Cheng, Schneeweiss, Thamerus:

A Small Sample Estimator for a Polynomial Regression with Errors in the Variables

Projektpartner

A Small Sample Estimator for a Polynomial Regression with Errors in the Variables

Chi-Lun Cheng[1], Hans Schneeweiss[2] and Markus Thamerus[2]

[1] Institute of Statistical Science, Academia Sinica, Taipei, Republic of China
[2] Institute for Statistics, University of Munich, Munich, Germany

Abstract. An adjusted least squares estimator, introduced by Cheng and Schneeweiss (1998) for consistently estimating a polynomial regression of any degree with errors in the variables, is modified such that it shows good results in small samples without losing its asymptotic properties for large samples. Simulation studies corroborate the theoretical findings. The new method is applied to analyse a geophysical law relating the depth of earthquakes to their distance from a trench where one of the earth's plates is submerged beneath another one.

Keywords: Polynomial regression, measurement error, adjusted least squares, modified adjusted least squares, small sample properties

1   Introduction

Cheng and Schneeweiss (1998), hereafter CS, building on the work of Chan and Mak (1985) and Stefanski (1989), developed a consistent estimator of the parameters of a polynomial regression with errors in the variables. In this model the latent true regressor variables ξ_i are taken to be non-stochastic (the so-called functional variant of the errors-in-variables model), so that no distributional assumption on the $\xi's$ is needed. The errors need not necessarily be normally distributed, although the estimation procedure simplifies considerably in the normal case. The errors of regressor and dependent variables are allowed to be correlated. Knowledge of certain higher moments of the error variables is required, but the variance (and higher moments) of the errors of the dependent variable need not be known. This is justified by the observation that for pure measurement errors knowledge or at least estimates of their variance and higher moments may well be available, e.g. through replicated measurements or through validation studies, whereas errors in the dependent variable very often comprise the errors in the equation and for these the variance is typically not known. For a further discussion of models with and without errors in the equation see CS.

The (error) adjusted least squares (ALS) estimator developed in CS can be viewed as resulting from the principle of corrected unbiased estimating equations; see Nakamura (1990) and Bounaccorsi (1996). This estimator is asymptotically normal under general conditions. However, in small samples devia-

1

tions from normality can be crucial. Indeed, simulation studies have shown that the ALS estimator of a polynomial regression can sometimes produce extremely large estimating errors. The purpose of this paper is to modify the ALS estimator such that this deficiency is remedied.

Fuller (1987 p. 212f.) constructed an error adjusted estimator for a quadratic errors-in-variables regression using another principle. His approach can be generalized to the polynomial case and then turns out to give identical results as in CS. However, Fuller's approach can be used as a starting point to modify the estimator in such a way that the small sample deficiencies mentioned above are considerably alleviated, whereas the large sample properties are left unaltered. The device to modify the estimator is taken again from Fuller (1987, p. 164ff.), but must be adapted to our case because, as will be seen, the estimated error covariance matrices involved in our approach need not necessarily be positive semi definite as required by Fuller.

In the next section the polynomial errors-in-variables model is introduced and the unmodified ALS estimator from CS is presented. It is shown that Fuller's approach, if properly generalized, leads to the same result. In Section 3 the ALS estimator is modified so that it produces a positive (semi) definite estimate of the error covariance matrix. Further modifications which lead to estimators with finite moments are considered in Section 4. Section 5 gives some simulation results showing the effect of the modifications. An empirical example is given in Section 6 and Section 7 contains some concluding remarks.

2   The unmodified ALS estimator

Consider a polynomial functional relationship with errors in the equation:

$$
\begin{aligned}
y_i &= \eta_i + \varepsilon_i = \beta_0 + \beta_1 \xi_i + \beta_2 \xi_i^2 + \ldots + \beta_k \xi_i^k + \varepsilon_i, \\
x_i &= \xi_i + \delta_i,
\end{aligned}
$$

$i = 1, \ldots, n$, where $(\delta_i, \varepsilon_i)$ are i.i.d. random errors with expectation 0 and covariance matrix

$$
\Omega = \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix}.
$$

The $\xi_i, i = 1, \ldots, n$, are unobservable (latent) nonstochastic variables. The regressor error variance σ_δ^2 and the covariance $\sigma_{\delta\varepsilon}$ of δ and ε are assumed to be known. We assume $\sigma_\delta^2 > 0$. An often encountered special case is $\sigma_{\delta\varepsilon} = 0$. The variance σ_ε^2, which contains the error-in-the-equation variance component, is unknown. If the error variables are jointly normally distributed, i.e.,

$$(N) \qquad\qquad (\delta_i, \varepsilon_i) \sim N(0, \Omega),$$

knowledge of σ_δ^2 and $\sigma_{\delta\varepsilon}$ is sufficient. Otherwise higher moments $\mathsf{E}\delta^r$, $r \leq 2k$, and $\mathsf{E}(\delta^r\varepsilon)$, $r \leq k$, have to be known.

2

It is well-known that replacing the latent variable ξ by its observable counterpart x in the polynomial relationship and estimating the parameters β_j in the resulting polynomial regression by ordinary least squares (OLS) yields inconsistent estimates; Grilliches and Ringstad (1970).

