Schneeweiss, Cheng, Wolf:

On the bias of structural estimation methods in a polynomial regression with measurement error when the distribution of the latent covariate is a mixture of normals

Sonderforschungsbereich 386, Paper 281 (2002)

On the bias of structural estimation methods in a polynomial regression with measurement error when the distribution of the latent covariate is a mixture of normals

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

Chi-Lun Cheng
Institute of Statistics, Academia Sinica, Taipeh, Republic of China
email: clcheng@stat.sinica.edu.tw

Roland Wolf
Department of Statistics, University of Munich, Germany
email: wolf@stat.uni-muenchen.de

Abstract

The structural variant of a regression model with measurement error is characterized by the assumption of an underlying known distribution of the latent covariate. Several estimation methods, like regression calibration or structural quasi score estimation, take this distribution into account. In the case of a polynomial regression, which is studied here, structural quasi score takes the form of structural least squares (SLS). Usually the underlying latent distribution is assumed to be the normal distribution because then the estimation methods take a particularly simple form. SLS is consistent as long as this assumption is true. The purpose of the paper is to investigate the amount of bias that results from violations of the normality assumption for the covariate distribution. Deviations from normality are introduced by switching to a mixture of normal distributions. It turns out that the bias reacts only mildly to slight deviations from normality.

Keywords: Structural Quasi Score Estimation, Polynomial Regression, Misspecification, Bias, Mixture of Normal Distributions, Measurement Errors.

1 Introduction

It is well-known that errors in the measurement of the exogenous variables of a regression produce inconsistent estimates of the regression parameters when one of the conventional estimation methods, like least squares or maximum likelihood, is applied to the model without taking the measurement errors into account. On the other hand, measurement errors prevail in most empirical data, even if in many cases they may be small enough to be safely ignored.

A classical example in economics where such errors cannot be ignored is Friedman's permanent income hypothesis in consumption theory. Another more recent example is the notoriously ill-measured schooling variable in a wage determination equation in labor economics, Card (1999). Many more examples can be found in economics as well as in other fields, notably in epidemiology.

To deal with such error-ridden regression models, one typically needs to know the error process. Here it is assumed that the error δ is additive and is normally distributed with mean zero (i.e., with no systematic error) and with a known variance σ_δ^2. Knowledge of the error variance can come from replicated measurements or from validation data.

While there is an extensive body of literature for the linear regression model with measurement error, where standard methods have been developed - see Schneeweiss and Mittag (1986), Fuller(1987), Cheng and Van Ness (1999), Wansbeek and Meijer (2000) - the situation is not so clear for nonlinear models. One finds a number of competing methods for nonlinear models, the properties of which are not completely understood and may depend on the kind of nonlinearity, see Carroll et al (1995). Here we focus on a polynomial regression in one variable, although our results may apply to a wider class of nonlinear models. A recent example for the use of a quadratic model to study a nonlinear phenomenon in economics can be found in Kuha and Temple (1999). A survey of methods for the polynomial model is given in Cheng and Schneeweiss (2002), see also Cheng and Van Ness (1999).

Among the various estimation methods we want to study a structural method which has been termed Structural Least Squares (SLS) in the context of a polynomial regression, Kukush et al (2001), and also Structural Quasi Score (SQS) in a wider context of nonlinear models, Kukush et al (2002). It is a structural method in so far as it is based on the knowledge of the distribution of the error free (latent) regressor variable ξ, which here is assumed to follow a normal distribution, see also Armstrong (1985) and Carroll et al. (1995).

There are other methods, like Adjusted Least Squares (ALS) in Cheng and Schneeweiss (1998), which do not depend on such knowledge and are called functional methods.

Structural methods may be more efficient than functional ones, as they utilize more information given in the model. On the other hand, structural methods become inconsistent when the distributional assumptions for the latent variable ξ are not met. The question is, how sensitive is the SLS estimator with respect to deviations from the presumed normality of ξ.

Note that in a polynomial model without measurement errors the estimation method (which is OLS) as well as its consistency property do not depend on the regressor distribution. One can therefore surmise that in an error-ridden model the (asymptotic) bias of SLS induced by deviations from normality is almost negligible if the deviation is slight in a sense. Simulation studies by Schneeweiss and Nittner (2001) seem to corroborate this conjecture.

