J. R. Statist. Soc. B (1998)
60, *Part* 1, *pp.* 189–199

Polynomial regression with errors in the variables

Chi-Lun Cheng

Academia Sinica, Taipei, Republic of China

and Hans Schneeweiss†

University of Munich, Germany

[Received August 1996. Revised March 1997]

Summary. A polynomial functional relationship with errors in both variables can be consistently estimated by constructing an ordinary least squares estimator for the regression coefficients, assuming hypothetically the latent true regressor variable to be known, and then adjusting for the errors. If normality of the error variables can be assumed, the estimator can be simplified considerably. Only the variance of the errors in the regressor variable and its covariance with the errors of the response variable need to be known. If the variance of the errors in the dependent variable is also known, another estimator can be constructed.

Keywords: Adjusted least squares; Errors in variables; Estimating equations; Functional relationship; Method of moments; Polynomial regression

1. Introduction

Whereas the linear functional relationship has found extensive treatment in the literature — for some reviews see, for example, Madansky (1959), Kendall and Stuart (1979), chapter 29, Schneeweiss and Mittag (1986), Fuller (1987), chapter 2, Cheng and Van Ness (1994) — the polynomial functional relationship has received only a little attention despite the fact that it is perhaps the most natural extension of the linear model towards a non-linear model.

The polynomial functional relationship is given by the equations

$$y_i = \eta_i + \epsilon_i = \beta_0 + \beta_1 \xi_i + \beta_2 \xi_i^2 + \ldots + \beta_k \xi_i^k + \epsilon_i,$$
$$x_i = \xi_i + \delta_i,$$

where the errors (δ_i, ϵ_i), $i = 1, \ldots, n$, are independent and identically distributed (IID) pairs of random variables with expectation 0 and with covariance matrix

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

The ξ_i, $i = 1, \ldots, n$, are unobservable (latent) non-stochastic variables, sometimes also called incidental parameters. This is the so-called functional case of an errors-in-variables model. In the structural case, which we shall only briefly touch on in Section 8, the ξ_i are assumed to be IID random variables. The x_i and y_i are the only observable variables. The parameters to be

†*Address for correspondence*: Seminar für Ökonometrie und Statistik, Ludwig Maximilians-Universität München, Akademiestraße 1/I, D-80799 München, Germany.
E-mail: schneew@stat.uni-muenchen.de

 1369–7412/98/60189

estimated are $\beta_0, \ldots, \beta_k$ and possibly also all or some elements of Ω. Typically knowledge of Ω or at least of part of Ω is required.

Kendall and Stuart (1979), chapter 29, applied maximum likelihood to the functional quadratic model assuming the errors δ and ϵ to be uncorrelated normally distributed and their variance ratio known. O'Neill *et al.* (1969) refined the estimation procedure, so that it can be effectively carried out on a computer; for an application in botany see Causton and Venus (1981). However, because of the incidental parameters it is not clear whether the resulting estimator has the usual large sample properties of maximum likelihood estimators; see Egerton and Laycock (1979) and also Fuller (1987), p. 247. Wolter and Fuller (1982) proposed an estimator for the parameters in the quadratic model when the errors are normally distributed with Ω known; for extensions to the polynomial case see Moon and Gunst (1994, 1995). Van Montfort (1988) has discussed the identification and estimation of a structural quadratic relationship with normally distributed ξ_i and unknown error variances; see also Barton and David (1960) for an example in astronomy. Griliches and Ringstad (1970) investigated the bias of ordinary least squares (OLS) in a structural quadratic relationship. Recently Carroll *et al.* (1995) presented general approaches to estimating non-linear measurement error models such as regression calibration and SIMEX. However, these techniques give only approximately consistent estimates in general. When applied to a structural quadratic model, regression calibration gives consistent estimates if ξ and δ are normally distributed, $\sigma_{\delta\epsilon} = 0$, and σ_δ^2 known; see Carroll *et al.* (1995), p. 68. Fuller (1987), p. 212 (and following pages), estimated a functional quadratic relationship with $\sigma_{\delta\epsilon} = 0$ and σ_δ^2 known by a method which differs from our approach but gives the same result. It can be generalized to a method for estimating a polynomial of any degree and with $\sigma_{\delta\epsilon} \neq 0$, but we shall not do so here.

Chan and Mak (1985) found estimates for the β_j by first applying a variant of least squares, which they call generalized least squares (GLS), to the model by assuming the ξ_i to be known and then replacing the resulting moments in the unknown ξ_i by unbiased estimates, which can be constructed as polynomials in x_i with coefficients depending on higher moments of the errors assumed to be known. In the important case of jointly normally distributed errors, i.e. under the assumption

$$(\delta_i, \epsilon_i) \sim N(0, \Omega), \tag{1}$$

it suffices to know Ω.