A consistent estimator, of course, is given by the OLS normal equations constructed from the *true* variables, supposing for a moment that these were known. Let $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ and $\zeta_i = (1, \xi_i, \ldots, \xi_i^k)'$, then

$$\sum \zeta_i \zeta_i' \hat{\beta}_T = \sum \zeta_i y_i \tag{1}$$

are the normal equations for a theoretical OLS estimator $\hat{\beta}_T$ of β (all sums, unless otherwise stated, are taken over $i = 1, \ldots, n$) and σ_ε^2 is estimated by

$$\hat{\sigma}_{\varepsilon T}^2 = \overline{y^2} - \overline{y\zeta'}\hat{\beta}_T, \tag{2}$$

where the bar, here and in the sequel, denotes averages over $i = 1, \ldots, n$, e.g., $\overline{y^2} = \frac{1}{n}\sum y_i^2$.

Of course, $\hat{\beta}_T$ and $\hat{\sigma}_{\varepsilon T}^2$ cannot be computed if the ξ_i have measurement errors. CS now replace the matrix $\zeta_i \zeta_i'$ by a $(k+1) \times (k+1)$ matrix H_i the elements of which are certain polynomials in the observable variable x_i such that $\mathsf{E}H_i = \zeta_i \zeta_i'$. Similarly $\zeta_i y_i$ is replaced by a vector h_i, which can also be computed from the observable variables, such that $\mathsf{E}h_i = \mathsf{E}(\zeta_i y_i) = \zeta_i \eta_i$. It then follows that the solution $\hat{\beta}_A$ of

$$\sum H_i \hat{\beta}_A = \sum h_i \tag{3}$$

is a consistent estimator of β assuming that $\lim \overline{\zeta\zeta'}$ exists and is nonsingular. Its asymptotic covariance matrix can be estimated by

$$\hat{\Sigma}_{\hat{\beta}_A} = \frac{1}{n}\bar{H}^{-1}\bar{U}\bar{H}^{-1}, \tag{4}$$

where $\bar{H} = \frac{1}{n}\sum H_i$ and $\bar{U} = \frac{1}{n}\sum u_i u_i'$ with $u_i = H_i \hat{\beta}_A - h_i$. Once β has been consistently estimated, a consistent estimate for σ_ε^2 can also be found:

$$\hat{\sigma}_{\varepsilon A}^2 = \overline{y^2} - \bar{h}'\hat{\beta}_A, \tag{5}$$

$\hat{\beta}_A$ and $\hat{\sigma}_{\varepsilon A}^2$ are the ALS estimates of β and σ_ε^2, respectively.

The matrices $H_i, i = 1, \ldots, n$, are constructed as follows; for details see CS. First compute the polynomials

$$t_{ri} = \sum_{j=0}^{r} a_{ri} x_i^j, \quad r = 0, \ldots, k,$$

by solving the linear system

$$x_i^m = \sum_{r=0}^{m} \binom{m}{r} t_{ri}\mathsf{E}\delta^{m-r}, \quad m = 0, \ldots, k.$$

It can be shown that $\mathsf{E}t_{ri} = \xi_i^r$. The matrix H_i is then given by its elements $H_{rsi} = t_i^{r+s}$; $r, s = 0, \ldots, k$.

The elements h_{ri} of the vector h_i, $r = 0, \ldots, k$, are constructed as

$$
\begin{aligned}
h_{ri} &= t_{ri}y_i - \sum_{j=0}^{r} b_{rj}t_{ji}, \\
b_{rj} &= \sum_{s=j}^{r} a_{rs} \binom{s}{j} \mathsf{E}(\delta^{s-j}\varepsilon).
\end{aligned}
$$

These computations are greatly simplified if (N) can be assumed; see CS. In fact, in this case, the t_{ri} can be computed from the recursive equation $t_{r+1,i} = x_i t_{ri} - \sigma_\delta^2 r t_{r-1,i}$, $r = 0, 1, 2, \ldots$, with $t_{0i} = t_{-1i} = 1$, and the elements of h_i are given by $h_{ri} = t_{ri}y - \sigma_{\delta\varepsilon} r t_{r-1,i}$, $r = 0, 1, \ldots, k$.

Fuller (1987, p. 212f.) solved the problem of constructing a consistent estimator of β by a different approach, which he described, however, only for the quadratic case ($k = 2$) under (N) and with $\sigma_{\delta\varepsilon} = 0$. It can be generalized to the polynomial case without (N) and with $\sigma_{\delta\varepsilon}$ arbitrary. Fuller suggested to view the powers of ξ as $k + 1$ different latent regressor variables, for which, as their observable counterparts, the unbiased estimates t_r are available, so that a linear functional relationship results:

$$
\begin{aligned}
y_i &= \beta_0 \xi_i^0 + \beta_1 \xi_i^1 + \ldots + \beta_k \xi_i^k + \varepsilon_i, \\
t_{ri} &= \xi_i^r + e_{ri}, \quad r = 0, \ldots, k, \quad i = 1, \ldots, n,
\end{aligned}
$$

where the e_{ri} are the new measurement errors, with $\mathsf{E}e_{ri} = 0$. Let $t_i = (t_{0i}, \ldots, t_{ki})'$, $e_i = (e_{0i}, \ldots, e_{ki})'$, then the model can be written as

$$
\begin{aligned}
y_i &= \zeta_i'\beta + \varepsilon_i \\
t_i &= \zeta_i + e_i
\end{aligned} \qquad i = 1, \ldots, n.
$$

For this linear functional relationship a consistent estimator of β can now be constructed if unbiased estimates $\hat{V}_i$ and $\hat{v}_i$ of, respectively, the covariance matrix $V_i = \mathsf{E}(e_i e_i')$ and the covariance vector $v_i = \mathsf{E}(e_i \varepsilon_i)$ are available. Fuller's error adjusted least squares normal equations are given by

$$
\Sigma(t_i t_i' - \hat{V}_i)\hat{\beta}_F = \Sigma(t_i y_i - \hat{v}_i), \tag{6}
$$

and σ_ε^2 is estimated by

$$
\hat{\sigma}_{\varepsilon F}^2 = \overline{y^2} - (\overline{ty} - \bar{v})'\hat{\beta}_F, \tag{7}
$$

where $\bar{v}$ is short term for $\bar{\hat{v}}$.

