.25% (16.00%)	41.00% (32.75%)
	100	95.50% (89.75%)	96.25% (90.25%)	19.75% (16.25%)	44.25% (36.25%)
	500	95.25% (89.25%)	95.25% (89.50%)	19.00% (14.00%)	43.25% (34.50%)
100	50	93.00% (86.75%)	94.50% (86.50%)	14.75% (12.00%)	30.00% (24.25%)
	100	93.00% (88.25%)	93.50% (88.50%)	20.50% (17.00%)	42.50% (35.00%)
	500	95.00% (88.50%)	93.75% (88.25%)	18.00% (15.50%)	39.50% (31.25%)
500	50	92.75% (87.75%)	93.75% (89.00%)	89.25% (83.75%)	97.50% (94.75%)
	100	94.00% (89.50%)	94.75% (90.25%)	85.00% (76.50%)	98.25% (96.25%)
	500	96.00% (89.25%)	96.50% (91.00%)	85.75% (76.50%)	99.50% (98.25%)

As we can see from the Table 1, the coverage probabilities of the newly proposed confidence sets are close to their nominal levels, while the coverage probabilities of the previous confidence sets (Ding and Hwang, 1999) assuming i.i.d. errors vary wildly. Generally, the previous sets have very low coverage probabilities because they ignore the deviation of error distributions from i.i.d. assumptions. Since we let three main sources of measurement error to be 10-20 times larger than the rest, the deviation from equal variance is bigger for smaller m. As m increases, the measurement errors looks more like from equal variance sources, and the coverage probabilities for the previous confidence sets increase. However, the coverage probabilities for the previous confidence sets are not converging to the nominal level in such cases. In some cases ($m = n = 500$ for the cubic-shaped sets), it is clear that they have coverage probabilities far higher than the nominal level.

The confidence sets proposed in this paper, however, do have correct coverage probabilities in all cases, even in the cases where $\hat{\Sigma}$ is inconsistent.

6 Conclusion and Discussions

In this paper, we generalized the HELP technique to cases with dependent error structure. It is shown that the HELP estimator $\hat{\mu}$ based on mean measurements of the devices is consistent when the measurement errors do not come from a single source of infinite variance. However, ignoring the error structure, the confidence sets for $\hat{\mu}$ are no longer valid. We estimated the error covariance matrix Σ from the repeated measurements of devices. It is shown that valid confidence sets for $\hat{\mu}$ can be constructed based on the estimated error covariance matrix $\hat{\Sigma}$ when the average error magnitude $\bar{\sigma}^2$ can be consistently estimated. This requires a large number of repeated measurements but can be much smaller than the number m required for a consistent estimation of Σ. Simulation studies were conducted and the results confirmed our asymptotic theory.

7 Appendix

Proof of Theorem 2.1:

Professor Hans R. Künsch communicated to us that Theorem 2.1 holds for any second-order stationary time series with a bounded spectral density. He also informed us the following proof which is much simpler than our original proof.

It is an easy corollary of Theorem 3.1.1 and Theorem 4.4.2 of Brockwell and Davis (1987) that the above process is second-order stationary with a bounded spectral density. Let $R(h)$ denote the autocovariance function and let $f(\lambda)$ be the spectral density. Then the covariance matrix of $\varepsilon = (e_1, e_2, ..., e_m)'$ is $\Sigma = (R(i-j); 1 \le i, j \le m)$. Thus the trace of Σ equals to

$mR(0)$, hence $\bar{\sigma} = \sqrt{R(0)}$. Moreover, for any m-dimensional vector α,

$$\begin{array}{ll} \alpha'\Sigma\alpha & = \int |\sum_{j=1}^{m} \alpha_j e^{ij\lambda}|^2 f(\lambda) d\lambda \\ & \leq sup[f(\lambda)] \int |\sum_{j=1}^{m} \alpha_j e^{ij\lambda}|^2 d\lambda \\ & = 2\pi sup[f(\lambda)] \alpha'\alpha. \end{array}$$