We shall follow this general idea of Chan and Mak (1985) but our treatment of the polynomial functional relationship will differ in several important ways from theirs.

(a) Chan and Mak constructed their estimator of $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ under assumption (1). We can give a more general result without this assumption.
(b) Whereas Chan and Mak assumed the whole of Ω to be known, we can show that their estimator can also be derived assuming, under assumption (1), only knowledge of σ_δ^2 and $\sigma_{\delta\epsilon}$.
(c) Whereas Chan and Mak used a GLS approach, we start by using OLS for the polynomial regression with ξ_i known. This approach leads to a simpler and, we hope, more lucid derivation of the estimator and ultimately also to a simpler formula for the estimator. For further details see Cheng and Schneeweiss (1996).
(d) The simplicity is enhanced by treating the intercept β_0 in the same way as the other regression parameters β_i and not separating β_0 from the rest of the β_i as in Chan and Mak (1985).

(e) Under assumption (1) we can greatly simplify the estimating equations by finding a simple explicit formula for these equations.
(f) Since we do not assume the whole of Ω to be known, σ_ϵ^2 is an unknown parameter for us, which we can estimate by the same principle as is used in estimating the β_i.
(g) The estimators of β and of σ_ϵ^2 serve as a starting point for an iterative procedure of estimating β and Ω when Ω is known only up to a factor of proportionality. This case is a new case for us, which is different from the main case of this paper where Ω is assumed to be known except for σ_ϵ^2. For Chan and Mak the case of Ω known up to a factor of proportionality is a generalization of their Ω-known case.
(h) Chan and Mak gave a separate treatment of the quadratic relationship. They argued that in this case they can construct an estimator without the need for assumption (1). However, this is not quite true. They do need the assumption that the third-order moments are those of expression (1), i.e. are 0. It is true, however, that the fourth-order moments need not be known. We give a new, simple and transparent derivation of this estimator and show how it can be generalized for a polynomial relationship of any degree. The relationship of this special estimator for the quadratic relationship to the general procedure as developed by Chan and Mak is not clear in their treatment as both estimators apparently use the same information, i.e. knowledge of Ω. Our approach makes it clear that the two estimators differ not only in the way that they are derived and constructed but also in the information that is used for their construction: the new estimator uses the whole of Ω; the previous estimator uses only σ_δ^2 and $\sigma_{\delta\epsilon}$. A similar distinction can be made for the extension of this new estimator to polynomials of higher degree.

In our view, point (b) is the most important improvement on Chan and Mak's method. The difference between the two cases 'Ω known' and 'Ω known except for σ_ϵ^2' seems slight but is, in fact, fundamental.

In general, the error term ϵ will typically contain an error in the equation ϵ^{**} in addition to the measurement error ϵ^* of η, i.e. ϵ is decomposable as

$$\epsilon = \epsilon^{**} + \epsilon^*,$$

where ϵ^{**} will usually be independent of (ϵ^*, δ). Now, it is quite possible that an investigator has knowledge, or at least an estimate, of the joint distribution of the measurement errors ϵ^* and δ, but he will hardly ever have any prior knowledge about ϵ^{**}. In particular, under assumption (1), knowledge of

$$\Omega^* = \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\epsilon} \\ \sigma_{\delta\epsilon} & \sigma_{\epsilon^*}^2 \end{pmatrix}$$

can be gained from replicated measurements or from former experience about the working of the measurement process. But, when there is an error in the equation ϵ^{**}, its variance will typically not be known beforehand. (We may gain some knowledge about $\sigma_{\epsilon^{**}}^2$ *after* having estimated all the other parameters.) This variance, however, is part of $\sigma_\epsilon^2 = \sigma_{\epsilon^*}^2 + \sigma_{\epsilon^{**}}^2$, and so, even if Ω^* is known, Ω will not be known, not even up to a factor of proportionality. However, if Ω^* is known, so is the first row of Ω, i.e. $(\sigma_\delta^2, \sigma_{\delta\epsilon})$. (Note that $\sigma_{\delta\epsilon^*} = \sigma_{\delta\epsilon}$.)

It is only in the case where errors in the equations are missing that Ω can reasonably be assumed to be known ($\Omega = \Omega^*$), or at least known up to a factor of proportionality. So the two major cases of prior knowledge about the error process, i.e. the cases Ω known (possibly only up to a factor of proportionality) and Ω known except for σ_ϵ^2 are strongly related, though not exactly the same, to the two cases of a relationship without and with errors in the equation respectively. No doubt the difference between these two cases is fundamental. For further discussions of the role of errors in the equation see Fuller (1987), p. 106, and Carroll *et al*. (1995), p. 30.