The covariance matrix V_i is given by

$$
\begin{aligned}
V_i &= \mathsf{E}(t_i - \zeta_i)(t_i - \zeta_i)' \\
&= \mathsf{E}(t_i t_i') - \zeta_i \zeta_i'.
\end{aligned}
$$

4

As $\mathsf{E}H_i = \zeta_i \zeta_i'$, an unbiased estimate of V_i is given by

$$\hat{V}_i = t_i t_i' - H_i. \tag{8}$$

Similarly,

$$v_i = \mathsf{E}\{(t_i - \zeta_i)\varepsilon_i\} = \mathsf{E}(t_i \varepsilon_i) = \mathsf{E}\{t_i(y_i - \eta_i)\} = \mathsf{E}(t_i y_i) - \zeta_i \eta_i = \mathsf{E}(t_i y_i - h_i).$$

Thus an unbiased estimate of v_i is given by

$$\hat{v}_i = t_i y_i - h_i. \tag{9}$$

Substituting (8) and (9) into the error adjusted normal equations (6) results in

$$\sum H_i \hat{\beta}_F = \sum h_i.$$

These are the same estimating equations for β as (3). Similarly (7) is seen to be equivalent to (5). So Fuller's approach and that of CS give rise to the same estimators for β and σ_ε^2, i.e., $\hat{\beta}_F = \hat{\beta}_A$ and $\hat{\sigma}_{\varepsilon F}^2 = \hat{\sigma}_{\varepsilon A}^2$. Nevertheless, Fuller's approach will be useful as a starting point to improve the ALS estimator.

3 Small sample modification

For large samples $(n \to \infty)$ $\bar{H}$ tends to $\overline{\zeta \zeta'}$ and therefore will be positive definite (p.d.) with high probability. For small samples, however, $\bar{H}$ may well be indefinite, despite the fact that it is an unbiased estimate of $\overline{\zeta \zeta'}$, which is p.d.. As a consequence, the estimate $\hat{\beta}_A$ can sometimes be far off the true parameter vector β. Even if $\bar{H}$ is p.d., the estimator of σ_ε^2 given by (5) is deficient in that the estimated covariance matrix

$$\hat{\Omega}_A = \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \hat{\sigma}_{\varepsilon A}^2 \end{pmatrix} \tag{10}$$

may not be p.d., and, what is more, $\hat{\sigma}_{\varepsilon A}^2$ may even turn out to be negative. Thus one would like to modify the estimation procedure so that it yields estimates of $\overline{\zeta \zeta'}$ and Ω which are p.d. or at least positive semi definite (p.s.d.).

Fuller (1987, pp. 164ff., 171f., 193) introduced such a modification for the general *linear* model with errors in the variables, which can be applied to our case, although, as will be seen, some adaptations are necessary. (Fuller (1987, p. 213) also tried to modify the *quadratic* errors-in-variables model, but only partially succeeded in doing so.)

First note that the estimating equations (3) and (5) – or equivalently (6) and (7) – for β and σ_ε^2 can be put together into one system of equations. Define $\hat{\theta} = (1, -\hat{\beta}')'$, then (6) and (7) are equivalent to

$$\left[\begin{pmatrix} \overline{y^2} & \overline{yt'} \\ \overline{ty} & \overline{tt'} \end{pmatrix} - \begin{pmatrix} \hat{\sigma}_\varepsilon^2 & \bar{v}' \\ \bar{v} & \bar{V} \end{pmatrix} \right] \hat{\theta} = 0, \tag{11}$$

where $\bar{v}$ and $\bar{V}$ are short terms for $\bar{\bar{v}}$ and $\bar{\bar{V}}$, respectively. In the sequel we assume the first matrix in the square brackets of (11) to be p.d., which it is with probability 1 if the ξ_i properly vary.

Now let λ be the smallest *positive* root of

$$\left| \left(\begin{array}{cc} \overline{y^2} & \overline{yt'} \\ \overline{ty} & \overline{tt'} \end{array} \right) - \lambda \left(\begin{array}{cc} \sigma^2_{\delta\varepsilon}/\sigma^2_{\delta} & \bar{v}' \\ \bar{v} & \bar{V} \end{array} \right) \right| = 0. \tag{12}$$

It can be shown that there is always at least one positive root of (12), and therefore λ exists; see proof of Prop. 1.

The following estimation procedure is proposed:

a) If $\lambda > 1$, then solve (10) for $\hat{\beta}$ and $\hat{\sigma}^2_\varepsilon$. The solution will be the ALS estimates given by (3) and (5), respectively.

b) If $\lambda \leq 1$, then, instead of (11), solve

$$\left[\left(\begin{array}{cc} \overline{y^2} & \overline{yt'} \\ \overline{ty} & \overline{tt'} \end{array} \right) - \lambda \left(\begin{array}{cc} \hat{\sigma}^2_\varepsilon & \bar{v}' \\ \bar{v} & \bar{V} \end{array} \right) \right] \hat{\theta} = 0. \tag{13}$$

Proposition 1

The estimates $\hat{\beta}$ and $\hat{\sigma}^2_\varepsilon$ resulting from this procedure have the following properties:

In case (a), $\hat{\Omega}$ and $\overline{tt'} - \bar{V}$ are both p.d..

