한국통계학회논문집 제2권 1호
1995년 4월 pp. 1-12

On Fitting Polynomial Measurement Error Models with Vector Predictor[1)]
- When Interactions Exist among Predictors -

Myung-Sang Moon[2)]

Abstract

An estimator of coefficients of polynomial measurement error model with vector predictor and first-order interaction terms is derived using Hermite polynomial. Asymptotic normality of estimator is provided and some simulation study is performed to compare the small sample properties of derived estimator with those of OLS estimator.

1. Introduction

In traditional regression with error-free predictors, polynomial model is a linear model. But it belongs to a nonlinear model in measurement error model(MEM), and it is not easy to derive consistent estimators of parameters due to the power terms of the error-free predictors. Some consistent estimation methods for the polynomial MEM were proposed in recent years. They include Wolter & Fuller(1982) for the quadratic functional model, Chan & Mak(1985), Stefanski(1985, 1989) and Moon & Gunst(1995) for the k-th order polynomial MEM. Their results were obtained without assuming decreasing error variances. Moon and Gunst(1994) derived estimator for the k-th order polynomial MEM assuming decreasing error variance and showed that it is asymptotically normally distributed. They also presented small-sample simulation results for cubic model and showed that their estimator possesses better properties than those of OLS estimator in MSE sense, especially when sample size is relatively large.

All the estimation methods mentioned above were derived for the polynomial MEM with one predictor and one response variable. Since polynomial MEM with vector predictor is more practical one and is of more interest, it deserves to be studied. In this paper, estimation method for the polynomial MEM with vector predictor is investigated. To make it more practical, first-order interaction terms among predictors are included in the model. In Section 2, notation and some calculation results needed to define an estimator are introduced. Basically, notation used in this paper is similar to that of Moon and Gunst(1994). The model contains more terms and the dimension of vectors and matrices is changed accordingly.

1) This work was supported by a Yonsei University Faculty Research Grant for 1993.
2) Department of Statistics, Yonsei University, Wonju-Kun, Kangwon-Do, 222-701, KOREA.

However, this minor change results in much more complex calculation in deriving an estimator. Derived estimator and its properties are presented in Section 3. Simulation results comparing the small-sample properties of derived estimator with those of OLS estimator, and some concluding remarks are reported in Section 4.

2. Model and Estimator

Polynomial MEM studied in this work is no-equation-error functional model containing $p(p \geq 2)$ predictors and first-order interaction terms among them. However, the model with 2 predictors is enough since the results of that model can be easily extended to the model with $p(p \geq 3)$ predictors. Therefore, p is set equal to 2 in this paper. The specific form of the model investigated is given as follows:

$$\begin{aligned}\psi_i &= \beta_0 + \beta_{11}\pi_{1i} + \beta_{12}\pi_{1i}^2 + \beta_{13}\pi_{1i}^3 + \cdots + \beta_{1k}\pi_{1i}^k \\ &\quad + \beta_{21}\pi_{2i} + \beta_{22}\pi_{2i}^2 + \beta_{23}\pi_{2i}^3 + \cdots + \beta_{2l}\pi_{2i}^l \\ &\quad + \beta_3\pi_{1i}\pi_{2i}, \\ &= \boldsymbol{\pi}_i^t\boldsymbol{\beta}, \qquad i = 1, 2, 3, \cdots, n, \end{aligned} \tag{2.1}$$

where

$$\boldsymbol{\pi}_i = (1,\ \pi_{1i},\ \pi_{1i}^2,\ \pi_{1i}^3, \cdots, \pi_{1i}^k,\ \pi_{2i},\ \pi_{2i}^2,\ \pi_{2i}^3, \cdots, \pi_{2i}^l,\ \pi_{1i}\pi_{2i})^t,$$

$$\boldsymbol{\beta} = (\beta_0,\ \beta_{11},\ \beta_{12},\ \beta_{13}, \cdots, \beta_{1k},\ \beta_{21},\ \beta_{22},\ \beta_{23}, \cdots, \beta_{2l},\ \beta_3)^t.$$

Note that in this model, an unobservable response variable ψ_i is related to an unobservable nonstochastic predictors π_{1i} and π_{2i} through a k-th and l-th order polynomial equation respectively. Also first-order interaction term $\pi_{1i}\pi_{2i}$ is included. Bold-face letters denote vectors or matrices and all vectors are column ones in this work. As mentioned, true response and true predictor variables are unobservable ones and they are contaminated as follows due to measurement errors:

$$y_i = \psi_i + v_i, \qquad x_{1i} = \pi_{1i} + u_{1i}, \qquad x_{2i} = \pi_{2i} + u_{2i}.$$

