Cheng, Schneeweiss:

On the Polynomial Measurement Error Model

Sonderforschungsbereich 386, Paper 259 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

On the Polynomial Measurement Error Model

Chi-Lun Cheng
Institute of Statistical Science, Academia Sinica
Taipei, Taiwan, R.O.C
email: `clcheng@stat.sinica.edu.tw`

Hans Schneeweiss
Department of Statistics, University of Munich
Munich, Germany
email: `schneew@stat.uni-muenchen.de`

Abstract

This paper discusses point estimation of the coefficients of polynomial measurement error (errors-in-variables) models. This includes functional and structural models. The connection between these models and total least squares (TLS) is also examined. A compendium of existing as well as new results is presented.

Keywords: Polynomial errors-in-variables, total least squares, structural relationship, functional relationship, adjusted least squares, structural least squares

1 Introduction: The model

Perhaps the most widely used models in applied statistics are regression models, linear or nonlinear. Measurement error (errors-in-variables) models are important alternatives to ordinary regression models because measurement error (ME) models assume both response and explanatory variables are measured with error, while ordinary regression models assume only the response variable to be measured with error. In many situations, ME models are more realistic than ordinary regression models.

The linear ME model has found extensive treatment in the literature, see, for example, the early review papers by Madansky (1959), Moran (1971), Kendall and Stuart (1979), Chapter 29, and, more recently, Cheng and Van Ness (1994); see also the monographs by Schneeweiss and Mittag (1986), Fuller (1987), and Cheng and Van Ness (1999).

Since the mid-80's, non-linear measurement error models have been extensively studied. Most results are summarized in Carroll et al (1995). The present article reviews the polynomial ME model, which is the most natural extension of the linear model towards a nonlinear model.

The model we investigate can be described as follows. Assume that there are unobservable 'true' variables (ξ_i, η_i) that satisfy a polynomial relation,

$$\eta_i = \beta_0 + \beta_1 \xi_i + \cdots + \beta_k \xi_i^k, \qquad i = 1, \ldots, n, \tag{1}$$

for $k \geq 1$. One observes (x_i, y_i), which are the true variables plus additive errors $(\delta_i, \varepsilon_i)$; that is

$$x_i = \xi_i + \delta_i, \qquad y_i = \eta_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $(\delta_i, \varepsilon_i)$ are independent identically distributed with mean zero and covariance matrix

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

If the ξ_i's are unknown constants, then the model is known as a *functional relationship*, while if the ξ_i's are independent identically distributed random variables and independent of the δ_i's and ε_i's, then the model is known as a *structural relationship*. The "structural parameters" are the intercept β_0, the slope coefficients $\beta_1, \ldots, \beta_k$, the error variances and covariance, σ_δ^2, σ_ε^2, and $\sigma_{\delta\varepsilon}$, and (in the structural model only) the parameters of the distribution of the ξ_i. In the functional model, the unknown constants ξ_i are known as "incidental parameters".

In addition to the measurement errors δ and ε, we may also have an "error in the equation", q say, which is to be added to the right hand side of (1). Although this error can be absorbed into the measurement error ε of the response variable, it is often worthwhile to distinguish between the cases with and without errors in the equation, see Section 3.1, see also Carroll et al (1995, Section 2.3.2) for further comments.

The focus of this paper is the estimation of the structural parameters, especially the intercept and the slope coefficients, which are the parameters of major interest. It is well known that ordinary least squares (OLS) will produce inconsistent estimates, e.g., Griliches and Ringstad (1970). We are interested in constructing consistent estimators.

2 Identifiability and estimation without prior knowledge

We discuss the problem of identifiability only in the context of the structural relationship. For the functional relationship, identifiability is always satisfied, in a sense, but is of no use, as it does not guarantee the existence of consistent estimates, see Cheng and Van Ness, 1999, p. 162 and Appendix A.

We assume that $(\delta_i, \varepsilon_i)$ is jointly normal and, for simplicity, that $\sigma_{\delta\varepsilon} = 0$ unless specified otherwise.

Recall that in the simple linear ME model, Reiersøl (1950) showed that the model is not identifiable if and only if ξ_i is normally distributed, see Bekker (1986) for an extension of this result to the multivariate case. Therefore, we need extra information in order to find estimators that are consistent also when ξ is normal. The most commonly used additional assumptions are: $\sigma_\varepsilon^2/\sigma_\delta^2 = \lambda$ is known or σ_δ^2 is known.