Chan and Mak's (1985) replacement principle can be generalized to the idea of replacing the whole likelihood function or the score function of an error-free model by an 'estimate' in terms of the observable, error-ridden, variables; see Nakamura (1990), Stefanski (1989) and Buonaccorsi (1996). In particular, Stefanski treated the polynomial regression in this way, but his theory is restricted to the case of normally distributed errors, and he did not consider an error in the dependent variable. Buonaccorsi gave a solution to the case of a quadratic regression but did not treat polynomial regressions.

After a short presentation of our model in Section 2, we shall introduce the error adjusted least squares (ALS) estimator in Section 3 and show how it can be computed under general conditions. Section 4 demonstrates how the estimator dramatically simplifies under the assumption of normality in the errors. Section 5 briefly deals with the case when Ω is known up to a factor of proportionality, and Section 6 introduces a new method of moments that is applicable for the case when Ω is known. Section 7 has some Monte Carlo results, and Section 8 contains some concluding remarks. We shall use a notation similar to that of Chan and Mak (1985) so that our results can be better compared with theirs.

2. The model

We consider the polynomial functional relationship

$$\begin{aligned} y_i &= \zeta_i'\beta + \epsilon_i = \eta_i + \epsilon_i, \\ x_i &= \xi_i + \delta_i \end{aligned} \tag{2}$$

with $\zeta_i' = (1, \xi_i, \xi_i^2, \ldots, \xi_i^k)$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$, where the ξ_i, $i = 1, \ldots, n$, are the n sample values of a latent non-stochastic variable ξ and the (δ_i, ϵ_i), $i = 1, \ldots, n$, are IID pairs of random errors with expectation 0 and covariance matrix Ω as given in Section 1.

It is assumed that all moments $E(\delta^l)$, $l = 1, \ldots, 2k$, and $E(\delta^l\epsilon)$, $l = 1, \ldots, k$, are known. With the further assumption (1), which we shall adopt in Sections 5 and 6, only knowledge of $\sigma_\delta^2 = E(\delta^2)$ and $\sigma_{\delta\epsilon} = E(\delta\epsilon)$ is required. The variance σ_ϵ^2 of ϵ need not be known, except for an alternative estimator proposed in Section 6.

3. The adjusted least squares estimator

If the ξ_i, $i = 1, \ldots, n$, were known, OLS could be used to estimate β, yielding the estimating equations

$$\overline{\zeta\zeta'}\hat{\beta}_{\mathrm{OLS}} = \overline{\zeta y}, \tag{3}$$

where the bars denote averages, i.e.

$$\overline{\zeta\zeta'} = \frac{1}{n}\sum \zeta_i\zeta_i'$$

etc.

Following Chan and Mak (1985) a least squares estimator adjusted for errors (ALS) is constructed by replacing the unknown matrix $\overline{\zeta\zeta'}$ and the unknown vector $\overline{\zeta y}$ by certain estimates, i.e. by a known matrix $\overline{H}$ and a known vector $\overline{h}$ respectively, such that $E(\overline{H}) = \overline{\zeta\zeta'}$ and $E(\overline{h}) = E(\overline{\zeta y}) = \overline{\zeta\eta}$.

Let us first consider the $(k+1)\times(k+1)$ matrix $\zeta\zeta'$ with $\zeta = (\xi^0, \ldots, \xi^k)'$. Here and in what follows we omit the observational index i for ease of notation. The (p, q) element of $\zeta\zeta'$ is given by ξ^{p+q}; $p, q = 0, \ldots, k$. We are thus led to look for a variable t_r to be computed from the observable x such that $E(t_r) = \xi^r$. Indeed there are uniquely defined polynomials t_r in x of degree r, $r = 0, \ldots, k$, such that $E(t_r) = \xi^r$. These can be found by noting that, because of equations (2),

$$E(x^r) = \sum_{j=0}^{r} \binom{r}{j} \xi^j \, E(\delta^{r-j}) = \sum_{j=0}^{r} c_{rj}\, \xi^j$$

is a polynomial in ξ with coefficients $c_{rj} = \binom{r}{j}\, E(\delta^{r-j})$. If we replace $E(x^r)$ by x^r and ξ^j by a new variable t_j, then the resulting recursive system of equations