Thus $\sigma_1^2 \leq 2\pi sup[f(\lambda)]$. This completes the proof. □

Lemma 7.1 *Under conditions of Theorem 3.5,*

$$\|\mathcal{E}\| = O_p(\sqrt{m} + \sqrt{n}) \tag{35}$$

$$\|M_1\mathcal{E}\| = O_p(\sqrt{t} + \sqrt{n}) \tag{36}$$

$$\widehat{S} = S[I_k + o_p(1)] + \frac{1}{\sqrt{mn}} M\mathcal{E}R\widehat{V}'\widehat{D}^{-1}, \tag{37}$$

$$\widehat{S}_1 = S_1[I_k + o_p(1)] + \frac{1}{\sqrt{mn}} M_1\mathcal{E}R\widehat{V}'\widehat{D}^{-1} = S_1 + o_p(\sqrt{\frac{t}{m}}), \tag{38}$$

$$\widehat{V} = [I_k + o_p(1)]V + \frac{1}{\sqrt{mn}} \widehat{D}^{-1}\widehat{S}'M\mathcal{E}R, \tag{39}$$

$$(\widehat{S}_1'\widehat{S}_1)^{-1} = (S_1'S_1)^{-1} + o_p(\frac{m}{t}), \tag{40}$$

$$VR\mathcal{E}'M^2\mathcal{E}RV' = m\bar{\sigma}^2 VR^2V' + o_p(m) \tag{41}$$

$$M\mathcal{E}R(\widehat{V} - V)' = o_p(\sqrt{m}) \tag{42}$$

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'\Sigma_1 M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} = (S_1'S_1)^{-1}S_1'M_1M_1'S_1(S_1'S_1)^{-1} + o_p(\frac{m}{t}) \tag{43}$$

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'\widehat{M}_1\widehat{M}_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} = (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} + o_p(\frac{m}{t}) \tag{44}$$

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}' = O_p(\sqrt{\frac{m}{t}} + \sqrt{\frac{n}{t}} + \sqrt{\frac{m}{n}}) \tag{45}$$

$$\widehat{S}'M\mathcal{E}RV' = O_p(\sqrt{\frac{m}{n}} + 1) \tag{46}$$

$$\widehat{D}^{-1}\widehat{\eta} = D^{-1}\eta + o_p(1). \tag{47}$$

Proof of Lemma 7.1:

(a). $\|\mathcal{E}\|^2/m$ is the largest eigenvalue of the matrix $\mathcal{E}'\mathcal{E}/m$, which is asymptotically $\sigma^2(1+\sqrt{\frac{n}{m}})^2$ by Yin, Bai and Krishnaiah (1988). Therefore $\|\mathcal{E}\|^2 = O_p((\sqrt{m}+\sqrt{n})^2)$. Hence (35) is established.

(b). Use a singular value decomposition and write $M_1 = P\Lambda Q$, where P, Λ and Q are of sizes $t \times t$, $t \times t$ and $t \times m$ respectively, $PP' = QQ' = I_t$, and Λ is a diagonal matrix.

Then $Q\mathcal{E}$ is a $t \times n$ matrix with i.i.d. entries following standard normal distribution. Hence $\|Q\mathcal{E}\| = O_p(\sqrt{t} + \sqrt{n})$ similar to (35).

Therefore, $\|M_1\mathcal{E}\| \leq \|\Lambda\| \cdot \|Q\mathcal{E}\| \leq \sigma_1 \|Q\mathcal{E}\| = O_p(\sqrt{t} + \sqrt{n})$.

(c). Combine (12) and (13), we have

$$\widehat{S}\widehat{D}\widehat{V} + \widehat{S}_0\widehat{D}_0\widehat{V}_0 = \frac{\bar{Y}}{\sqrt{mn}} = SDV + \frac{M\mathcal{E}R}{\sqrt{mn}}. \tag{48}$$

Post-multiplying by $\widehat{V}'$ on both sides, we establish

$$\widehat{S}\widehat{D} = SDV\widehat{V}' + \frac{M\mathcal{E}R\widehat{V}'}{\sqrt{mn}}.$$

Hence

$$\begin{aligned} \widehat{S} &= SDV\widehat{V}'\widehat{D}^{-1} + \tfrac{1}{\sqrt{mn}} M\mathcal{E}R\widehat{V}'\widehat{D}^{-1} \\ &= S(I_k + o_p(1)) + \tfrac{1}{\sqrt{mn}} M\mathcal{E}R\widehat{V}'\widehat{D}^{-1}. \end{aligned}$$

The last line is because $DV\widehat{V}'\widehat{D}^{-1} = I_k + o_p(1)$ by Theorem 3.2 and Theorem 3.4. Thus (37) is established.