In case (b), $\hat{\Omega}$ is p.s.d. and singular with probability 1, and $\overline{tt'} - \lambda\bar{V}$ is p.d. with probability 1.

Thus in both cases, $\hat{\beta}$ is uniquely determined (with probability 1) from (11) or (13), respectively.

Proof

Let
$$A = \left(\begin{array}{cc} \overline{y^2} & \overline{yt'} \\ \overline{ty} & \overline{tt'} \end{array} \right), \quad B = \left(\begin{array}{cc} \sigma^2_{\delta\varepsilon}/\sigma^2_{\delta} & \bar{v}' \\ \bar{v} & \bar{V} \end{array} \right), \quad \hat{B} = \left(\begin{array}{cc} \hat{\sigma}^2_\varepsilon & \bar{v}' \\ \bar{v} & \bar{V} \end{array} \right).$$

As already said, we suppose A to be p.d.. Let A be decomposed as

$$A = CC'$$

with a nonsingular matrix C (e.g., by a Cholesky decomposition) and define

$$\tilde{B} = C^{-1}BC'^{-1}.$$

6

Let d be the largest eigenvalue of $\tilde{B}$. As the second diagonal element of $\bar{V}$, i.e., $\overline{t_1^2} - \bar{t}_2 = \sigma_\delta^2$, is positive, $\tilde{B}$ has at least one positive eigenvalue, and therefore $d > 0$. It follows that $\lambda = 1/d$. Case (a) corresponds to $d < 1$ and case (b) to $d \geq 1$.

Let D be the diagonal matrix of the eigenvalues of $\tilde{B}$ and E a matrix the columns of which are the corresponding normalized eigenvectors. Then

$$\tilde{B} = EDE', \ EE' = I.$$

It follows that $B = CEDE'C'$, and, with $T = CE$, we can write

$$A = TT', \ B = TDT'.$$

Hence for any scalar a

$$A - aB = T(I - aD)T', \tag{14}$$

with a nonsingular matrix T.

In case (a), $d < 1$, i.e., $D < I$, and because of (14)

$$A - B > 0.$$

Obviously then also $\overline{tt'} - \bar{V} > 0$. In order to prove $\hat{\Omega} > 0$ note that, because of (11) and the positive definiteness of $A - B$,

$$\hat{\theta}'(A - \hat{B})\hat{\theta} = 0$$
$$\hat{\theta}'(A - B)\hat{\theta} > 0$$

and therefore

$$\hat{\theta}'(\hat{B} - B)\hat{\theta} = \hat{\sigma}_\varepsilon^2 - \sigma_{\delta\varepsilon}^2/\sigma_\delta^2 > 0.$$

This is equivalent to $\hat{\Omega} > 0$.

In case (b), $d \geq 1$, and therefore $I - \frac{1}{d}D$ is singular and ≥ 0. It follows from (14) that $A - \frac{1}{d}B = A - \lambda B$ is singular and p.s.d.. With probability 1, rank $(A - \lambda B) = k + 1$ and the first row of $A - \lambda B$ is a linear combination of the remaining $k + 1$ last rows, i.e., the rows of the matrix

$$(\overline{ty}, \overline{tt'}) - \lambda(\bar{v}, \bar{V}) = C,$$

say. Since by (13) $C\hat{\theta} = 0$, so also $(A - \lambda B)\hat{\theta} = 0$. But by (13) also $(A - \lambda\hat{B})\hat{\theta} = 0$. It follows that $(\hat{B} - B)\hat{\theta} = 0$, i.e.,

$$\hat{\sigma}_\varepsilon^2 = \sigma_{\delta\varepsilon}^2/\sigma_\delta^2,$$

which is equivalent to $\hat{\Omega}$ singular and p.s.d.. $\diamond$

Note that the first paragraph of the proof gives a method of how to compute λ by standard programs; see also Amemiya (1985).

Note also that if $\sigma_{\delta\varepsilon} = 0$, $\hat{\sigma}_\varepsilon^2 = 0$ in case (b).

Finally note that $\hat{V}_i$, although it is an unbiased estimate of the covariance matrix of the errors e_i of the derived linear errors-in-variables model, need not be p.s.d.. The same is true for $\bar{V}$. But since we are not interested in an estimate of V_i per se, this should not bother us too much. But as a consequence of the possible indefiniteness of $\bar{V}$ we had to adapt Fuller's modification to our case by taking for λ not just the smallest root of (12) but rather the smallest positive root.

<u>Proposition 2:</u>

> Let Ω and $\lim \overline{\zeta\zeta'}$ be p.d.. The modified estimator $\hat{\beta}$ is asymptotically equivalent to the unmodified adjusted estimator $\hat{\beta}_A$, i.e., $\mathrm{plim}_{n\to\infty} \sqrt{n}(\hat{\beta} - \hat{\beta}_A) = 0$, and therefore both estimators have the same asymptotic properties.

<u>Proof:</u>

From the construction of $\hat{\beta}$ we know that $\hat{\beta} = \hat{\beta}_A$ if case (a) occurs or $\lambda = 1$. The case $\lambda = 1$ can be neglected, as it occurs only with probability 0. Case (a) is equivalent to $A > B$; see the proof of Prop. 1. Therefore

$$P\{\sqrt{n}(\hat{\beta} - \hat{\beta}_A) \neq 0\} = 1 - P(A > B).$$

But $A > B$ is equivalent to $\hat{\Omega}_A > 0$ and $\overline{tt'} - \bar{V} > 0$; see the proof of Prop. 1. As $\hat{\Omega}_A$ converges to $\Omega > 0$ in probability and as $\overline{tt'} - \bar{V} = \bar{H}$ converges to $\lim \overline{\zeta\zeta'} > 0$ in probability, therefore $P(A > B)$ converges to 1, and consequently $\sqrt{n}(\hat{\beta} - \hat{\beta}_A)$ converges to 0 in probability. $\diamond$

4 Further modifications

Although the modification of the ALS estimator introduced so far guarantees that the linear system of estimating equations for $\hat{\beta}$ resulting from (11) or (13), respectively, has a p.d. system matrix, viz., $\overline{tt'} - \bar{V}$ or $\overline{tt'} - \lambda\bar{V}$, respectively, this matrix may still come arbitrarily close to the case of singularity and may thus produce estimates with large deviations from the true parameter values. This difficulty can be partly overcome by a further modification of the estimator.