Here we want to study the effect of deviations from the normality of ξ on the bias of the SLS estimator in a more systematic way. We replace the normal distribution of ξ with a mixture of two normals of equal variance but differing means. Denote the difference of the means by ϑ, then the bias of SLS is a function of ϑ. It is zero for $\vartheta = 0$ and increases in absolute value with growing ϑ, at least for small ϑ. If the bias function $\boldsymbol{b}(\vartheta)$ were to increase linearly for small ϑ, we would consider SLS to be sensitive with regard to deviations from normality, if it increases in a quadratic way for small ϑ, it may be considered insensitive. The question of which of the two is true can be settled by computing the derivative of $\boldsymbol{b}(\vartheta)$ at $\vartheta = 0$.

In a similar way we can also study the dependence of the bias on the error variance σ_δ^2. It is zero for $\sigma_\delta^2 = 0$. We compute the derivative of the bias function with respect to σ_δ at $\sigma_\delta = 0$.

It turns out that the SLS method is insensitive in the sense above under both variations.

A Monte Carlo simulation study corroborates these results.

In the next section we introduce the model and the SLS method with the same notation as in Cheng and Schneeweiss (1998). In the third section the bias of SLS under the distributional misspecification described above is evaluated and is further studied for small deviations in Section 4, where our main result is presented and proved. Section 5 contains some simulation results, and Section 6 has some concluding remarks.

The appendix provides some additional proofs, a formula for the second derivative with respect to σ_δ, and a study of the linear case.

2 The model and its estimate

We study the polynomial regression model:

$$y_i = \sum_{j=1}^{k} \beta_i \xi_i^j + \epsilon_i \tag{1}$$

with additive measurement errors δ_i in the observed regressors x_i:

$$x_i = \xi_i + \delta_i, \tag{2}$$

where $(\xi_i, \epsilon_i, \delta_i) \backsim$ i.i.d. $\mathrm{N}((\mu_\xi, 0, 0), \Sigma)$ with $\Sigma = \mathrm{diag}(\sigma_\xi^2, \sigma_\epsilon^2, \sigma_\delta^2), \sigma_\xi^2 > 0, \sigma_\epsilon^2 > 0, i = 1, \dots, n$. The variables $\xi_i, \epsilon_i, \delta_i$ are unobservable, the variables y_i, x_i are observable. The measurement error variance σ_δ^2 is assumed to be known.

The SLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta} := (\beta_0, \beta_1, \dots, \beta_k)$ is defined by way of the following reasoning: We have to reduce model (1) to a corresponding model in the observable variable x instead of the latent variable ξ. We cannot do so by simply replacing ξ with x in equation (1). This would lead to the so-called naive estimator $\hat{\beta}_N$, which is (asymptotically) biased. Instead we construct the conditional mean and variance of y given x, see Kukush et al (2001), see also 7.1:

$$\mathrm{E}(y|x) = \boldsymbol{\mu}(x)'\boldsymbol{\beta} \tag{3}$$

$$\mathrm{V}(y|x) = \sigma_\epsilon^2 + \boldsymbol{\beta}'\{\boldsymbol{M}(x) - \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'\}\boldsymbol{\beta} =: v(x, \boldsymbol{\beta}, \sigma_\epsilon^2), \tag{4}$$

where $\boldsymbol{\mu}(x) = (1, \mu^1(x), \dots, \mu^k(x))$ with

$$\mu^r(x) = \mathrm{E}(\xi^r|x), \tag{5}$$

and $\boldsymbol{M}(x)$ is a matrix with $(\boldsymbol{M}(x))_{r,s} = \mu^{r+s}(x), r, s = 0, \dots, k$.

$\mu^r(x)$ is easily computed using the fact that due to (2) and by the normality assumption

$$\xi|x \sim \mathrm{N}(\mu(x), \tau^2)$$

with

$$\mu(x) = \mu_x + \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right)(x - \mu_x) \tag{6}$$

and

$$\tau^2 = \sigma_\delta^2 \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right). \tag{7}$$

It then follows that

$$\mu^r(x) = \sum_{j=0}^{r} \binom{r}{j} \mu_j^* \mu(x)^{r-j} \tag{8}$$

$$\mu_j^* = \begin{cases} 0 & \text{if } j \text{ is odd} \\ 1 \cdot 3 \cdot 5 \cdots (j-1)\tau^j & \text{if } j \text{ is even, } \mu_0^* = 1. \end{cases} \tag{9}$$

These expressions depend on μ_x and σ_x^2. For simplicity let us assume that these (nuisance) parameters are known. In practice, of course, they have to estimated, which is easily done by computing sample mean and variance of the $x_i, i = 1, \ldots, n$.