As for the general nonlinear ME model, that is,

$$\eta_i = g(\xi_i, \boldsymbol{\beta}), \qquad i = 1, \ldots, n, \tag{3}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T$ is the unknown parameter vector and g is a real-valued known function, there is no identifiability characterization corresponding to that in the linear ME model. However, in the polynomial ME model, Kendall and Stuart (1979, pp. 435-437) use a method of cumulants to find a consistent estimator in the structural model without additional assumptions. Their results imply that the structural polynomial model with $k > 1$ is identifiable if the errors $(\delta_i, \varepsilon_i)$ are normal.

Van Montfort (1989, Chapter 2) uses higher moments instead of cumulants to estimate the parameters of a structural polynomial regression, see also Cheng and Van Ness (1999), Chapter 6. Recently, Huang and Huwang (2001) constructed a consistent estimator by regressing both y and y^2 on the powers of x when ξ is normal. In the latter regression the parameters depend on those of the first regression. See also Huwang and Huang (2000) for the corresponding Berkson model.

If one wishes to use ML estimation methods, the immediate problem encountered is that one needs to find the joint distribution of (x_i, y_i). This is not as straightforward as in the linear case. To the authors' best knowledge, there are no articles that deal with ML estimation in the polynomial ME model.

Although the above estimators are consistent, their efficiency may be rather low because they do not use knowledge of the error variances. In many situations such knowledge may well be available and can be used to construct more efficient estimators. We now turn to such estimators.

3 Functional methods

We start with the functional variant of the ME model, i.e., the ξ_i's are taken to be fixed unknown quantities. The estimation methods developed on this basis can, however, also be used in the structural case, in particular, if no specific distribution has been assumed for the random ξ_i's. We consider two major cases: when σ_δ^2 is known and when both error variances σ_δ^2 and σ_ε^2 (or their ratio) are known.

3.1 Regressor error variance known

If there is an error in the equation, q, it is very unlikely that its variance would be known *a priori*. On the other hand, the variance of the measurement error, δ, might well be known or at least estimable either from replicated data or from validation data. In this subsection we assume σ_δ^2 to be known (and possibly also higher moments of δ). In case $\sigma_{\delta\varepsilon} \neq 0$, this parameter should also be known.

If only σ_δ^2 is known, maximum likelihood estimation breaks down and does not yield sensible estimators, just as it happens in the linear case (Moberg and Sundbert, 1978). The problem with maximum likelihood estimation in the functional relationship is due to the presence of the incidental parameters ξ_i, whose number increases with sample size (Neyman and Scott, 1948). It is well known that in the presence of incidental parameters, the maximum likelihood estimator need not exist. Even if it exists, it need not be consistent.

Nevertheless consistent estimation of the model parameters is possible. One approach is based on a variant of the corrected score function approach, see Stefanski (1989), Nakamura (1990), and Buonaccorsi (1996). The following discussion is adapted from Cheng and Schneeweiss (1998a), see also Cheng and Van Ness (1999), Chapter 6.

If we provisionally assume that the ξ_i's are known, then the model will return to the ordinary polynomial regression model, and ordinary least squares (OLS) induces the following unbiased estimating equation

$$\sum_{i=1}^{n}(y_i\xi_i^j - \beta_0\xi_i^j - \beta_1\xi_i^{j+1} - \cdots - \beta_k\xi_i^{j+k}) = 0, \tag{4}$$

for $j = 0, \ldots, k$. The problem with (4) is that the terms ξ_i^r, are not observable. If $\sigma_{\delta\varepsilon} = 0$ and if we can find unbiased estimates t_{ri} of ξ_i^r which do not involve any unknown parameters, then we can replace ξ_i^r by t_{ri} in (4) and solve the system for the estimators of $\beta_0, \beta_1, \ldots, \beta_k$.