The idea for this modification again stems from Fuller (1987). But due to the fact that B need not be p.d., his approach cannot be adopted unchanged. We first modify cases (a) and (b) to (a') $\lambda > 1 + 1/n$ and (b') $\lambda \leq 1 + 1/n$, respectively. The modified estimators $\hat{\beta}_M$ and $\hat{\sigma}_{\varepsilon M}^2$ are then defined as the solution of the equations system

$$(A - a\hat{B})\hat{\theta}_M = 0, \tag{15}$$

8

where

$$a = \begin{cases} (n - \alpha)/n & \text{in case } (a') \\ \lambda(n - \alpha)/(n + 1) & \text{in case } (b') \end{cases}$$

with some $\alpha < n$ to be chosen so that an estimator with good small sample properties results. Following Fuller (1987, p. 172) – but note that he determined α in a somewhat different context – it is suggested to choose $\alpha = k + 1$.

It should be noted that Fuller used a different modification in case (b'). His "additive" modification cannot be applied to our model. We use instead a "multiplicative" modification.

Proposition 3:
The system matrix of (15) for $\hat{\beta}_M$ satisfies the inequality

$$\overline{tt'} - a\bar{V} \geq \frac{\alpha + 1}{n + 1} \, \overline{tt'} > 0, \tag{16}$$

and $\hat{\sigma}_{\varepsilon M}^2$ satisfies

$$\hat{\sigma}_{\varepsilon M}^2 - \sigma_{\delta\varepsilon}^2/\sigma_\delta^2 \geq \frac{\alpha + 1}{(n + 1)a} \, \hat{\theta}_M' A \hat{\theta}_M > 0. \tag{17}$$

Proof:

In case (a'), $d < n/(n + 1)$ and therefore $D < n/(n + 1)I$; see the proof of Prop. 1. It follows that

$$I - aD = I - \frac{n - \alpha}{n} D > \frac{\alpha + 1}{n + 1} I.$$

In case (b'), $d \geq n/(n + 1) > 0$ and $d^{-1}D \leq I$. It follows that

$$I - aD = I - \frac{n - \alpha}{n + 1} d^{-1}D \geq \frac{\alpha + 1}{n + 1} I$$

because $\lambda = d^{-1}$. So in both cases, because of (14),

$$A - aB \geq \frac{\alpha + 1}{n + 1} A > 0. \tag{18}$$

Deleting the first row and column of these matrices results in (16).

To prove (17) note that (18) implies

$$\hat{\theta}_M'(A - aB)\hat{\theta}_M \geq \frac{\alpha + 1}{n + 1}\hat{\theta}_M' A \hat{\theta}_M.$$

On the other hand, (15) implies

$$\hat{\theta}_M'(A - a\hat{B})\hat{\theta}_M = 0.$$

9

Hence
$$a\hat{\theta}'_M(\hat{B} - B)\hat{\theta}_M = a(\hat{\sigma}^2_{\varepsilon M} - \sigma^2_{\delta\varepsilon}/\sigma^2_\delta) \geq \frac{\alpha + 1}{n + 1}\hat{\theta}'_M A\hat{\theta}_M,$$

which proves (17). $\diamond$

By similar arguments as in the proof of Prop. 2 it can also be shown that $\hat{\beta}_M$ has the same asymptotic properties as $\hat{\beta}$ and $\hat{\beta}_A$.

Up to now we assumed that the error of measurement in the dependent variable η was not known. This measurement error is part of ε. In general ε can be decomposed into the error of the equation, φ say, and the measurement error of η, ε_* say, so that
$$\varepsilon = \varepsilon_* + \varphi.$$

We assume ε_* and φ to be independent, so that

$$\sigma^2_\varepsilon = \sigma^2_{\varepsilon_*} + \sigma^2_\varphi. \tag{19}$$

The measurement error covariance matrix is therefore

$$\Omega_* = \begin{pmatrix} \sigma^2_\delta & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \sigma^2_{\varepsilon_*} \end{pmatrix}.$$

Note that $\sigma_{\delta\varepsilon_*} = \sigma_{\delta\varepsilon}$. Contrary to what was assumed before, we now suppose, for the remaining part of this section, that besides σ^2_δ and $\sigma_{\delta\varepsilon}$ also $\sigma^2_{\varepsilon_*}$ and therefore the whole covariance matrix Ω_* is known. With this knowledge, once σ^2_ε has been estimated by one of the methods dealt before, the variance of the error in the equation, σ^2_φ, can also be estimated, viz., due to (19), by

$$\hat{\sigma}^2_\varphi = \hat{\sigma}^2_\varepsilon - \sigma^2_{\varepsilon_*}.$$

However, this estimate may be negative. In order to avoid this possibility the estimators of β and σ^2_ε must be modified such that $\hat{\sigma}^2_{\varepsilon M} > \sigma^2_{\varepsilon_*}$.