We observe only $\boldsymbol{z}_i = (y_i, x_{1i}, x_{2i})^t$; i.e., $\boldsymbol{z}_i = \boldsymbol{\xi}_i + \boldsymbol{w}_i$ with $\boldsymbol{\xi}_i = (\psi_i, \pi_{1i}, \pi_{2i})^t$ denoting the vector of error-free true variates and $\boldsymbol{w}_i = (v_i, u_{1i}, u_{2i})^t$ the vector of measurement errors. The vector of measurement errors $\boldsymbol{w}_i$ are assumed to be i.i.d. $N(\mathbf{0}, \boldsymbol{\Sigma}_{ww})$ with known covariance matrix

$$\boldsymbol{\Sigma}_{ww} = \begin{bmatrix} \sigma_{vv} & \sigma_{v1} & \sigma_{v2} \\ \sigma_{1v} & \sigma_{11} & \sigma_{12} \\ \sigma_{2v} & \sigma_{21} & \sigma_{22} \end{bmatrix} .$$

Let $H_m(Z)$ be the m-th Hermite polynomial in normal random variate Z with mean μ and variance σ^2, and let's define $P_m(Z) = \sigma^m H_m(Z/\sigma)$. Then, we have

$$E\{P_m(Z)\} = \mu^m . \tag{2.2}$$

Let's define $\boldsymbol{p}_i$ as an unbiased estimator of $\boldsymbol{\pi}_i$ obtained using $P_m(Z)$ and (2.2), and $\boldsymbol{f}_i$ as a vector of deviations from the corresponding powers of the error-free predictors. That is,

$$\begin{aligned} \boldsymbol{p}_i &= \boldsymbol{\pi}_i + \boldsymbol{f}_i \\ &= (1, P_1(x_{1i}), P_2(x_{1i}), \cdots, P_k(x_{1i}), P_1(x_{2i}), P_2(x_{2i}), \cdots, P_l(x_{2i}), P_1(x_{1i})P_1(x_{2i}) - \sigma_{12})^t, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{f}_i = (0,\ & P_1(x_{1i}) - \pi_{1i}, P_2(x_{1i}) - \pi_{1i}^2, \cdots, P_k(x_{1i}) - \pi_{1i}^k, \\ & P_1(x_{2i}) - \pi_{2i}, P_2(x_{2i}) - \pi_{2i}^2, \cdots, P_l(x_{2i}) - \pi_{2i}^l, P_1(x_{1i})P_1(x_{2i}) - \sigma_{12} - \pi_{1i}\pi_{2i})^t. \end{aligned}$$

Also define $(k+l+3)$ x 1 vectors $\boldsymbol{\varsigma}_i$, $\boldsymbol{r}_i$ and $\boldsymbol{g}_i$ by

$$\boldsymbol{\varsigma}_i = (\psi_i, \boldsymbol{\pi}_i^{\,t})^t, \qquad \boldsymbol{r}_i = (v_i, \boldsymbol{f}_i^{\,t})^t \text{ and } \qquad \boldsymbol{g}_i = \boldsymbol{\varsigma}_i + \boldsymbol{r}_i = (y_i, \boldsymbol{p}_i^{\,t})^t.$$

The symmetric $(k+l+3)$ x $(k+l+3)$ covariance matrix of $\boldsymbol{r}_i$ (vector of measurement errors of $\boldsymbol{g}_i$) is necessary to derive an estimator of $\boldsymbol{\beta}$, and it is obtained through tedious calculations.

$$\boldsymbol{\Omega}_i = \begin{bmatrix} \Omega_{11i} & 0 & \boldsymbol{\Omega}_{13i} & \boldsymbol{\Omega}_{14i} & \Omega_{15i} \\ & 0 & \mathbf{0} & \mathbf{0} & 0 \\ & & \boldsymbol{\Omega}_{33i} & \boldsymbol{\Omega}_{34i} & \Omega_{35i} \\ & & & \boldsymbol{\Omega}_{44i} & \Omega_{45i} \\ & & & & \Omega_{55i} \end{bmatrix},$$

where $\Omega_{11i} = Cov(v_i, v_i) = \sigma_{vv}$,

$$\boldsymbol{\Omega}_{13i} = Cov(v_i, P_j(x_{1i}) - \pi_{1i}^j) = j\pi_{1i}^{j-1}\sigma_{v1}, \quad j = 1, 2, \cdots, k,$$

$$\boldsymbol{\Omega}_{14i} = Cov(v_i, P_j(x_{2i}) - \pi_{2i}^j) = j\pi_{2i}^{j-1}\sigma_{v2}, \quad j = 1, 2, \cdots, l,$$

$$\Omega_{15i} = Cov(v_i, x_{1i}x_{2i} - \sigma_{12} - \pi_{1i}\pi_{2i}) = \pi_{1i}\sigma_{v2} + \pi_{2i}\sigma_{v1},$$

$$\boldsymbol{\Omega}_{35i} = m\pi_{1i}^{m-1}\pi_{2i}\sigma_{11} + m\pi_{1i}^{m}\sigma_{12} + m(m-1)\pi_{1i}^{m-2}\sigma_{11}\sigma_{12}, \quad m = 1, 2, \cdots, k,$$

$$\boldsymbol{\Omega}_{45i} = m\pi_{2i}^{m-1}\pi_{1i}\sigma_{22} + m\pi_{2i}^{m}\sigma_{12} + m(m-1)\pi_{2i}^{m-2}\sigma_{22}\sigma_{12}, \quad m = 1, 2, \cdots, l,$$

$$\Omega_{55i} = \pi_{1i}^2\sigma_{22} + \pi_{2i}^2\sigma_{11} + 2\pi_{1i}\pi_{2i}\sigma_{12} + \sigma_{12}^2 + \sigma_{11}\sigma_{22}.$$