In the following discussion, we omit the observation index i for ease of notation. Because $x = \xi + \delta$, we have, by the binomial theorem,

$$x^r = \sum_{j=0}^{r}\binom{r}{j}\xi^j\delta^{r-j}. \tag{5}$$

After replacing ξ^j in (5) by t_j and δ^{r-j} by its expectation, we can solve for t_j. The first five terms of t_j, for example, are

$$t_0 = 1, \quad t_1 = x, \quad t_2 = x^2 - \sigma_\delta^2, \quad t_3 = x^3 - 3x\sigma_\delta^2 - \mathbf{E}\delta^3,$$
$$t_4 = x^4 - 6x^2\sigma_\delta^2 - 4x\mathbf{E}\delta^3 - \mathbf{E}\delta^4 + 6\sigma_\delta^4,$$

Under the normality assumption of the error δ, we only need to know σ_δ^2 in order to find t_j. In this case we also have a recursion formula for obtaining t_j:

$$t_{j+1} = xt_j - \sigma_\delta^2 j t_{j-1},$$

with $t_0 = t_{-1} = 1$, see also Stefanski (1989). Once the β's have been estimated, the unknown error variance σ_ε^2 can also be estimated consistently. If $\sigma_{\delta\varepsilon} \neq 0$, the term $y\xi^j$ has to be replaced in (4) by an expression h_j involving $\sigma_{\delta\varepsilon}$ such that $Eh_j = \eta\xi^j$. For $\sigma_{\delta\varepsilon} = 0$, $h_j = yt_j$ and for normal (δ, ε), $h_j = yt_j - \sigma_{\delta\varepsilon} j t_{j-1}$.

To sum up, let us introduce the matrices $\boldsymbol{H}_i$ with elements $(\boldsymbol{H}_i)_{jl} = t_{j+l,i}$, $j, l = 0, \ldots, k, i = 1, \ldots, n$, so that $\mathbf{E}\boldsymbol{H}_i = \left(\xi_i^{j+l}\right)_{j,l=0,\ldots,k}$, and the vectors $\boldsymbol{h}_i$, with $\mathbf{E}\boldsymbol{h}_i = \left(\eta_i \xi_i^j\right)_{j=0,\ldots,k}$. The "corrected" or "adjusted" estimating equation that replaces (4) can then be written as

$$\sum_{i=1}^n \boldsymbol{H}_i \hat{\boldsymbol{\beta}}_{ALS} = \sum_{i=1}^n \boldsymbol{h}_i.$$

Its solution, $\hat{\boldsymbol{\beta}}_{ALS}$, is the so-called *adjusted least squares* (ALS) estimate of $\boldsymbol{\beta}$. Chan and Mak (1985) proposed a similar approach for the $\boldsymbol{\Omega}$ known case.

Under some regularity conditions (Huber, 1967), the resulting estimators are consistent and asymptotically normal. See also Carroll et al (1995, p 261 ff.), Fazekas et al (1999) and Baran (2000). The asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{ALS}$ is given by the formula

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_{ALS}} = \frac{1}{n} \bar{\boldsymbol{H}}^{-1} \bar{\boldsymbol{U}} \bar{\boldsymbol{H}}^{-1} \tag{6}$$

with $\bar{\boldsymbol{H}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{H}_i$, $\bar{\boldsymbol{U}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{H}_i\hat{\boldsymbol{\beta}}_{ALS} - \boldsymbol{h}_i)(\boldsymbol{H}_i\hat{\boldsymbol{\beta}}_{ALS} - \boldsymbol{h}_i)^T$. Because of its form, (6) is often called a sandwich formula, see Carroll et al (1995).

For small samples, $\hat{\boldsymbol{\beta}}_{ALS}$ turns out to be rather unstable in particular if σ_δ^2 is large and if $k \geq 3$. In fact, the moments of $\hat{\boldsymbol{\beta}}_{ALS}$ do not exist, see Cheng and Van Ness, 1999, Chapter 2, for the linear case. Using an idea of Fuller's (1997), the estimator can be improved for small samples by the following modification, see Cheng et al (2000). We use the concepts and notations of Subsection 3.2, see in particular (9). Then the modified ALS estimator of β is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MALS} &= (\boldsymbol{M}_{tt} - a\,\bar{\hat{\boldsymbol{V}}})^{-1}(\boldsymbol{m}_{ty} - a\bar{\hat{\boldsymbol{v}}}), \\ a &= \begin{cases} (n-\alpha)/n & \text{if} \quad \hat{\rho} > 1 + \frac{1}{n} \\ \hat{\rho}(n-\alpha)/(n+1) & \text{if} \quad \hat{\rho} \leq 1 + \frac{1}{n}, \end{cases} \end{aligned}$$

where $\hat{\rho}$ is the smallest positive root of $\det(\boldsymbol{M} - \rho\,\tilde{\boldsymbol{W}}) = 0$ with

$$\tilde{\boldsymbol{W}} = \begin{pmatrix} \sigma_{\delta\varepsilon}^2/\sigma_\delta^2 & \bar{\hat{\boldsymbol{v}}}^T \\ \bar{\hat{\boldsymbol{v}}} & \bar{\hat{\boldsymbol{V}}} \end{pmatrix}$$

and some α to be chosen so that an estimator of good small sample properties results, e.g., $\alpha = k + 5$.