This can be easily done by using the same estimator as defined in (15) but with λ replaced by λ_* defined as the smallest positive root of

$$\mid A - \lambda_* B_* \mid = 0,$$

where
$$B_* = \begin{pmatrix} \sigma^2_{\varepsilon_*} & \bar{v}' \\ \bar{v} & \bar{V} \end{pmatrix}$$

and where cases (a') and (b') are also redefined with λ_* in place of λ.

It can be shown, in the same way as before, that for the estimators resulting from this method (16) is satisfied as before and (17) is strengthened to

$$\hat{\sigma}^2_{\varepsilon M} \geq \sigma^2_{\varepsilon_*} + \frac{\alpha + 1}{(n + 1)a}\hat{\theta}'_M A\hat{\theta}_M > \sigma^2_{\varepsilon_*}.$$

10

5 Simulation

In order to study the small sample properties of the estimators and in particular the effect of the various modifications proposed we performed a small simulation study. We studied five models. The first two are almost identical to those of CS, except that σ_δ^2 has been given a larger value and that now the new estimators are included, the other three have been investigated by Moon and Gunst (1995), but without considering the modifications in this article. It turns out that these modifications are crucial for the precision of the estimates. The models are defined by the following parameter values given in table 1. In

Model	I	II	III	IV	V
k	2	3	2	3	3
β_0	0	0	0.25	0	0
β_1	1	1	1	1	1
β_2	-0.5	-0.5	1	0	0
β_3	–	0.5	–	-1	-1
range	[-1,1]	[-1,1]	[-0.5,0.5]	[-5,5]	[-5,5]
n	201	201	33	100	100
σ_δ^2	0.1	0.1	0.0324	1.5	0.1
τ	0.3	0.3	0.39	0.18	0.01

Table 1: *Parameter values of simulated models I–V (τ= noise–to–signal ratio).*

all models ε and δ were specified as normally distributed variables with $\sigma_\varepsilon^2 = \sigma_\delta^2$ and $\sigma_{\delta\varepsilon} = 0$. The ξ_i were taken as fixed equidistant values within the interval indicated by 'range', with the exception of model III where the ξ_i were placed at the points $-0.5 + i/10, i = 0, \ldots, 10$, each point taken 3 times. The parameter τ is the noise–to–signal ratio and is computed as $\tau = 12\sigma_\delta^2/r^2$, where r is the range of the ξ_i. For each model 1000 samples were simulated and used to estimate the β_i by the various methods: the naive method M1, which consists in replacing ξ_i by x_i and applying least squares, the ALS method M2 of Section 2, the modified ALS method M3 of Section 3, and the further modification M4 of Section 4. In each case the 1000 estimates were averaged and their standard deviations were computed. In each simulation run we also estimated the asymptotic covariance matrix of $\hat{\beta}$ by (4), where however $\hat{\beta}_A$ was replaced by the estimate of β that resulted from the particuliar estimation method used in that run. From the covariance matrix an estimate of the asymptotic standard deviation of each $\hat{\beta}_i$ was derived. The standard deviations were averaged over the 1000 simulation runs. The averages can be compared with the standard deviations directly derived from the 1000 estimates of each β_i.

The results are presented in the following tables 2 and 3. In almost all cases the naive estimator (M1) is strongly biased. Even in model V with the rather small noise–to–signal ratio τ=0.01 the coefficient β_1 is extremely biased. Typically

<div align="center">Model I ($\tau = 0.3$)</div>

β_i	M1	M2	M3	M4	Estimation method
0.00	-0.09	0.03	0.00	-0.00	Averages of
1.00	0.78	1.02	1.00	1.00	estimates
-0.50	-0.19	-0.59	-0.52	-0.50	
	0.03	0.10	0.06	0.06	Standard deviations
	0.03	0.08	0.06	0.06	of estimates
	0.04	0.30	0.18	0.17	
	0.03	0.09	0.08	0.08	Averages of estimated
	0.03	0.07	0.07	0.07	asymptotic standard
	0.05	0.26	0.24	0.23	deviations of estimates

<div align="center">Model II ($\tau = 0.3$)</div>

β_i	M1	M2	M3	M4	Estimation method
0.00	-0.09	0.43	-0.08	-0.01	Averages of
1.00	1.17	2.54	0.11	0.94	estimates
-0.50	-0.19	-1.67	-0.15	-0.47	
0.500	-0.16	-2.47	2.30	0.63	
	0.03	8.30	2.11	0.11	Standard deviations
	0.07	65.85	9.23	0.53	of estimates
	0.05	23.24	8.65	0.38	
	0.06	119.19	19.49	0.91	
	0.04	506.28	2.18	1.12	Averages of estimated
	0.08	3959.80	11.54	6.41	asymptotic standard
	0.06	1346.54	7.36	3.54	deviations of estimates
	0.07	7077.31	20.72	11.58	

Table 2: *Simulation results for the models I and II.*

the adjusted estimators reduce the bias substantially, often practically to zero. The latter specifically holds for models with very small τ, like model V, but also in models I' and II' (not presented here) of CS, which differ from models I and II only by a smaller σ_δ^2, viz., $\sigma_\delta^2 = 0.01$ instead of 0.1 (and thus $\tau = 0.03$ instead of 0.3). In these models the ALS estimator (M2) exhibits almost no bias. The modified estimators M3 and M4 perform just as well in these cases or only slightly better, also in terms of their variability.