$\boldsymbol{\Omega}_{33i}$ and $\boldsymbol{\Omega}_{44i}$ are obtained following the same step presented in Moon & Gunst(1994). The remaining element of $\boldsymbol{\Omega}_i$, that is $\boldsymbol{\Omega}_{34i}$, is derived as follows.

i) $Cov\{P_{2m-1}(x_{1i}) - \pi_{1i}^{2m-1},\ P_{2n-1}(x_{2i}) - \pi_{2i}^{2n-1}\}$

$$= \sum_{j=1}^{m}\sum_{c=1}^{n}(-1)^{m+n-(j+c)} \frac{(2m-1)!\,(2n-1)!\,\sigma_{11}^{m-j}\sigma_{22}^{n-c}E(x_{1i}^{2j-1}x_{2i}^{2c-1})}{(2j-1)!\,(m-j)!\,(2c-1)!\,(n-c)!\,2^{m+n-(j+c)}},$$

ii) $Cov\{P_{2m}(x_{1i}) - \pi_{1i}^{2m},\ P_{2n}(x_{2i}) - \pi_{2i}^{2n}\}$

$$= \sum_{j=0}^{m}\sum_{c=0}^{n}(-1)^{m+n-(j+c)} \frac{(2m)!\,(2n)!\,\sigma_{11}^{m-j}\sigma_{22}^{n-c}E(x_{1i}^{2j}x_{2i}^{2c})}{(2j)!\,(m-j)!\,(2c)!\,(n-c)!\,2^{m+n-(j+c)}},$$

iii) $Cov\{P_{2m-1}(x_{1i}) - \pi_{1i}^{2m-1},\ P_{2n}(x_{2i}) - \pi_{2i}^{2n}\}$

$$= \sum_{j=1}^{m}\sum_{c=0}^{n}(-1)^{m+n-(j+c)} \frac{(2m-1)!\,(2n)!\,\sigma_{11}^{m-j}\sigma_{22}^{n-c}E(x_{1i}^{2j-1}x_{2i}^{2c})}{(2j-1)!\,(m-j)!\,(2c)!\,(n-c)!\,2^{m+n-(j+c)}}.$$

$\boldsymbol{\Omega}_i$ contains cross-product terms of powers of π_{1i} and π_{2i}. An unbiased estimator of $\boldsymbol{\Omega}_i$, $\widehat{\boldsymbol{\Omega}}_i$, is obtained by using the relations given in i), ii) and iii) in reverse direction and the equation (2.2). Before developing an estimator, it is convenient to introduce the following notations.

$$M_{gg} = \frac{1}{n}\sum_{i=1}^{n} g_i\, g_i^t = \begin{bmatrix} M_{yy} & M_{yp} \\ M_{py} & M_{pp} \end{bmatrix}, \qquad r_{cc} = \frac{1}{n}\sum_{i=1}^{n} c_i\, c_i^t = \begin{bmatrix} r_{\phi\phi} & r_{\phi\pi} \\ r_{\pi\phi} & r_{\pi\pi} \end{bmatrix},$$

$$\Omega_i = \begin{bmatrix} \sigma_{vv} & \Omega_{vf(i)} \\ \Omega_{fv(i)} & \Omega_{ff(i)} \end{bmatrix}, \qquad \Omega_{rr} = \frac{1}{n}\sum_{i=1}^{n} \Omega_i = \begin{bmatrix} \sigma_{vv} & \Omega_{vf} \\ \Omega_{fv} & \Omega_{ff} \end{bmatrix},$$

$$\widehat{\Omega}_{rr} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Omega}_{i.}$$

3. Estimator and Its Properties

Wolter and Fuller's(1982) estimation method for the quadratic model can be extended to derive an estimator. To do so, define an estimator of $\boldsymbol{\beta}$ to be that which minimizes the function

$$h(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^t M_{gg}\, \boldsymbol{\theta}}{\boldsymbol{\theta}^t\, \widehat{\Omega}_{rr}\, \boldsymbol{\theta}}, \tag{3.1}$$

where $\boldsymbol{\theta} = (1,\ -\ \boldsymbol{\beta}^t)^t$. Fuller showed that the minimization of (3.1) yields MLE for linear functional MEM. However, it is not true in polynomial MEM. The minimization of $h(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ results in

$$\widehat{\boldsymbol{\beta}} = (\ M_{pp} - \hat{\alpha}\ \widehat{\Omega}_{ff})^{-1}(\ M_{py} - \hat{\alpha}\ \widehat{\Omega}_{fv}), \tag{3.2}$$

where $\hat{\alpha}$ is the smallest root of $|\ M_{gg} - \hat{\alpha}\ \widehat{\Omega}_{rr}| = 0$.