A multivariate extension of the ALS procedure including a small sample modification has recently been provided by Wall and Amemiya (2000). Actually the authors study a factor analysis (or "structural") model where the common factors obey a polynomial relation instead of the usual linear relation.

3.2 Ratio of error variances known

Consider now model (1) and (2), i.e., a polynomial without an error in the equation but with measurement errors in both variables. We can assume that in such a model $\mathbf{\Omega}$ is known.

Let us first consider maximum likelihood (ML), assuming the errors are normally distributed. For the general nonlinear model (3), with $\sigma_{\delta\varepsilon} = 0$, ML is equivalent to minimizing the function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \min_{\xi_i \in R} [(y_i - g(\xi_i, \boldsymbol{\beta}))^2 + \lambda(x_i - \xi_i)^2]$$

with respect to $\boldsymbol{\beta}$, where $\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$. This method is called total least squares (TLS). Geometrically this means that in case $\lambda = 1$ one has to minimize the sum of squares of the perpendicular distances from the data points to the regression curve. TLS was first proposed by Adcock (1877) and then re-discovered by many authors since. Golub and Van Loan (1980) proposed numerical methods for TLS in a general linear framework. The more recent work is summarized in Van Huffel and Vandewalle (1991). In the linear case, minimizing $Q(\boldsymbol{\beta})$ reduces to minimizing the function

$$q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma_e^2} \tag{7}$$

with $e_i = \varepsilon_i - \beta_1 \delta_i$. This procedure, which corresponds to the algebraic side of TLS, is also called generalized or weighted least squares (WLS) and was introduced by Sprent (1966).

Kendall and Stuart (1979, p. 434) applied ML to the quadratic ($k = 2$)

measurement error model. They did not find a closed form solution but had to resort to iterative methods just as in the general TLS approach. For numerical procedures in the polynomial ME context see O'Neill et al (1969), see also Britt and Luecke (1973).

In the linear model, the ML estimates of intercept and slope parameters are consistent. This is no more true for the nonlinear model, see Fazekas et al (1998).

Despite the inconsistency of the TLS estimate of β, it might still be useful to use this method in practice, especially when σ_δ^2 is small. In such cases the bias, though still existent, might be almost negligible. In fact, there exists a small-σ_δ theory for TLS, see Wolter and Fuller (1982b).

However, consistent estimation is possible. One can extend the WLS approach to the polynomial ME model. The idea comes from Wolter and Fuller (1982a), which they applied to the quadratic functional model with normal errors and known covariance matrix. Cheng and Schneeweiss (1998b) extended it to a polynomial of any degree, and the restriction of normally distributed errors was lifted. To be more specific, model (1)-(2) can be re-written as

$$y_i = \boldsymbol{\zeta}_i^T \boldsymbol{\beta} + \varepsilon_i, \qquad \boldsymbol{t}_i = \boldsymbol{\zeta}_i + \boldsymbol{\pi}_i, \tag{8}$$

where $\boldsymbol{\zeta}_i = (\xi_i^0, \ldots, \xi_i^k)^T$, $\boldsymbol{t}_i = (t_{0i}, \ldots, t_{ki})^T$; and $\boldsymbol{\pi}_i = (\pi_{0i}, \ldots, \pi_{ki})^T$ is implicitly defined by (8). Note that the new measurement errors π_{ri} have mean zero, covariance matrix $\boldsymbol{V}_i = \mathbf{E}(\boldsymbol{\pi}_i \boldsymbol{\pi}_i^T)$, and covariance vector $\boldsymbol{v}_i =$