(d). Taking the first t rows of both sides of equation (37) yields

$$\widehat{S}_1 = S_1[I_k + o_p(1)] + \frac{1}{\sqrt{mn}} M_1\mathcal{E}R\widehat{V}'\widehat{D}^{-1}.$$

Notice that

$$\frac{1}{\sqrt{mn}}\|M_1\mathcal{E}R\widehat{V}'\widehat{D}^{-1}\| \leq \frac{1}{\sqrt{mn}}\|M_1\mathcal{E}\| \cdot \|R\| \cdot \|\widehat{D}^{-1}\|,$$

which by (36) is of order

$$O_p(\frac{1}{\sqrt{mn}}(\sqrt{t} + \sqrt{n})) = O_p(\sqrt{\frac{t}{m}}(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{t}})) = o_p(\sqrt{\frac{t}{m}})$$

Hence $\widehat{S}_1 = S_1 + o_p(\sqrt{\frac{t}{m}})$.

(e). (39) is derived similarly to (37) by pre-multiplying (48) by $\widehat{S}'$ and using the fact $\widehat{D}^{-1}\widehat{S}'SD = I_k + o_p(1)$ from Theorem 3.2 and Theorem 3.4.

(f). (38) implies that

$$\widehat{S}_1'\widehat{S}_1 = S_1'S_1 + o_p(\frac{t}{m}).$$

This together with (9) implies

$$(\widehat{S}_1'\widehat{S}_1)^{-1} = (S_1'S_1)^{-1} + o_p(\frac{m}{t}).$$

(g). Conduct a singular value decomposition on RV', we have

$$RV' = P\Lambda Q,$$

where P, Λ and Q are of sizes $n \times k$, $k \times k$ and $k \times k$ respectively, $P'P = Q'Q = I_k$, and Λ is a diagonal matrix.

Hence $VR^2V' = Q'\Lambda^2 Q$.

On the other hand, $\mathcal{E}P$ is a $m \times k$ matrix with i.i.d. entries that comes from standard normal distribution. Hence Lemma 7.2 (whose proof is at the end of Appendix) states that

$$P'\mathcal{E}'M^2\mathcal{E}P = m\bar{\sigma}^2 I_k + o_p(m).$$

This implies that

$$\begin{aligned} VR\mathcal{E}'M^2\mathcal{E}RV' &= Q'\Lambda P'\mathcal{E}'M^2\mathcal{E}P\Lambda Q \\ &= m\bar{\sigma}^2 Q'\Lambda^2 Q + o_p(m). \end{aligned}$$

Hence (41) is proven.

(h). Using (39),

$$M\mathcal{E}R(\widehat{V} - V)' = M\mathcal{E}RV'o_p(1) + \frac{1}{\sqrt{mn}}M\mathcal{E}R^2\mathcal{E}'M\widehat{S}'\widehat{D}^{-1}.$$

According to (41), the first term is of order $o_p(\sqrt{m})$.

The second term is bounded by $\frac{1}{\sqrt{mn}}\|M\|\|\mathcal{E}\|\|R^2\|\|\mathcal{E}'\|\|M\|\|\widehat{S}'\|\|\widehat{D}^{-1}\|$, which is of order

$$O_p(\frac{1}{\sqrt{mn}}\|\mathcal{E}\|^2) = O_p(\sqrt{\frac{m}{n}} + \sqrt{\frac{n}{m}}) = o_p(\sqrt{m})$$

according to (35), (19) and $n = o(m^2)$.

(i). (43) follows directly from (38), (40) and (9).

(j). Taking the first t rows of $\widehat{M}$, (17) implies

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'\widehat{M}_1\widehat{M}_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} = (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\frac{\sum_{i=1}^{r-n} e_ie_i'}{r-n}M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1}.$$

Conduct a singular value decomposition,

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1 = P\Lambda Q,$$

where P, Λ and Q are of sizes $k\times k$, $k\times k$ and $k\times m$, $PP' = QQ' = I_k$, and Λ is a diagonal matrix.

Since e_i's are independent of $\widehat{S}_1$, condition on $\widehat{S}_1$, Qe_i are i.i.d. standard k-dimensional normal random vectors. Hence $Q\frac{\sum_{i=1}^{r-n} e_ie_i'}{r-n}Q' = I_k + o_p(1)$. Therefore,

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'\widehat{M}_1\widehat{M}_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} = P\Lambda^2P' + o_p(\frac{m}{t}) = (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1M_1'\widehat{S}_1(\widehat{S}_1'\widehat{S}_1)^{-1} + o_p(\frac{m}{t}).$$