Asymptotic properties of $\widehat{\boldsymbol{\beta}}$ are stated in the following Theorem. It is adapted from Wolter and Fuller(1982) for the quadratic model except assumption about the convergence of $n^{-1}\sum_{i=1}^{n}\left(|\pi_{1i}|^j\, |\pi_{2i}|^h\right)$ and dimension of some vectors and matrices. Brief proof of Theorem including the comment on the development of different assumption is provided.

Theorem Let the model (2.1) define a polynomial functional MEM containing 2 predictors with first-order interaction term. Assume the followings.

(a) For all γ in an open sphere containing the true parameter, β,

$$0 < L < (1, -\gamma^t)\, \Omega_{rr}\, (1, -\gamma^t)^t$$

for any $n > k+l+2$, where L is a fixed constant.

(b) $\Upsilon_{\pi\pi}$ is a positive definite matrix for all $n > k+l+2$ and

$$\lim_{n\to\infty} \Upsilon_{\pi\pi} = \Gamma_{\pi\pi},$$

where $\Gamma_{\pi\pi}$ is positive definite.

(c) $n^{-1}\sum_{i=1}^{n} |\pi_{1i}|^j$ converges for $j = 2k+1$ to $4k-2+\delta$,

$n^{-1}\sum_{i=1}^{n} |\pi_{2i}|^h$ converges for $h = 2l+1$ to $4l-2+\delta$, and

$n^{-1}\sum_{i=1}^{n} |\pi_{1i}|^j |\pi_{2i}|^h$ converges for $j = 1, 2, 3, \cdots, 3k-1+\delta$, and

$h = 1, 2, 3, \cdots, 3l-1+\delta$, where $\delta > 0$.

(d) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} abs(\Omega_i) = \Xi^*$, where $abs(\Omega_i)$ denotes a $(k+l+3)$x$(k+l+3)$ matrix whose elements are the absolute values of the elements of Ω_i and where each element of Ξ^* is finite.

Then, $\hat{\beta} \xrightarrow{p} \beta$ and $n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Gamma_{\pi\pi}^{-1} G\, \Gamma_{\pi\pi}^{-1})$, where

$$G = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E(\phi_i \phi_i^t), \quad \phi_i = p_i e_i - (\theta^t \Omega_{rr} \theta)^{-1}(e_i^2 - \theta^t \hat{\Omega}_i \theta)\, \Omega_{fe} - \hat{\Omega}_{fe(i)},$$

$$e_i = g_i^t \theta = r_i^t \theta, \quad \Omega_{fe} = n^{-1}\sum_{i=1}^{n} \Omega_{fe(i)} = n^{-1}\sum_{i=1}^{n} (\Omega_{fv(i)} - \Omega_{ff(i)}\beta) \text{ and}$$

$$\hat{\Omega}_{fe(i)} = \hat{\Omega}_{fv(i)} - \hat{\Omega}_{ff(i)}\beta.$$

Proof Since $\hat{\alpha}$ is the smallest root of $| \boldsymbol{M}_{gg} - \hat{\alpha}\, \widehat{\boldsymbol{\Omega}}_{rr}| = 0$, we have

$\hat{\alpha} = \min_{\boldsymbol{\theta}^*} \dfrac{\boldsymbol{\theta}^{*t}\, \boldsymbol{M}_{gg}\, \boldsymbol{\theta}^*}{\boldsymbol{\theta}^{*t}\, \widehat{\boldsymbol{\Omega}}_{rr}\, \boldsymbol{\theta}^*}$ where $\boldsymbol{\theta}^* = (1, -\boldsymbol{\gamma}^t)^t$. Let $\boldsymbol{\Gamma}_{cc} = \lim_{n\to\infty} \boldsymbol{r}_{cc}$ and