$$x^r = \sum_{j=0}^{r} c_{rj} t_j, \tag{4}$$

$r = 0, \ldots, k$, can be easily solved, one by one, for the t_j, yielding

$$t_r = \sum_{j=0}^{r} a_{rj} x^j, \tag{5}$$

say, where the a_{rj} are functions of $E(\delta^l)$, $l = 0, \ldots, r$. Obviously $E(t_r) = \xi^r$ by construction. Using this definition of the t_r we find the first five t_r as $t_0 = 1$, $t_1 = x$, $t_2 = x^2 - \sigma_\delta^2$, $t_3 = x^3 - 3x\sigma_\delta^2 - E(\delta^3)$ and $t_4 = x^4 - 6x^2\sigma_\delta^2 - 4x\, E(\delta^3) - E(\delta^4) + 6\sigma_\delta^4$.

Now let $H = H(x)$ be a $(k+1)\times(k+1)$ matrix the (p, q) element of which is t_{p+q}, $p, q = 0, \ldots, k$; then obviously $E(H) = \zeta\zeta'$ and, letting

$$\overline{H} = \frac{1}{n} \sum_{i=1}^{n} H(x_i),$$

also $E(\overline{H}) = \overline{\zeta\zeta'}$.

Next consider the $(k+1)$-vector $\zeta\eta$, which has elements $\xi^r\eta$, $r = 0, \ldots, k$. We have

$$\xi^r\eta = E(t_r\eta) = E(t_r y) - E(t_r\epsilon)$$

and by equation (5)

$$E(t_r\epsilon) = E\left\{ \sum_{j=0}^{r} a_{rj} (\xi+\delta)^j \epsilon \right\} = \sum_{j=0}^{r} b_{rj}\, \xi^j$$

with coefficients

$$b_{rj} = \sum_{s=j}^{r} a_{rs} \binom{s}{j} \, E(\delta^{s-j}\epsilon), \tag{6}$$

which depend only on $E(\delta^l)$ (via the coefficients a_{rs}) and $E(\delta^l\epsilon)$, $l = 0, \ldots, r$, and can be easily computed once these moments are known and the a_{rs} have been determined as in equation (5). A natural estimate of $E(t_r\epsilon)$ therefore is $\hat{E}(t_r\epsilon) = \Sigma\, b_{rj} t_j$.

Hence if we define

$$h_r = t_r y - \hat{E}(t_r \epsilon) = t_r y - \sum_{j=0}^{r} b_{rj} t_j, \tag{7}$$

then obviously $E(h_r) = \xi^r \eta$. Now let $h = h(x, y) = (h_0, h_1, \ldots, h_k)'$; then $E(h) = \zeta\eta$ and hence $E(\overline{h}) = \overline{\zeta\eta} = E(\overline{\zeta y})$, where

$$\overline{h} = \frac{1}{n} \sum_{i=1}^{n} h(x_i, y_i).$$

The first three elements of h are $h_0 = y$, $h_1 = xy - \sigma_{\delta\epsilon}$, $h_2 = (x^2 - \sigma_\delta^2)y - E(\delta^2\epsilon) - 2\sigma_{\delta\epsilon}x$. In the important special case of independence between δ and ϵ the term $\hat{E}(t_r\epsilon)$ vanishes and $\overline{h}$ reduces to $\overline{ty}$, where $t = (t_0, t_1, \ldots, t_k)'$.

The estimating equation for the ALS estimator $\hat{\beta} = \hat{\beta}_{\mathrm{ALS}}$ of β is now constructed by replacing $\overline{\zeta\zeta'}$ and $\overline{\zeta y}$ in equation (3) by $\overline{H}$ and $\overline{h}$ respectively:

$$\overline{H}\hat{\beta} = \overline{h}. \tag{8}$$

In the quadratic case, $k = 2$, the estimating equations (8) are easily set up by constructing $\overline{H}$ and $\overline{h}$, with the first five t_r and the first three h_r as given above.

Once $\hat{\beta}$ has been found as the solution of equation (8), i.e. $\hat{\beta} = \overline{H}^{-1}\overline{h}$, an estimator of σ_ϵ^2 can also be constructed. The OLS estimator of σ_ϵ^2 in model (2) with ξ known is

$$\hat{\sigma}^2_{\epsilon,\mathrm{OLS}} = \overline{y^2} - \overline{y\zeta'}\,\hat{\beta}_{\mathrm{OLS}}.$$

Replacing $\overline{y\zeta'}$ by $\overline{h'}$ and $\hat{\beta}_{\mathrm{OLS}}$ by $\hat{\beta}$ we obtain the ALS estimator

$$\hat{\sigma}^2_\epsilon = \overline{y^2} - \overline{h}'\hat{\beta} = \overline{y^2} - \overline{h}'\overline{H}^{-1}\overline{h}. \tag{9}$$