(k). Using (38) and (39),

$$\begin{aligned}
& (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}' \\
= \ & (\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}RV'(I_k+o_p(1)) + \tfrac{1}{\sqrt{mn}}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R^2\mathcal{E}M\widehat{S}\widehat{D}^{-1} \\
= \ & (\widehat{S}_1'\widehat{S}_1)^{-1}(I_k+o_p(1))S_1'M_1\mathcal{E}RV'(I_k+o_p(1)) + \tfrac{1}{\sqrt{mn}}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{D}^{-1}\widehat{V}R\mathcal{E}'M_1'M_1\mathcal{E}RV'(I_k+o_p(1)) \\
& + \tfrac{1}{\sqrt{mn}}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R^2\mathcal{E}M\widehat{S}\widehat{D}^{-1} \\
= \ & O_p(\|(\widehat{S}_1'\widehat{S}_1)^{-1}S_1'M_1\mathcal{E}RV'\|) + O_p(\tfrac{1}{\sqrt{mn}}\|(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{D}^{-1}\widehat{V}R\mathcal{E}'M_1'M_1\mathcal{E}RV'\|) \\
& + O_p(\tfrac{1}{\sqrt{mn}}\|(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R^2\mathcal{E}M\widehat{S}\widehat{D}^{-1}\|) \\
= \ & O_p(\tfrac{m}{t}\|S_1'M_1\mathcal{E}RV'\|) + O_p(\tfrac{1}{\sqrt{mn}}\tfrac{m}{t}\|\mathcal{E}'M_1'\|\cdot\|M_1\mathcal{E}RV'\|) + O_p(\tfrac{1}{\sqrt{mn}}\sqrt{\tfrac{m}{t}}\|M_1\mathcal{E}\|\cdot\|\mathcal{E}\|)
\end{aligned}$$

$S_1'M_1\mathcal{E}RV'$ is a $k\times k$ matrix. Each of its entries is a zero mean normal random variable. The variance of each entry is at most $\|S_1'M_1\|^2\cdot\|RV'\|^2 = O_p(\frac{t}{m})$. Hence $\|S_1'M_1\mathcal{E}RV'\| = O_p(\sqrt{\frac{t}{m}})$.

Similar to Lemma 7.2, $\|M_1\mathcal{E}RV'\| = O_p(\sqrt{t})$.

Using the above two facts together with (35) and (36), we calculate the order of $(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\mathcal{E}R\widehat{V}'$ as

$$\begin{aligned}
& O_p(\tfrac{m}{t}\sqrt{\tfrac{t}{m}}) + O_p(\tfrac{1}{\sqrt{mn}}\tfrac{m}{t}(\sqrt{t}+\sqrt{n})\sqrt{t}) + O_p(\tfrac{1}{\sqrt{mn}}\sqrt{\tfrac{m}{t}}(\sqrt{t}+\sqrt{n})(\sqrt{m}+\sqrt{n})) \\
= \ & O_p(\sqrt{\tfrac{m}{t}}) + O_p(\sqrt{\tfrac{m}{n}}+\sqrt{\tfrac{m}{t}}) + O_p(\sqrt{\tfrac{m}{n}}+1+\sqrt{\tfrac{m}{t}}+\sqrt{\tfrac{n}{t}}) \\
= \ & O_p(\sqrt{\tfrac{m}{t}}+\sqrt{\tfrac{m}{n}}+\sqrt{\tfrac{n}{t}})
\end{aligned}$$

(l). Using (37), (41) and (42), we have

$$\begin{aligned}
&\widehat{S}'M\mathcal{E}RV' \\
=\ & [I_k + o_p(1)]S'M\mathcal{E}RV' + \tfrac{1}{\sqrt{mn}}\widehat{D}^{-1}\widehat{V}R\mathcal{E}'M^2\mathcal{E}RV' \\
=\ & [I_k + o_p(1)]S'M\mathcal{E}RV' + \tfrac{1}{\sqrt{mn}}\widehat{D}^{-1}[m\bar{M}^2VR^2V' + o_p(m)] \\
=\ & O_p(\|S'M\mathcal{E}RV'\|) + O_p(\sqrt{\tfrac{m}{n}}).
\end{aligned}$$

Again, $S'M\mathcal{E}RV'$ is a $k \times k$ matrix whose entries are zero mean random variables. The variance of each entry is at most $\|S'M\| \cdot \|RV'\| \leq \sigma_1^2$. Hence

$$\|S'M\mathcal{E}RV'\| = O_p(1).$$

This completes the proof for (46).