$\mathbf{E}(\boldsymbol{\pi}_i \varepsilon_i)$. Cheng and Schneeweiss (1998a, 1998b) gave unbiased estimates of $\boldsymbol{V}_i$ and $\boldsymbol{v}_i$, which were

$$\hat{\boldsymbol{V}}_i = \boldsymbol{t}_i \boldsymbol{t}_i^T - \boldsymbol{H}_i, \qquad \hat{\boldsymbol{v}}_i = \boldsymbol{t}_i y_i - \boldsymbol{h}_i.$$

Let $\boldsymbol{z}_i = (y_i, \boldsymbol{t}_i^T)^T$ and let $\boldsymbol{W}_i$ be the error covariance matrix of $\boldsymbol{z}_i$ and

$$\hat{\boldsymbol{W}}_i = \begin{pmatrix} \sigma_\varepsilon^2 & \hat{\boldsymbol{v}}_i{}^T \\ \hat{\boldsymbol{v}}_i & \hat{\boldsymbol{V}}_i \end{pmatrix}$$

its estimate. Moreover, let $\boldsymbol{M} = n^{-1} \Sigma \boldsymbol{z}_i \boldsymbol{z}_i^T$, $\boldsymbol{m}_{ty} = n^{-1} \Sigma \boldsymbol{t}_i y_i$, $\boldsymbol{M}_{tt} = n^{-1} \Sigma \boldsymbol{t}_i \boldsymbol{t}_i^T$ and $m_{yy} = n^{-1} \Sigma y_i^2$, then the estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{M}_{tt} - \hat{\kappa} \bar{\hat{\boldsymbol{V}}})^{-1} (\boldsymbol{m}_{ty} - \hat{\kappa} \bar{\hat{\boldsymbol{v}}}), \tag{9}$$

where $\hat{\kappa}$ is the smallest positive root of

$$\det(\boldsymbol{M} - \kappa \bar{\hat{\boldsymbol{W}}}) = 0;$$

see also Moon and Gunst (1995).

As pointed out by Wolter and Fuller (1982a), the estimate (9) is the $\boldsymbol{\beta}$ that minimizes the function $w(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{M} \boldsymbol{\theta} / (\boldsymbol{\theta}^T \bar{\hat{\boldsymbol{W}}} \boldsymbol{\theta})^{-1}$, where $\boldsymbol{\theta}^T = (1, -\boldsymbol{\beta}^T)$. In other words, the estimate (9) can be viewed as a WLS estimate of the ("linearized") polynomial ME model (8). Note that if $k = 1$, then $w(\boldsymbol{\theta})$ reduces to (7).

The estimate (9) is consistent and asymptotically normal and its asymptotic covariance matrix is given in Cheng and Schneeweiss (1998b).

When the errors $(\delta_i, \varepsilon_i)$ are normal, then it is possible to relax the assumption of $\boldsymbol{\Omega}$ being known to knowing $\boldsymbol{\Omega}$ up to a proportionality factor. One can assume that, without loss of generality, this unknown factor is σ_ε^2 and

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} a & b \\ b & 1 \end{pmatrix},$$

where a and b are known constants. Note that the unknown factor σ_ε^2 can be estimated by

$$\hat{\sigma}_\varepsilon^2 = \hat{\boldsymbol{\theta}}^T \begin{pmatrix} m_{yy} & \boldsymbol{m}_{ty}^T \\ \boldsymbol{m}_{ty} & \bar{\boldsymbol{H}} \end{pmatrix} \hat{\boldsymbol{\theta}}.$$

Therefore, we need to estimate $\boldsymbol{\beta}$ and σ_ε^2 simultaneously. The exact numerical procedure has not been studied yet. It should be noted that we need the normality assumption of $(\delta_i, \varepsilon_i)$ because then all moments of $(\delta_i, \varepsilon_i)$ will be known up to some powers of σ_ε^2 once a and b are known.

4 Structural methods

In this section, we study the structural variant of the polynomial ME model. We assume that the ξ_i are iid $\mathrm{N}(\mu_\xi, \sigma_\xi^2)$. The errors are also assumed to be normal with $\sigma_{\delta\varepsilon} = 0$. σ_δ^2 is supposed to be known. $\mu_\xi (= \mu_x)$ and $\sigma_\xi^2 (= \sigma_x^2 - \sigma_\delta^2)$ can be estimated from the x_i alone without having resort to the model.