Under general assumptions, the ALS estimators of β and σ_ϵ^2 are consistent. In fact, we need only to assume that higher moments of δ and ϵ (more specifically $E(\delta^{4k})$ and $E(\delta^{2k}\epsilon^2)$) exist, that the limits of higher moments of ξ, i.e. $\lim(\overline{\xi^r})$, exist for $r = 0, \ldots, 4k$ and that $\lim(\overline{H})$ is non-singular. Under similar assumptions $\hat{\beta}$ is asymptotically normal, i.e. $n^{1/2}(\hat{\beta} - \beta) \rightarrow N(0, \Sigma)$, and the variance of the approximating normal distribution of $\hat{\beta}$ can be estimated by

$$\hat{\Sigma} = \frac{1}{n}\,\overline{H}^{-1}\hat{V}\overline{H}^{-1}$$

with

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}(H_i\hat{\beta} - h_i)(H_i\hat{\beta} - h_i)'.$$

(For a general procedure of finding the asymptotic covariance matrix in the context of unbiased estimating equations or M-estimators see Huber (1967) or Carroll *et al.* (1995), p. 262.)

4. The adjusted least squares estimator under assumption (1)

If (δ, ϵ) is jointly normally distributed, only $\sigma_{\delta\epsilon}$ and σ_δ^2 need to be known. In this case simple formulae for the computation of t_r and $\hat{E}(t_r\epsilon)$ and hence of H and h can be derived. In particular, the t_r, $r = 0, 1, 2, \ldots,$ can be computed by a simple recursion formula.

Lemma 1. Under assumption (1)

$$t_{r+1} = xt_r - \sigma_\delta^2 r t_{r-1} \tag{10}$$

for $r = 0, 1, 2, \ldots$ with $t_0 = t_{-1} = 1$.

Proof. For $\sigma_\delta^2 = 1$ the recursion formula (10) defines the Hermite polynomials in x, and for these the expectation equals ξ^r if $x \sim N(\xi, 1)$; see Stulajter (1978). It then follows easily that $E(t_r) = \xi^r$ for general σ_δ^2; see also Stefanski (1989). For a different, more explicit, proof see Cheng and Schneeweiss (1996). □

As for $\hat{E}(\overline{t\epsilon})$, we find the following lemma.

Lemma 2. Under assumption (1)

$$\hat{E}(t_r\epsilon) = \sigma_{\delta\epsilon} r t_{r-1}.$$

Proof. First note that, using equation (5),

$$E(t_r\epsilon) = \sum_{j=0}^{r} a_{rj}\, E\{(\xi + \delta)^j \epsilon\}$$

can be written as

$$E(t_r\epsilon) = \sum_{l=0}^{r} d_{rl}(\xi)\, E(\delta^l \epsilon),$$

where the $d_{rl}(\xi)$ are polynomials in ξ. But also $E(t_r\delta) = \Sigma\, d_{rl}(\xi) E(\delta^{l+1})$. Under assumption (1), $E(\delta^l\epsilon) = (\sigma_{\delta\epsilon}/\sigma_\delta^2)\, E(\delta^{l+1})$ and hence $E(t_r\epsilon) = (\sigma_{\delta\epsilon}/\sigma_\delta^2)\, E(t_r\delta)$.

Now, because of equation (10), $E(t_r\delta) = E(t_r x - t_r\xi) = E(t_{r+1} + \sigma_\delta^2 r t_{r-1} - t_r\xi) = \sigma_\delta^2 r \xi^{r-1}$. It follows that $E(t_r\epsilon) = \sigma_{\delta\epsilon} r \xi^{r-1}$. □

We can now write the estimating equation (8) for $\hat{\beta}$ as

$$\overline{H}\hat{\beta} = \overline{ty} - \sigma_{\delta\epsilon}\overline{p}, \tag{11}$$

where $p = (0, t_0, 2t_1, \ldots, kt_{k-1})'$.