$\boldsymbol{\Xi}_{rr} = \lim_{n\to\infty} \boldsymbol{\Omega}_{rr}$. Since $\boldsymbol{M}_{gg}$ and $\widehat{\boldsymbol{\Omega}}_{rr}$ are consistent estimators of $\boldsymbol{\Gamma}_{cc} + \boldsymbol{\Xi}_{rr}$ and $\boldsymbol{\Xi}_{rr}$ respectively,

$$plim\ \hat{\alpha} = \min_{\boldsymbol{\theta}^*} \frac{\boldsymbol{\theta}^{*t}(\boldsymbol{\Gamma}_{cc} + \boldsymbol{\Xi}_{rr})\, \boldsymbol{\theta}^*}{\boldsymbol{\theta}^{*t}\, \boldsymbol{\Xi}_{rr}\, \boldsymbol{\theta}^*} = 1 + \min_{\boldsymbol{\theta}^*} \frac{\boldsymbol{\tau}^t\, \boldsymbol{\Gamma}_{\pi\pi}\, \boldsymbol{\tau}}{\boldsymbol{\theta}^{*t}\, \boldsymbol{\Xi}_{rr}\, \boldsymbol{\theta}^*} = 1, \tag{3.3}$$

where $\boldsymbol{\tau} = [\ \boldsymbol{\beta}\ \ \boldsymbol{I}]\ \ \boldsymbol{\theta}^*$. Therefore, we have

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}. \tag{3.4}$$

Now, by definition of $\hat{\boldsymbol{\theta}}$ and $\hat{\alpha}$

$$(\boldsymbol{M}_{gg} - \hat{\alpha}\, \widehat{\boldsymbol{\Omega}}_{rr})\, \hat{\boldsymbol{\theta}} = \boldsymbol{0}. \tag{3.5}$$

Premultiplying (3.5) by $\boldsymbol{\theta}^t$ leads to

$$\hat{\alpha} - 1 = \frac{\boldsymbol{\theta}^t\{(\Delta\, \boldsymbol{M}_{gg}) - (\Delta\, \widehat{\boldsymbol{\Omega}}_{rr})\}\, \hat{\boldsymbol{\theta}}}{\boldsymbol{\theta}^t\{(\Delta\, \widehat{\boldsymbol{\Omega}}_{rr}) + \boldsymbol{\Omega}_{rr}\}\, \hat{\boldsymbol{\theta}}} = \frac{\boldsymbol{\theta}^t\{\boldsymbol{M}_{gg} - E(\boldsymbol{M}_{gg}) + \boldsymbol{\Omega}_{rr} - \widehat{\boldsymbol{\Omega}}_{rr}\}\, \hat{\boldsymbol{\theta}}}{\boldsymbol{\theta}^t\, \widehat{\boldsymbol{\Omega}}_{rr}\, \hat{\boldsymbol{\theta}}}, \tag{3.6}$$

where $(\Delta\, \boldsymbol{M}_{gg}) = \boldsymbol{M}_{gg} - E(\boldsymbol{M}_{gg}) = O_p(n^{-1/2})$ and $(\Delta\, \widehat{\boldsymbol{\Omega}}_{rr}) = \widehat{\boldsymbol{\Omega}}_{rr} - \boldsymbol{\Omega}_{rr} = O_p(n^{-1/2})$.
Since
$(\Delta\, \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \begin{bmatrix} 0 \\ -(\Delta\, \hat{\boldsymbol{\beta}}) \end{bmatrix} = o_p(1)$ by (3.4), we have

$$\hat{\alpha} - 1 = O_p(n^{-1/2}) \tag{3.7}$$

from the mid-side of (3.6). Rewrite $\hat{\boldsymbol\beta}-\boldsymbol\beta$ as follows:

$$\hat{\boldsymbol\beta}-\boldsymbol\beta=(\boldsymbol M_{pp}-\hat\alpha\,\hat{\boldsymbol\Omega}_{ff})^{-1}\{\boldsymbol M_{py}-\hat\alpha\,\hat{\boldsymbol\Omega}_{fv}-(\boldsymbol M_{pp}-\hat\alpha\,\hat{\boldsymbol\Omega}_{ff})\boldsymbol\beta\}. \tag{3.8}$$

By multiplying (3.7) to both terms of right-hand side of (3.8), we get $\hat{\boldsymbol\theta}=\boldsymbol\theta+O_p(n^{-1/2})$. Since $\boldsymbol\theta^t\,\hat{\boldsymbol\Omega}_{rr}\,\hat{\boldsymbol\theta}=\boldsymbol\theta^t\,\boldsymbol\Omega_{rr}\,\boldsymbol\theta+O_p(n^{-1/2})$ and $\boldsymbol\theta^t\{\boldsymbol M_{gg}-E(\boldsymbol M_{gg})+\boldsymbol\Omega_{rr}-\hat{\boldsymbol\Omega}_{rr}\}\hat{\boldsymbol\theta}=n^{-1}\sum_{i=1}^n e_i^2-\boldsymbol\theta^t\,\hat{\boldsymbol\Omega}_{rr}\,\boldsymbol\theta+O_p(n^{-1})$, it is easy to derive from the right-side of (3.6) that

$$\hat\alpha-1=\left(n^{-1}\sum_{i=1}^n e_i^2-\boldsymbol\theta^t\,\hat{\boldsymbol\Omega}_{rr}\,\boldsymbol\theta\right)\left(\boldsymbol\theta^t\,\boldsymbol\Omega_{rr}\,\boldsymbol\theta\right)^{-1}+O_p(n^{-1}).$$