(m). First, $\widehat{D}^{-1}\widehat{\eta} = \widehat{D}^{-1}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'S_1\eta + \frac{1}{\sqrt{m}}\widehat{D}^{-1}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon$.

According to (38) and (40),

$$(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'S_1 = (S_1'S_1)^{-1}S_1'S_1 + o_p(1) = I_k + o_p(1).$$

This and Theorem 3.2 implies that

$$\widehat{D}^{-1}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'S_1\eta = D\eta + o_p(1).$$

Because ε is independent of $\widehat{D}$ and $\widehat{S}_1$, the remainder term $\frac{1}{\sqrt{m}}\widehat{D}^{-1}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\varepsilon$ is of order $\frac{1}{\sqrt{m}}\|\widehat{D}^{-1}(\widehat{S}_1'\widehat{S}_1)^{-1}\widehat{S}_1'M_1\|$, which according to (43) is $O_p(\frac{1}{\sqrt{m}}\sqrt{\frac{m}{t}}) = o_p(1)$. □

Lemma 7.2 *Suppose the entries of a $m \times k$ matrix E are i.i.d. standard normal variables. Then*

$$E'\Sigma^2E = m\bar{\sigma}^2I_k + O_p(\sqrt{m}).$$

Proof of Lemma 7.2: Since Σ is a covariance matrix, it can be written as

$$P'\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{pmatrix}P,$$

where P is an orthogonal matrix of size $m \times m$.

Hence PE is also an $m \times k$ matrix with i.i.d. entries of standard normal distribution. Therefore without loss of generality, we can assume that

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{pmatrix}$$

here.

Let e_{ij} and a_{ij} denote the entries of E and $A = E'\Sigma E$ respectively. Obviously,

$$a_{ij} = \sum_{\ell=1}^{m} \sigma_\ell^2 e_{\ell i} e_{\ell j} \qquad i = 1, ..., k, \; j = 1, ..., k.$$

Hence

$$E(a_{ij}) = \begin{cases} m\bar{\sigma}^2 & i = j \\ 0 & i \neq j \end{cases} \qquad \text{and} \qquad Var(a_{ij}) = \begin{cases} 2\sum_{\ell=1}^{m} \sigma_\ell^4 & i = j \\ \sum_{\ell=1}^{m} \sigma_\ell^4 & i \neq j \end{cases}$$

Therefore

$$a_{ij} = \begin{cases} m\bar{\sigma}^2 + O_p(\sqrt{m}) & i = j \\ 0 + O_p(\sqrt{m}) & i \neq j \end{cases}$$

Since k is fixed, this means that

$$E'\Sigma^2 E = m\bar{\sigma}^2 I_k + O_p(\sqrt{m}).$$

□

References

Bai, Z. D. and Yin, Y. Q. (1993). Limit of smallest eigenvalue of a large dimensional sample covariance matrix. *Annals of Probability* **21** 1275-1294.

Brockwell, P.J. and Davis, R.A. (1987). Time Series: Theory and Methods. Springer-Verlag, New York.

Ding, A. A. and Hwang, J. T. G. (1999). Prediction Intervals, Factor Analysis Models and HELP. *Journal of American Statistical Association* **94** 446-455.

Golub, G.H. and Van Loan, C.F. (1989). Matrix Computations. 2nd ed. Johns Hopkins University Press.

Hwang, J. T. G. and Liu, H. K. (1995a). Asymptotic Theory for Prediction through Empirical Linear Models. *Spring Research Conference Proceedings* 207-216.

Hwang, J. T. G. and Liu, H. K. (1995b). Application of Empirical Linear Prediction. *Proceedings of the 1994 Joint Statistical Meetings, Section of Physical and Engineering Sciences* 175-180.

Souders, T. M. and Stenbakken, G. N. (1987). Test Point Selection and Testability Measures via QR Factorization of Linear Models. *IEEE Trans. Instum. Meas.* **IM-36** 406-410.

Souders, T. M. and Stenbakken, G. N. (1991). Cutting the High Cost of Testing. *IEEE Spectrum* March 1991, p 48-51.

Souders, T. M. and Stenbakken, G. N. (1994). Developing Linear Error Models for Analog Devices. *IEEE Transaction on Instrumentation and Measurement* **43** 2, 157-163.

Stenbakken, G. N. (1996). Effects of Nonmodel Errors on Model-based Testing. *IEEE Transaction on Instrumentation and Measurement* **45** 2, 384-388.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields* **78** 509-521.