For the quadratic functional relationship ($k = 2$) equation (11) becomes

$$\left.\begin{aligned} \beta_0 + \overline{x}\beta_1 + (\overline{x^2} - \sigma_\delta^2)\beta_2 &= \bar{y}, \\ \overline{x}\beta_0 + (\overline{x^2} - \sigma_\delta^2)\beta_1 + (\overline{x^3} - 3\overline{x}\sigma_\delta^2)\beta_2 &= \overline{xy} - \sigma_{\delta\epsilon}, \\ (\overline{x^2} - \sigma_\delta^2)\beta_0 + (\overline{x^3} - 3\overline{x}\sigma_\delta^2)\beta_1 + (\overline{x^4} - 6\overline{x^2}\sigma_\delta^2 + 3\sigma_\delta^4)\beta_2 &= \overline{x^2 y} - \overline{y}\sigma_\delta^2 - 2\sigma_{\delta\epsilon}\overline{x}. \end{aligned}\right\} \tag{12}$$

5. Ω known up to a factor of proportionality

Chan and Mak (1985) also studied the case where, under assumption (1),

$$\Omega = \kappa W$$

with W known and κ an unknown parameter. To them this case is a generalization of their case Ω known. To us it is simply a different case of prior knowledge about Ω: instead of the assumption that Ω is known except for σ_ϵ^2 we now assume that Ω is known up to a factor of proportionality κ.

Nevertheless, this new case can be dealt with by reducing it to the previous case, though other approaches are also possible; see, for example, Wolter and Fuller (1982) for the quadratic functional relationship.

Following an idea of Chan and Mak (1985), but with our own estimation procedure for $\hat{\beta}$ and $\hat{\sigma}_\epsilon^2$ as developed in Section 3, we propose to solve equations (8), (9) and

$$\hat{\Omega} = \hat{\kappa} W \tag{13}$$

simultaneously for $\hat{\beta}$, $\hat{\Omega}$ and $\hat{\kappa}$, where in equations (8) and (9) $\overline{H}$ and $\overline{h}$ are to be considered as functions of $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_{\delta\epsilon}$ (see lemmas 1 and 2).

More specifically, the following iterative procedure is suggested. Begin with a starting value κ_0 for κ, e.g. $\kappa_0 = 1$. If κ_m is the value for κ after the mth iteration, compute $\Omega_m = \kappa_m W$, use $\sigma_{\delta m}^2$ and $\sigma_{\delta\epsilon m}$ to compute $\overline{H}_m$ and $\overline{h}_m$ with the help of lemmas 1 and 2 and derive the mth iteration value for $\hat{\beta}$ and $\hat{\sigma}_\epsilon^2$, i.e. $\hat{\beta}_m$ and $\hat{\sigma}_{\epsilon m}^2$, from equations (8) and (9). Finally find the next iteration value for κ from $\kappa_{m+1} = \hat{\sigma}_{\epsilon m}^2 / w_\epsilon^2$, where w_ϵ^2 is the element of W corresponding to σ_ϵ^2. The iterative cycle is then repeated until (hopefully) convergence is attained.

For the linear functional relationship ($k = 1$) this procedure will boil down to the well-known linear functional relationship estimator with Ω known up to a factor of proportionality; see, for example Fuller (1987), chapter 2.

6. A method of moments

Chan and Mak (1985) treated the quadratic functional relationship with another method, specially developed for this case. They claimed that they need no normality assumption for this case, but they do need to know the complete error covariance matrix Ω. Though they do not state it that way, their estimation procedure stems from a method of moments, which can be easily generalized to polynomial relationships of any degree.

Let us introduce the null variate (again omitting the observational index i)

$$u = y - \sum_{r=0}^{k} \beta_r t_r = y - t'\beta. \tag{14}$$

Since

$$u = \epsilon - (t - \zeta)'\beta, \tag{15}$$

we obviously have $E(u) = 0$. We can also compute $E(x^r u) = E(x^r \epsilon) - E\{x^r (t - \zeta)'\beta\}$ as a linear function of $\beta_0, \ldots, \beta_k$ with coefficients depending on the moments $E(\delta^l)$, $E(\delta^l \epsilon)$ and on the powers of ξ. If we replace these powers by the corresponding t-variates, i.e. ξ^r by t_r, and denote the resulting expression by $\hat{E}(x^r u)$, then the system of equations

$$\overline{x^r u} = \hat{E}\,(\overline{x^r u}), \tag{16}$$

$r = 0, \ldots, k$, serves as a linear system of estimating equations for $\beta_0, \ldots, \beta_k$.