Therefore, we have $\sqrt{n}(\hat{\boldsymbol\beta}-\boldsymbol\beta)=\boldsymbol\Gamma_{xx}^{-1}\frac{1}{\sqrt n}\sum_{i=1}^n\boldsymbol\phi_i+O_p(n^{-1/2})$. Limiting distribution of $\sqrt{n}(\hat{\boldsymbol\beta}-\boldsymbol\beta)$ is the same as that of $\boldsymbol\Gamma_{xx}^{-1}\frac{1}{\sqrt n}\sum_{i=1}^n\boldsymbol\phi_i$. To investigate their limiting distribution, let $\boldsymbol\lambda$ be an arbitrary nonzero $(k+l+2)$x1 vector and consider $\frac{1}{\sqrt n}\sum_{i=1}^n\boldsymbol\lambda^t\boldsymbol\phi_i$. The term containing the highest power of $\pi_{1i}^j\pi_{2i}^h$ in E $(\boldsymbol\phi_i\boldsymbol\phi_i^t)$ is $(\boldsymbol\theta^t\,\boldsymbol\Omega_i\,\boldsymbol\theta)\ \boldsymbol\pi_i\,\boldsymbol\pi_i^t$, and it contains terms from π_{1i} to π_{1i}^{4k-2}, π_{2i} to π_{2i}^{4l-2} and from $\pi_{1i}\pi_{2i}$ to $\pi_{1i}^{3k-1}\pi_{2i}^{3l-1}$. Therefore,

$$\lim_{n\to\infty}\frac{\sum_{i=1}^n E\{|\boldsymbol\lambda^t\boldsymbol\phi_i|^{2+\delta}\}}{\left[\sum_{i=1}^n E\{(\boldsymbol\lambda^t\boldsymbol\phi_i)^2\}\right]^{(2+\delta)/2}}=\lim_{n\to\infty}\frac{n^{-1}\sum_{i=1}^n E\{|\boldsymbol\lambda^t\boldsymbol\phi_i|^{2+\delta}\}}{n^{\delta/2}\left\{\boldsymbol\lambda^t n^{-1}\sum_{i=1}^n E(\boldsymbol\phi_i\boldsymbol\phi_i^t)\boldsymbol\lambda\right\}^{(2+\delta)/2}}=0,$$

by assumptions (b) and (c). Therefore, using Liapounov CLT,

$$\frac{\frac{1}{\sqrt n}\sum_{i=1}^n\boldsymbol\lambda^t\boldsymbol\phi_i}{\left\{\boldsymbol\lambda^t E\left(n^{-1}\sum_{i=1}^n\boldsymbol\phi_i\boldsymbol\phi_i^t\right)\boldsymbol\lambda\right\}^{1/2}}\xrightarrow{d}N(0,1).$$

The result of the theorem follows from assumption (b) and multivariate CLT.

4. Simulation Results

In this Section, some simulation results are presented to compare the small sample properties of derived estimator with those of OLS estimator. Wolter and Fuller(1982) introduce a small-order modification in their estimator of the quadratic model since their estimator has heavy tail problems due to the presence of a small number of extreme observations in their preliminary simulation results. Same kind of problem happened in this work and the derived estimator is modified in the same manner as given by Wolter and Fuller. The modified form of the estimator is given by

$$\widehat{\beta}(h) = \left\{ \boldsymbol{M}_{pp} - (\hat{a} - \frac{h}{n})\ \widehat{\boldsymbol{\Omega}}_{ff} \right\}^{-1} \left\{ \boldsymbol{M}_{py} - (\hat{a} - \frac{h}{n})\ \widehat{\boldsymbol{\Omega}}_{fv} \right\},$$

and this one is used in the simulation for $h = 0, 4, 6$.

The polynomial MEM studied in the simulation is given by

$$\psi_i = \beta_0 + \beta_{11}\pi_{1i} + \beta_{12}\pi_{1i}^2 + \beta_{21}\pi_{2i} + \beta_{22}\pi_{2i}^2 + \beta_3\pi_{1i}\pi_{2i},$$

where $\boldsymbol{w}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{ww})$ with completely known $\boldsymbol{\Sigma}_{ww}$. The true $\boldsymbol{\beta}^t$ is set equal to (0.5, 0.0, 1.0, 0.0, -1.0, 1.0). Data sets, and the remaining parameters of the model are selected from one of the followings:

1) Data set and sample sizes:

i) Data set A, n=36: $\boldsymbol{\pi}_1^t = \boldsymbol{\pi}_2^t =$ (-0.5, -0.3, -0.1, 0.1, 0.3, 0.5), and each (π_{1i}, π_{2i}) is obtained as a permutation of element from each $\boldsymbol{\pi}_1^t$ and $\boldsymbol{\pi}_2^t$.

ii) Data set A, n=108: Duplicate Data set given in i) three times.

iii) Data set B, n=36: $\boldsymbol{\pi}_1^t = \boldsymbol{\pi}_2^t =$ (-0.4, -0.3, -0.15, -0.05, 0.4, 0.5), and each (π_{1i}, π_{2i}) is obtained as a permutation of element from each $\boldsymbol{\pi}_1^t$ and $\boldsymbol{\pi}_2^t$.

iv) Data set B, n=108: Duplicate Data set given in iii) three times.