The method we present here is based on a quasi score function, which is derived from a conditional mean-variance model corresponding to the polynomial model. The following presentation is essentially due to Thamerus (1998), see also Carroll et al (1995).

Note that the error-free model (1) can be written as a mean-variance model

$$\begin{aligned} \mathbf{E}(y_i \mid \xi_i) = \beta_0 + \beta_1\xi_i + \cdots + \beta_k\xi_i^k \\ V(y_i \mid \xi_i) = \sigma_\varepsilon^2. \end{aligned} \tag{10}$$

Now, the conditional distribution of ξ_i given x_i is also normal.

$$\begin{aligned} \xi_i \mid x_i &\sim \mathrm{N}(\mu_1(x_i), \tau^2) \\ \mu_1(x) &= \mu_x + (1 - \sigma_\delta^2/\sigma_x^2)(x - \mu_x) \\ \tau^2 &= \sigma_\delta^2(1 - \sigma_\delta^2/\sigma_x^2). \end{aligned}$$

Therefore we can easily derive from (10) a conditional mean-variance model of y given x:

$$\begin{aligned} \mathbf{E}(y_i \mid x_i) = \sum_{j=0}^{k} \beta_j \mu_j(x_i) =: m(x_i, \boldsymbol{\beta}) \\ V(y_i \mid x_i) = \sigma_\varepsilon^2 + \sum_{j,l=0}^{k} \beta_j\beta_l\{\mu_{j+l}(x_i) - \mu_j(x_i)\mu_l(x_i)\} =: v(x_i, \boldsymbol{\beta}, \sigma_\varepsilon^2) \end{aligned} \tag{11}$$

with

$$\mu_r(x) = \mathbf{E}(\xi^r \mid x) = \sum_{j=0}^{r} \binom{r}{j} \mu_j^* \mu_1(x)^{r-j}$$

$$\mu_j^* = \begin{cases} 0 & \text{if } \ j \text{ is odd} \\ 1 \cdot 3 \cdots (j-1)\tau^j & \text{if } \ j \text{ is even.} \end{cases}$$

The following estimating equations can then be set up:

$$\begin{aligned}
\sum_{i=1}^{n}[\{y_i - m(x_i,\boldsymbol{\beta})\}/v(x_i,\boldsymbol{\beta},\sigma_\varepsilon^2)]\boldsymbol{\mu}(x_i) = 0 \\
\sigma_\varepsilon^2 = \tfrac{1}{n}\sum_{i=1}^{n}[\{y_i - m(x_i,\boldsymbol{\beta})\}^2 - \sum_{j,l}\beta_j\beta_l\{\mu_{j+l}(x_i) - \mu_j(x_i)\mu_l(x_i)\}]
\end{aligned} \tag{12}$$

where $\boldsymbol{\mu}(x) = (\mu_0(x),\ldots,\mu_k(x))^T$. They can be solved by iteratively reweighted least squares and give rise to the so-called *structural least squares* (SLS) estimators of $\boldsymbol{\beta}$ and σ_ε^2. It should be noted that μ_x and σ_x^2 have to be replaced by their estimates in the estimating equations (12).

The estimator $\hat{\boldsymbol{\beta}}_{SLS}$ is consistent and asymptotically normal with asymptotic covariance matrix

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_{SLS}} = \tfrac{1}{n}[\mathbf{E}\{\boldsymbol{\mu}(x)\boldsymbol{\mu}^T(x)/v(x,\boldsymbol{\beta},\sigma_\varepsilon^2)\}]^{-1} + O(\sigma_\delta^4). \tag{13}$$

The term $O(\sigma_\delta^4)$ is due to the estimation of μ_x and σ_x^2 and can be neglected in a first approximation, see Kukush and Schneeweiss (2000).

The SLS method can be generalized to encompass models where the ξ_i follow a finite mixture of normal distributions and where the measurement errors could be heteroscedastic, Thamerus (1998).

A very simple approximate method to SLS is the regression calibration (RCAL) method, Carroll et al (1995). In this method, the ξ_i in equation (1) are simply replaced by their conditional means $\mu_1(x_i) = \mathbf{E}(\xi_i | x_i)$ and the resulting polynomial regression is estimated by OLS. The RCAL estimates are not consistent but their biases are typically very small.