It turns out that this system is the same as equation (8); see Cheng and Schneeweiss (1996). However, if the last equation of system (16) ($r = k$) is replaced by

$$\overline{yu} = \hat{E}(\overline{yu}), \tag{17}$$

where $\hat{E}(\overline{yu})$ is computed in a similar way as $\hat{E}(\overline{x^r u})$, then the resulting system, i.e. equations (16) for $r = 0, \ldots, k - 1$ and equation (17) taken together, is a new system of estimating equations of $\beta_0, \ldots, \beta_k$, which differs from system (16), $r = 0, \ldots, k$, in that it can be set up with somewhat smaller powers of x than are needed for system (16) alone. It will therefore be more efficient than system (16) alone. This gain in efficiency, however, is paid for by the requirement of the additional knowledge of σ_ϵ^2. Indeed, since

$$E(yu) = E[(\eta+\epsilon)\{\epsilon-(t-\zeta)'\beta\}] = \sigma_\epsilon^2 - E(\epsilon t')\beta,$$

$$\hat{E}(\overline{yu}) = \sigma_\epsilon^2 - \hat{E}(\overline{\epsilon t}')\beta, \tag{18}$$

and under assumption (1) (see equation (11) and lemma 2)

$$\hat{E}(\overline{yu}) = \sigma_\epsilon^2 - \sigma_{\delta\epsilon}\overline{p}'\beta.$$

Thus this estimator of β is feasible only if the whole of Ω is known, i.e. typically in the case of no errors in the equation.

Let us see how this method of moments works in the quadratic case, $k=2$. In this case

$$\begin{aligned} u &= y-\beta_0-\beta_1 x-\beta_2(x^2-\sigma_\delta^2) \\ &= \epsilon-\beta_1\delta-\beta_2(\delta^2-\sigma_\delta^2)-2\beta_2\delta\xi. \end{aligned}$$

Under assumption (1) it follows that

$$\begin{aligned} E(u) &= 0, \\ E(xu) &= \sigma_{\delta\epsilon}-\beta_1\sigma_\delta^2-2\beta_2\sigma_\delta^2\xi, \\ E(yu) &= \sigma_\epsilon^2-\beta_1\sigma_{\delta\epsilon}-2\beta_2\sigma_{\delta\epsilon}\xi, \end{aligned}$$

where, in fact, not all of assumption (1) was used but only the property $E(\delta^3)=E(\delta^2\epsilon)=0$. Replacing ξ by $t_1=x$, we find the following system of linear estimating equations corresponding to equations (16) and (17):

$$\left.\begin{aligned} \overline{y}-\hat\beta_0-\hat\beta_1\overline{x}-\hat\beta_2(\overline{x^2}-\sigma_\delta^2) &= 0, \\ \overline{xy}-\hat\beta_0\overline{x}-\hat\beta_1\overline{x^2}-\hat\beta_2\overline{x(x^2-\sigma_\delta^2)} &= \sigma_{\delta\epsilon}-\hat\beta_1\sigma_\delta^2-2\hat\beta_2\overline{x}\sigma_\delta^2, \\ \overline{y^2}-\hat\beta_0\overline{y}-\hat\beta_1\overline{yx}-\hat\beta_2\overline{y(x^2-\sigma_\delta^2)} &= \sigma_\epsilon^2-\hat\beta_1\sigma_{\delta\epsilon}-2\hat\beta_2\overline{x}\sigma_{\delta\epsilon}, \end{aligned}\right\} \tag{19}$$

which can be solved for β_0, β_1 and β_2. This system is equivalent to that of Chan and Mak (1985) in their section 3.

This result shows that Chan and Mak cannot completely dispense with the normality assumption; they must assume $E(\delta^3)=E(\delta^2\epsilon)=0$. But it is true that they need not know $E(\delta^4)$, in contrast with the corresponding ALS estimator. This saving in knowledge of higher moments of the errors in x is compensated for by extra knowledge of σ_ϵ^2, which was not needed in the ALS method and therefore places the new method into a totally different category: it is a method that is suitable for polynomial relationships with no errors in the equation.

Note that this method is not restricted to the quadratic relationship, as in Chan and Mak (1985), but is applicable to any polynomial relationship as long as the required moments of the error variables are known.

7. Some Monte Carlo results

To study the working of the ALS estimator for a medium sample size, in particular in comparison with the naïve OLS estimator, some Monte Carlo studies were performed. These

Table 1. Simulated estimates of β by ALS and OLS

	Results for the following models:					
	Model I			*Model II*		
	True	*ALS*	*OLS*	*True*	*ALS*	*OLS*
β_0	0	0.001	−0.017	0	−0.001	−0.019
		0.016	*0.014*		*0.019*	*0.015*
β_1	1	1.003	0.974	1	0.981	1.130
		0.020	*0.019*		*0.082*	*0.046*
β_2	−0.5	−0.506	−0.439	−0.5	−0.500	−0.433
		0.039	*0.031*		*0.066*	*0.042*
β_3				0.5	0.537	0.209
					0.150	*0.073*

are certainly not very extensive and are only meant as an example of what we might expect when the estimation method is applied to real data. For another simulation study see Moon and Gunst (1995).