2) Covariance matrices of measurement error vector, $\boldsymbol{\Sigma}_{ww}$:

$$\text{I}\ \begin{bmatrix} 0.06 & 0.00 & 0.00 \\ & 0.06 & 0.00 \\ & & 0.06 \end{bmatrix} \quad \text{II}\ \begin{bmatrix} 0.04 & 0.00 & 0.00 \\ & 0.04 & 0.00 \\ & & 0.04 \end{bmatrix} \quad \text{III}\ \begin{bmatrix} 0.04 & 0.02 & 0.02 \\ & 0.04 & 0.01 \\ & & 0.04 \end{bmatrix} \quad \text{IV}\ \begin{bmatrix} 0.02 & 0.00 & 0.00 \\ & 0.02 & 0.00 \\ & & 0.02 \end{bmatrix} \quad \text{V}\ \begin{bmatrix} 0.005 & 0.000 & 0.000 \\ & 0.005 & 0.000 \\ & & 0.005 \end{bmatrix}$$

Five covariance matrices are selected so that the ratios of the error variance to the mean square for π_j take the similar values as those of Wolter & Fuller(1982). Covariance matrix III is included in order to assess the performance of an estimator when the measurement errors are correlated. Multivariate normal variates $\boldsymbol{w}_i$ with covariance matrix Σ_{ww} were generated by IMSL subroutine RNMVN. Data sets A, B are chosen to see the effects of collinearity. Although collinearity is not severe in both cases, data set B contains stronger collinearity problem than data set A since it is not equally spaced around zero.

For each combination of four data sets and five Σ_{ww}, 200 replication results are summarized in Table 1. It includes the ratios of TSE{ $\hat{\beta}(h)$ } to TSE{ $\hat{\beta}_{OLS}$} where TSE is the sum of six MSE. As was expected, the results of data set A are better than those of data set B except Parameter set V. And as the sample size is increased, these ratios reduced much which implies that the performance of $\hat{\beta}(h)$ relative that of $\hat{\beta}_{OLS}$ becomes better and better for larger samples. For data set A, direct comparison of $\hat{\beta}(4)$ and $\hat{\beta}_{OLS}$ shows that $\hat{\beta}(4)$ performs better when n=108 but $\hat{\beta}_{OLS}$ does when n=36. But for data set B, $\hat{\beta}_{OLS}$ is better in almost cases regardless of sample size. The examination of Table 2 and Table 3, which contain more detailed information on the estimators, reveals that it is due to much less variances of $\hat{\beta}_{OLS}$ than those of $\hat{\beta}(4)$ although biases of $\hat{\beta}_{OLS}$ are relatively large. The situation is somewhat different for $\hat{\beta}(6)$. That is, $\hat{\beta}(6)$ is better than $\hat{\beta}_{OLS}$ except for data set B with n=36 and Parameter Set I, II and III, although variances of $\hat{\beta}_{OLS}$ is much less than those of $\hat{\beta}(6)$ as was in $\hat{\beta}(4)$.

As a conclusion, $\hat{\beta}(h)$, $h=4, 6$, is considered better than $\hat{\beta}_{OLS}$ in almost all cases because of large biases of $\hat{\beta}_{OLS}$. Although TSE{ $\hat{\beta}_{OLS}$} is less in some cases, it is mainly due to small variances of $\hat{\beta}_{OLS}$. Table 2 and Table 3 show those features clearly. Finally, the choice between $\hat{\beta}(4)$ and $\hat{\beta}(6)$ goes to $\hat{\beta}(6)$, especially for the data set with a mild or severe collinearity problem and/or with small samples.

References

[1] Chan, L. K. and Mak, T. K. (1985). On the Polynomial Functional Relationship, *Journal of the Royal Statistical Society, Series B,* Vol. 47, 510-518.

[2] IMSL User's Manual (1989). IMSL Inc., Houston, TX.

[3] Moon, M. S. and Gunst, R. F.(1994). Estimation of the Polynomial Errors-in-variables Model with Decreasing Error Variances, *Journal of the Korean Statistical Society,* Vol. 23, No. 1, 115-134.

[4] Moon, M. S. and Gunst, R. F.(1995). Polynomial Measurement Error Modeling, *Computational Statistics and Data Analysis, North Holland,* Vol. 19, 1-21.