However for models I–IV we notice that M2 does not always fare that well and that M4 is in most (but not in all) cases the best estimator from the point of view of bias reduction. But more importantly, M4 has usually the smallest standard deviation among the adjusted estimators, whereas M2 and, to a much lesser degree, also M3 show extremely large standard deviations, rendering the estimates very unstable. This is particularly true for models II and IV with $k = 3$, whereas for the quadratic models I and III the differences in the various

Model III ($\tau = 0.39$)

β_i	M1	M2	M3	M4	Estimation method
0.25	0.30	0.26	0.13	0.23	Averages of
1.00	0.77	1.10	1.10	1.03	estimates
1.00	0.40	0.69	2.42	1.22	
	0.05	2.02	1.02	0.11	Standard deviations
	0.11	2.46	1.86	0.22	of estimates
	0.30	22.18	14.03	1.17	
	0.06	13.31	1.42	0.56	Averages of estimated
	0.12	13.91	1.72	0.69	asymptotic standard
	0.32	144.45	16.03	5.97	deviations of estimates

Model IV ($\tau = 0.18$)

β_i	M1	M2	M3	M4	Estimation method
0.00	0.26	-17.01	0.29	0.20	Averages of
1.00	-9.20	-3.32	2.79	-0.43	estimates
0.00	-0.01	1.50	-0.00	0.01	
-1.00	-0.15	-0.88	-1.15	-0.91	
	2.34	846.40	11.27	8.47	Standard deviations
	1.41	1048.26	8.36	4.39	of estimates
	0.24	127.16	1.74	1.30	
	0.08	80.41	0.62	0.34	
	3.48	20200.51	73.37	60.15	Averages of estimated
	1.61	27139.23	103.19	86.45	asymptotic standard
	0.25	3161.42	11.29	9.22	deviations of estimates
	0.08	2070.57	7.56	6.32	

Model V ($\tau = 0.01$)

β_i	M1	M2	M3	M4	Estimation method
0.00	-0.04	-0.05	-0.05	-0.05	Averages of
1.00	-0.92	1.42	0.97	0.86	estimates
0.00	0.00	0.00	0.00	0.00	
-1.00	-0.85	-1.03	-1.00	-0.99	
	0.87	1.16	1.12	1.10	Standard deviations
	0.85	1.21	0.97	0.96	of estimates
	0.18	0.23	0.22	0.22	
	0.07	0.10	0.08	0.08	
	1.45	1.04	1.01	1.00	Averages of estimated
	0.81	1.04	1.01	1.00	asymptotic standard
	0.13	0.20	0.20	0.20	deviations of estimates
	0.05	0.08	0.08	0.08	

Table 3: *Simulation results for the models III–V.*

adjusted methods are not that prominent albeit still noticeable. The estimated asymptotic standard deviations are even larger than the observed ones and in some cases (M2 in model IV) completely out of any reasonable range. In these situations M4 produces the smallest asymptotic standard deviations.

For model V and similarly for models I' and II' (not presented here) with rather small measurement errors ($\tau = 0.01$ and $\tau = 0.03$, respectively) the asymptotic standard deviations do almost coincide with the empirical standard deviations of the estimates and, for the adjusted estimators, are only slightly larger than for the naive estimator. The same holds for the adjusted estimator M3, which in this case surpasses M4 with respect to the bias. Actually, if we examine the average estimates of the parameter β_1 in model V for both methods, we realize that M4 tends to overcorrect for bias if compared to M2 in this situation whereas M3 shows satisfactory results. This effect disappears when we shift the interval of the ξ_i values to the right and simulate a polynomial shifted in the same way. Then both estimators M3 and M4 perform almost in the same manner. This indicates exemplarily the dependency of the adjusted estimators not only on the amount of measurement error but also on the specific type of polynomial to be fitted.

It is interesting to note that Moon and Gunst (1995) considered the ALS method M2 as being very inefficient (for small samples), and this is born out by our simulation results. However its modifications, in particular M4, do not show that deficiency any more.

6 Earthquake Data

To see how our method works in practice we applied it to data that was analysed by Fuller (1987, p. 214) with the help of a quadratic model. The data consists of 43 measurements of the three-dimensional location of earthquake centers in a region near the Tonga trench. The variable y is the depth of the center of an earthquake, ξ_1 its distance on the earth's surface from the Tonga trench, and ξ_2 is a coordinate measuring the location of the earthquake parallel to the Tonga trench. For geophysical reasons the depth of an earthquake is nonlinearly related to its distance ξ_1 from the Tonga trench and linearly with respect to the other coordinate ξ_2. The nonlinear relation was approximated by a quadratic function in Fuller (1987). Here we experiment with a polynomial of third degree in order to find out whether a quadratic function is, in fact, sufficient or whether the third power of the distance variable needs to be introduced in the model. We thus work with the following regression

$$y_i = \beta_0 + \beta_1 \xi_{1i} + \beta_2 \xi_{1i}^2 + \beta_3 \xi_{1i}^3 + \beta_4 \xi_{2i} + \varepsilon_i, \ i = 1, \ldots, 43, \tag{20}$$

with the purpose in mind to test the hypothesis $\beta_3 = 0$. The data can be found in Fuller (1987, p. 289). All variables are measured in 100 km. The variables

ξ_{1i} and ξ_{2i} are measured with errors δ_{1i} and δ_{2i}, both with the same standard deviation σ_δ, which is known to be 10 km. Thus $x_{ki} = \xi_{ki} + \delta_{ki}$, $k = 1, 2$, with $\sigma_{\delta_k}^2 = \sigma_\delta^2 = 0.01$. The errors ε, δ_1, and δ_2 are independent, and the normality assumption (N) is adopted.

The model (20) differs from the basic polynomial model in that it contains a second (error-ridden) variable ξ_2 in addition to the variable ξ_1, which appears in the polynomial part of the model. Nevertheless the same approach as developed for the basic model can be chosen to analyse the present model.