The following two polynomial regressions of degrees 2 and 3 were investigated: model I with $\beta = (0, 1, -0.5)$ and model II with $\beta = (0, 1, -0.5, 0.5)$. For both models $\sigma_\delta^2 = \sigma_\epsilon^2 = 0.01$ and $\sigma_{\delta\epsilon} = 0$. The errors were taken to be normally distributed. The ξ_i were fixed at the points $\xi_i = -1 + 0.01i$, $i = 0, \ldots, 200$, i.e. they were uniformly placed in the interval $[-1, 1]$.

For each model 100 replicated samples were simulated and β was estimated by ALS and OLS. For each model and each estimation method, the mean and standard deviation (the latter in italics) of these 100 estimates were computed and are shown in Table 1.

Despite the small measurement error variance σ_δ^2 (compared with the variance of the ξ_i), the OLS estimator has a rather large bias at least for some of the parameters, notably for β_3 in model II. In all cases ALS reduces the bias to practically 0. As was to be expected, the variances are larger for ALS than for OLS, though not so large as to set off the reduction of the bias.

8. The structural case

Although we derived the ALS and the method-of-moments estimators for the functional polynomial relationship, the same estimators can also be applied to the structural relationship, i.e. to the case where the latent true regressor variable ξ is a random variable with an unknown distribution. The stochastic properties of the estimators (consistency and asymptotic normality) will not change.

Acknowledgements

Parts of this paper were written when the first author was visiting the Institute of Statistics, University of Munich. He gratefully acknowledges a fellowship from the National Science Council of Taiwan, Republic of China. The authors thank M. Thamerus for his help in carrying out the Monte Carlo simulation.

References

Barton, D. E. and David, F. N. (1960) Models of functional relationship illustrated on astronomical data. *Bull. Int. Statist. Inst.*, **37**, 9–33.

Buonaccorsi, J. P. (1996) A modified estimating equation approach to correcting for measurement error in regression. *Biometrika*, **83**, 433–440.

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

Causton, D. R. and Venus, J. C. (1981) *The Biometry of Plant Growth*. London: Arnold.

Chan, L. K. and Mak, T. K. (1985) On the polynomial functional relationship. *J. R. Statist. Soc.* B, **47**, 510–518.

Cheng, C.-L. and Schneeweiss, H. (1996) The polynomial regression with errors in the variables. *Discussion Paper 42, Sonderforschungsbereich 386*. University of Munich, Munich.

Cheng, C.-L. and Van Ness, J. W. (1994) On estimating linear relationships when both variables are subject to errors. *J. R. Statist. Soc.* B, **56**, 167–183.

Egerton, M. F. and Laycock, P. J. (1979) Maximum likelihood estimation of multivariate non-linear functional relationships. *Math. Operforsch. Statist. Ser. Statist.*, **10**, 273–280.

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.

Huber, P. J. (1967) The behavior of maximum likelihood estimates under nonstandard condition. In *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, pp. 221–233. Berkeley: University of California Press.

Kendall, M. G. and Stuart, A. (1979) *The Advanced Theory of Statistics*, 4th edn, vol. 2. London: Griffin.

Madansky, A. (1959) The fitting of straight lines when both variables are subject to error. *J. Am. Statist. Ass.*, **54**, 173–205.

Moon, M.-S. and Gunst, R. F. (1994) Estimation of the polynomial errors-in-variables model with decreasing error variances. *J. Kor. Statist. Soc.*, **23**, 115–134.

——— (1995) Polynomial measurement error modeling. *Comput. Statist. Data Anal.*, **19**, 1–21.

Nakamura, T. (1990) Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrika*, **77**, 127–137.

O'Neill, M., Sinclair, L. G. and Smith, F. J. (1969) Polynomial curve fitting when abscissa and ordinate are both subject to error. *Comput. J.*, **12**, 52–56.

Schneeweiss, H. and Mittag, H. J. (1986) *Lineare Modelle mit Fehlerbehafteten Daten*. Heidelberg: Physica.

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

Stulajter, F. (1978) Nonlinear estimators of polynomials in mean values of a Gaussian stochastic process. *Kybernetika*, **14**, 206–220.

Van Montfort, K. (1988) *Estimating in Structural Models with Non-normal Distributed Variables: Some Alternative Approaches*. Leiden: DSWO.

Wolter, K. H. and Fuller, W. A. (1982) Estimation of the quadratic errors-in-variables model. *Biometrika*, **69**, 175–182.