[5] Stefanski, L. (1985). The Effects of Measurement Error on Parameter Estimation, *Biometrika,* Vol. 72, 583-592.

[6] Stefanski, L. (1989). Unbiased Estimation of a Nonlinear Function of a Normal Mean with Application to Measurement Error Models, *Communications in Statistics-Theory and Methods,* Vol. 18, 4335-4358.

[7] Wolter, K. M. and Fuller, W. A. (1982). Estimation of the Quadratic Errors-in-variables Model, *Biometrika,* Vol. 69, 175-182.

Table 1. Ratios of TSE#

Data		Parameter Set									
		I		II		III		IV		V	
Set	$\hat{\beta}(h)$ n:	36	108	36	108	36	108	36	108	36	108
A	$\hat{\beta}(0)$	*※	*	*	173	138	298	230	57	98	39
	$\hat{\beta}(4)$	113	99	116	85	112	86	106	49	89	38
	$\hat{\beta}(6)$	95	76	93	70	91	69	91	45	86	38
B	$\hat{\beta}(0)$	*	*	*	*	*	*	*	194	100	35
	$\hat{\beta}(4)$	124	123	128	109	120	107	121	62	84	33
	$\hat{\beta}(6)$	107	96	107	87	103	82	98	53	79	32

$100 \cdot (TSE\{\hat{\beta}(h)\}/TSE\{\hat{\beta}_{OLS}\})$, $h=0,4,6$.

※ * inicates value > 1,000.

Table 2. Estimator Bias and Variance, Parameter Set III, n=36

Data Set	Estimator	β_0	β_1	β_2	β_3	β_4	β_5	Absolute Sum
				(a) Bias				
A	$\hat{\beta}_{OLS}$	.013	.105	-.684	.091	.579	-.353	1.825
	$\beta(0)$	-.129	.199	-.577	-.384	2.623	.935	4.847
	$\beta(4)$	.005	.021	-.286	.007	.280	-.028	.627
	$\beta(6)$	.009	.034	-.383	.020	.343	-.090	.879
				(b) Variance				
	$\hat{\beta}_{OLS}$	.003	.011	.061	.009	.065	.081	.230
	$\beta(0)$	3.294	2.157	172.060	20.014	1392.560	27.587	1617.672
	$\beta(4)$	.009	.027	.333	.025	.405	.352	1.151
	$\beta(6)$	.007	.022	.227	.020	.269	.254	.799
				(a) Bias				
B	$\hat{\beta}_{OLS}$	.042	.197	-.921	.036	.587	-.327	2.110
	$\beta(0)$	.098	-.067	.159	.205	-1.060	-.070	1.659
	$\beta(4)$	.041	.104	-.687	-.016	.286	-.058	1.192
	$\beta(6)$	.041	.125	-.757	-.010	.361	-.107	1.401
				(b) Variance				
	$\hat{\beta}_{OLS}$	.004	.011	.082	.011	.073	.075	.256
	$\beta(0)$	.684	1.978	36.911	4.292	79.933	13.186	136.984
	$\beta(4)$	.011	.033	.512	.037	.480	.269	1.342
	$\beta(6)$	.008	.025	.337	.029	.319	.198	.916

Table 3. Estimator Bias and Variance, Parameter Set III, n=108

Data Set	Estimator	β_0	β_1	β_2	β_3	β_4	β_5	Absolute Sum
				(a) Bias				
A	$\hat{\beta}_{OLS}$	.014	.116	-.735	.118	.605	-.413	2.001
	$\beta(0)$	.946	.522	-10.956	-1.991	3.898	-2.283	20.596
	$\beta(4)$	.008	.016	-.025	.014	-.033	-.009	.105
	$\beta(6)$	.010	.018	-.090	.017	.014	-.027	.176
				(b) Variance				
	$\hat{\beta}_{OLS}$	.001	.003	.015	.003	.017	.022	.061
	$\beta(0)$	180.74	51.52	24911.63	802.15	3261.84	1063.08	30270.97
	$\beta(4)$	.012	.013	.323	.013	.422	.213	.996
	$\beta(6)$	.009	.012	.252	.012	.325	.180	.790
				(a) Bias				
B	$\hat{\beta}_{OLS}$	.031	.216	-.906	.040	.645	-.402	2.240
	$\beta(0)$	.123	-.007	.165	.165	-1.193	-.075	1.728
	$\beta(4)$	-.006	.037	-.133	-.013	.152	.022	.363
	$\beta(6)$	.002	.055	-.245	-.016	.196	-.007	.521
				(b) Variance				
	$\hat{\beta}_{OLS}$	.001	.004	.018	.002	.019	.024	.068
	$\beta(0)$	6.147	1.712	46.389	7.817	302.675	6.352	371.092
	$\beta(4)$	.018	.021	.545	.025	.761	.206	1.576
	$\beta(6)$	.013	.017	.391	.018	.537	.164	1.140