The matrix H_i is modified by adding the column (and corresponding row)

$$(t_{oi}t_{1i}^*, \ldots, t_{3i}t_{1i}^*, t_{2i}^*)',$$

where $t_{1i}^* = x_{2i}$ and $t_{2i}^* = x_{2i}^2 - \sigma_\delta^2$. Similarly the vector t_i is augmented by the element $t_{1i}^* = x_{2i}$. With these modified matrices H_i and vectors t_i the matrices $\hat{V}_i$ are computed according to (8) and the matrices A, B, and $\hat{B}$ are constructed with $\bar{v} = 0$ and $\sigma_{\delta\varepsilon} = 0$. We then estimated the parameters of (20) in the same way as for the one variable polynomial model.

The results are presented in table 4. For the naive, the adjusted, and the second modified adjusted estimation method, estimates of the parameters $\beta_0, \ldots, \beta_4$ are shown together with their asymptotic standard deviations. In this case, probably due to the rather small error variances, the estimates of the two last methods do not differ very much, but both of them differ to some degree from the estimates of the naive method. Except for the last parameter β_4 all other parameters are not significant. This is certainly due to the high collinearity of the powers of ξ_1 and does not mean that ξ_1 does not have any influence on y. The smallest t-value is the one for β_3: $|\hat{\beta}_3|/\hat{\sigma}_{\hat{\beta}_3} = 0.517$, which suggests that ξ_1^3, being quite insignificant, should be dropped from the regression.

Estimation method	β_0	β_1	β_2	β_3	β_4
Naive	-0.129	0.274	0.274	-0.027	0.074
	(0.134)	(0.280)	(0.176)	(0.032)	(0.033)
Adjusted	-0.164	0.374	0.205	-0.013	0.075
	(0.092)	(0.293)	(0.176)	(0.029)	(0.028)
2nd Modified	-0.158	0.361	0.214	-0.015	0.075
	(0.092)	(0.294)	(0.176)	(0.029)	(0.028)

Table 4: *Earthquake data: Estimated parameters values and their asymptotic standard errors (given in parantheses) for the naive, the adjusted and the second modified adjusted estimation method.*

The resulting quadratic (in ξ_1) regression can be estimated. We present only the results for the modified adjusted estimation method:

$$\hat{y}_i = \quad -0.200 \quad +0.481\,\xi_{1i} \quad +0.132\,\xi_{1i}^2 \quad +0.077\,\xi_{2i},$$
$$\phantom{\hat{y}_i =} \quad (0.075) \quad (0.109) \quad\quad (0.032) \quad\quad (0.031)$$

where the numbers in parentheses are the estimated standard errors. They correspond closely to Fuller's result, who only studied the quadratic case. It is seen that the estimates are now all significant.

7   Conclusion and discussion

If the naive estimator of a regression with errors in the variables, i.e., the estimator which disregards the measurement errors, is adjusted for the errors with the help of the error variance and possibly also of higher moments of the error variable, then a consistent estimator results: the adjusted least squares (ALS) estimator. This estimator, however, is extremely unstable for small samples, particularly if the error variance is high relative to the variance of the regressor. Various modifications of the ALS estimator are possible to remedy this situation without destroying the consistency and other asymptotic properties of the estimator. We discussed two such modifications, though other modifications are also possible. They do not seem to differ very much in their results, but rather all agree in reducing the small sample variance considerably.

When Moon and Gunst (1995) in their simulation study judged the unmodified ALS estimator to be very unstable, they were right, but they were wrong in their general judgement that "specialized estimators for polynomial measurement error models do not appear to be needed." By "specialized estimators" they meant among others unmodified ALS. Had they considered modified ALS estimators (and not just the unmodified one) they would have been forced to qualify their statement. Also, what they called the general nonlinear estimator, which performed best in their simulations, is not so general after all. It uses more information than ALS, namely, knowledge of *both* error variances, of regressor and of regressand, and therefore cannot be compared with ALS or any of its modifications, because these methods use only the error variance of the regressor. In addition, the general nonlinear estimator, despite its good performance in their simulation study, is not consistent in the usual sense, see Wolter and Fuller (1982).

ACKNOWLEDGEMENTS

REFERENCES

AMEMIYA, Y. (1985) What should be done when an estimated between - group covariance matrix is not nonnegative definite? *The American Statistican* **39**, 112-117.

BUONACCORSI, J. P. (1996) A modified estimating equation approach to correcting for measurement error in regression. *Biometrika*, 433-440.

CHAN, L. K. and MAK, M. K. (1985) On the polynomial functional relationship. *J. R. Statist. Soc. B*, **47**, 510-518.

CHENG, C. L. and SCHNEEWEISS, H. (1998) The polynomial regression with errors in the variables. *J. R. Statist. Soc. B*, **60**, 189-199.

FULLER, W. A. (1987) *Measurement Error Models*. New York: Wiley.

GRILICHES, Z. and RINGSTAD, V. (1970) Errors-in-variables bias in nonlinear contexts. *Econometrica*, **38**, 368-370.

MOON, M.S. and GUNST, R.F. (1995) Polynomial measurement error modelling. *Computational Statistics & Data Analysis.* **19**, 1-21.

NAKAMURA, T. (1990) Corrected score function for errors-in-variables models. *Biometrika* **77**, 127-137.

STEFANSKI, L. A. (1989) Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Statist.–Theory Meth.* **18**, 4335-4358.

WOLTER, K. M. and FULLER, W. A. (1982) Estimation of nonlinear errors-in-variables models. *Ann. Statist.* **10**, 539-548.