5 Comparison of functional and structural methods

Since functional methods can also be used in a structural model, we can compare functional and structural methods within the context of a structural polynomial ME model. In particular Kukush et al (2001) compare ALS and SLS in a structural model with normally distributed ξ_i. It might be expected that SLS is more efficient than ALS in this case, as it uses the information on the distribution of the ξ_i. However, it turns out that the asymptotic covariance matrices of ALS and SLS, see (6) and (13), are equal up to the order of σ_δ^2.

Comparisons with respect to OLS and to RCAL can also be carried out, see also Moon and Gunst (1995).

When the ξ_i are not normally distributed, SLS breaks down, i.e., it leads to inconsistent estimates. Thus while SLS might be more efficient than ALS when the ξ_i are normal - but only in terms of order $O(\sigma_\delta^4)$ -, the latter procedure is more robust with respect to the distribution of the ξ_i, see Kuha and Temple (1999), Schneeweiss and Nittner (2000).

6 Miscellaneous topics

We only studied point estimation. Another important issue is interval estimation. However, little is known in the polynomial ME model. In principle, one can use the asymptotic normality of the estimates of the β's to obtain (approximate) confidence regions on the parameters. But there is a serious problem regarding this approach due to the zero confidence level effect (Gleser and Hwang, 1987). For details, see Cheng and Van Ness (1999), Section 2.4.

The models we studied so far are all characterized by having additive measurement errors. Iturria et al (1999) investigated a polynomial ME model with multiplicative measurement errors, that is, $x_i = \xi_i \delta_i$.

When we surveyed the structural polynomial ME model in Section 4, we restricted our attention to the case of normally distributed ξ_i. It is known that in the linear model a nonnormal distribution of ξ will produce a nonlinear regression function in the observables y and x, i.e., $\mathbf{E}(y|\, x)$ will be nonlinear, see Lindley (1947). Chesher (1998) extends this result to the polynomial regression and gives explicit formulas for $\mathbf{E}(y|\, x)$ for any distribution of x and normal measurement errors. These relations might be exploited to find consistent estimates for the β_j's as long as they are identifiable from $\mathbf{E}(y|\, x)$ even if σ_δ^2 is not known.

In this paper we only studied polynomial ME models where either no further knowledge was available or where some knowledge about the error variances could be used. A further possibility to consistently estimate the β's is the use of instrumental variables and of indicator variables. The usual instrumental variables approach for the linear model breaks down in nonlinear models, see Amemiya (1985). Nevertheless consistent estimation using higher moments is possible, see Hausman and Newey (1991) and Hausman et al (1995) with an application to the estimation of Engel curves.

References

Adcock, R.J. (1877), *Note on the method of least sqares*, Analyst 4, 183-184.

Amemiya, Y. (1985), *Instrumental variable estimator for the nonlinear errors-in-variables model*, Journ. of Econometrics 28, 273-289.

Baran, S. (2000), *A consistent estimator in general functional errors in variables models*, Metrika 51, 117-132.

Bekker, P.A. (1986), *Comment on identification in the linear errors in variables model*, Econometrica 54, 215-217.

Britt, H.I. and Luecke, R.H. (1973), *The estimation of parameters in nonlinear implicit models*, Technometrics 15, 233-247.

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*, Chapman and Hall, London.

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. (1998a), *Polynomial regression with errors in the variables*, J.R. Statist. Soc. B 60, 189-199.

Cheng, C.-L. and Schneeweiss, H. (1998b), *Note on two estimators for the polynomial regression with errors in the variables*, In: *Statistical Modeling: Procedings of the 13th International Workshop on Statistical Modeling*, 141-147, B. Marx and H. Friedl eds. New Orleans, Louisiana.

Cheng, C.-L., Schneeweiss, H. and Thamerus, M. (2000), *A small sample estimator for a polynomial regression with errors in the variables*, J.R. Statist. Soc. B 62, 699-709.

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.

Cheng, C.-L. and Van Ness, J.W. (1999), *Statistical Regression with Measurement Error*, London: Arnold

Chesher, A. (1998), *Polynomial regression with covariate measurement error*, Discusion Paper No.98/448, University of Bristol, Department of Economics.

Fazekas, I., Baran, S., Kukush, A. and Lauridsen, J. (1999), *Asymptotic properties in space and time of an estimator in nonlinear functional errors-in-variables models,* Random Oper. Stoch. Equ. 7, 389-412.

Fazekas, I., Kukush, A. and Zwanzig, S. (1998), *On inconsistency of the least squares estimator in nonlinear functional relations*, Research Report N1, Odense University, Denmark, Department of Statistic and Demography.

Fuller, W.A. (1987), *Measurment Error Models*, New York: Wiley

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

Golub, G.H. and Van Loan, C.F. (1980), *An analysis of total least squares problem*, SIAM J. Numer. Anal. 17, 883-893.

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

Hausman, J.A. and Newey, W.K. (1991), *Identification and estimation of polynomical errors-in-variables models*, J. of Econometrics 50, 273-295.

Hausman, J.A., Newey, W.K. and Powell, J.L. (1995), *Nonlinear errors in variables estimation of some Engel curves*, J. of Econometrics 65, 205-233.

Huang, Y.H. and Huwang, L.C. (2001) *On the polynomial structural relationship*, Canadian J. Statist., to appear.

Huber, P.J. (1967), *The behavior of maximum likelihood estimates under nonstandard condition*, Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 221-233.

Huwang, L.C. amd Huang, Y.H. (2000), *On errors-in-variables in polynomial regression-Berkson case*, Statistica Sinica 10, 923-936.

Iturria, S.J., Carroll, R.J. and Firth, D. (1999), *Polynomial regression and estimating functions in the presence of multiplicative measurement error*, J. R. Statist. Soc. B 61, 547-561.

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

Kuha, J.T. and Temple, J. (1999), *Covariate measurement error in quadratic regression*, Discussion Paper 1999-W2, Nuffield College Oxford.

Kukush, A. and Schneeweiss, H. (2000), *A Comparison of asymptotic covariance matrices of Adjusted Least Squares and Structural Least Squares in error ridden polynomial regression*, Discussion Paper 218, Sonderforschungsbereich 386, University of Munich.

Kukush, A., Schneeweiss, H. and Wolf, R. (2001), *Comparison of three estimators in a polynomial regression with measurement errors*, Discussion Paper 233, Sonderforschungsbereich 386, University of Munich.

Lindley, D.V. (1947), *Regression lines and the linear functional relationship*, J.R. Statist. Soc. Suppl. 9, 218-244.

Madansky, A. (1959), *The fitting of straight lines when both variables are subject to errors*, J. Amer. Statist. Assoc. 54, 173-205.

Moberg, L. and Sundberg, R. (1978), *Maximum likelihood estimation of a linear functional relationship when one of the departure variances is known*, Scand. J. Statist. 5, 61-64.

Moon, M.S. and Gunst, R.F. (1995). *Polynomial measurement error modeling*, Comput. Statist. Data Anal. 19, 1-21.

Moran, P.A.P. (1971), *Estimating structural and functional relationships*, J. Multivariate Anal. 1, 232-255.

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

Neyman, J. and Scott, E.L. (1948). *Consistent estimates based on partial consistent observations*, Econometrica 16, 1-32.

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

Reiersøl, O. (1950), *Identifiability of a linear relation between variables which are subject to errors*, Econometrica 18, 375-389.

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

Schneeweiss, H. and Nittner, T. (2000), *Estimating a polynomial regression with measurement errors in the structural and in the functional case - a comparison*, Discussion Paper 197, Sonderforschungsbereich 386, University of Munich.

Sprent, P. (1966), *A generalized least squares approach to linear functional relationships.* J. R. Statist. Soc. B 28, 278-297.

Stefanski, L.A. (1989), *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models*, Communications in Statistics, Part A - Theory and Methods 18, 4335-4358.

Thamerus, M. (1998), *Different nonlinear regression models with incorrectly observed covariates.* In R. Galata and H. Kuechenhoff (Eds.): Econometrics in Theory and Practice, Festschrift for Hans Schneeweiss. Physica, Heidelberg, New York.

Van Huffel, S. and Vandewalle, J. (1991), *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia.

Van Montfort, K. (1989), *Estimating in Structural Models with Nonnormal Distributed Variables: Some Alternative Approaches*, Leiden: DSWO

Wall, M.M. and Amemiya, Y. (2000), *Estimation for polynomial structural equation models*, J. Amer. Statist. Assoc. 95, 929-940.

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

Wolter, K.H. and Fuller, W.A. (1982b), *Estimation of nonlinear error-in-variables models*, The Annals of Statistics 10, 539-548.