Equations (3) and (4) can now be rewritten as a linear regression of y on the $\mu^r(x), r = 0, \ldots, k$:

$$y = \boldsymbol{\mu}(x)'\boldsymbol{\beta} + \varphi$$

with $E(\varphi|x) = 0$ and $V(\varphi|x) = v(x, \boldsymbol{\beta}, \sigma_\epsilon^2)$. Thus the equation error φ is heteroscedastic. Note also that $v(x, \boldsymbol{\beta}, \sigma_\epsilon^2) \geq \sigma_\epsilon^2$, as $\boldsymbol{M} - \boldsymbol{\mu\mu'}$ is positive semi-definite. The model can be estimated by an iteratively reweighted least squares algorithm. The resulting SLS (or SQS) estimator $\hat{\boldsymbol{\beta}}$ is the solution to the equations

$$\sum_{i=1}^{n} \frac{y_i - \boldsymbol{\mu}(x_i)'\hat{\boldsymbol{\beta}}}{v(x_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}_\epsilon^2)} \; \boldsymbol{\mu}(x_i) = 0 \tag{10}$$

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n}[(y - \boldsymbol{\mu}(x)'\hat{\boldsymbol{\beta}})^2 - \hat{\boldsymbol{\beta}}'\{\boldsymbol{M}(x_i) - \boldsymbol{\mu}(x_i)\boldsymbol{\mu}(x_i)'\}\hat{\boldsymbol{\beta}}]. \tag{11}$$

The SLS estimator $\hat{\boldsymbol{\beta}}$ is consistent as long as the normality assumption for ξ is satisfied, Kukush and Schneeweiss (2000). This is essentially due to the fact that the expected value of (10), with the estimates replaced by their true values, is zero. Of course, $\hat{\sigma}_\epsilon^2$ is also a consistent estimate.

3 Bias under nonnormality

Suppose that ξ is *not* normally distributed. Indeed, let the distribution of ξ be a mixture of two normals with equal variance σ^2 but unequal means:

$$\xi \sim p_1 \mathrm{N}(\mu_1, \sigma^2) + p_2 \mathrm{N}(\mu_2, \sigma^2). \tag{12}$$

If we still estimate $\hat{\boldsymbol{\beta}}$ by solving (10) and (11), then the resulting SLS estimator $\hat{\boldsymbol{\beta}}$ will be (asymptotically) biased except for the linear case ($k = 1$); for the latter see 7.3.

The bias is given by

$$\boldsymbol{b} = \mathrm{plim}\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} =: \boldsymbol{\beta}^* - \boldsymbol{\beta},$$

and $\boldsymbol{\beta}^*$ is found as the solution to the equations

$$\mathrm{E}\left[\frac{y - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*}{v(x, \boldsymbol{\beta}^*, \sigma_{\epsilon_*}^2)}\boldsymbol{\mu}(x)\right] = 0 \tag{13}$$

$$\begin{aligned} \sigma_{\epsilon_*}^2 &= \mathrm{E}[\{y - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*\}^2 - \boldsymbol{\beta}^{*'}\{\boldsymbol{M}(x) - \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'\}\boldsymbol{\beta}^*] \\ &=: \sigma_\epsilon^2(\boldsymbol{\beta}^*). \end{aligned} \tag{14}$$

If we substitute $\sigma_{\epsilon_*}^2$ from (14) in the expression for v, the latter becomes a function of x and $\boldsymbol{\beta}^*$, which we denote by

$$v_*(x, \boldsymbol{\beta}^*) := v(x, \boldsymbol{\beta}^*, \sigma_\epsilon^2(\boldsymbol{\beta}^*)).$$

Substituting $v_*(x, \boldsymbol{\beta}^*)$ in (13) we get the equation

$$\mathrm{E}\left[\frac{y - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*}{v(x, \boldsymbol{\beta}^*)}\ \boldsymbol{\mu}(x)\right] = 0, \tag{15}$$

which is to be solved for $\boldsymbol{\beta}^*$, assuming the solution exists and is unique.

The term in square brackets in (15) is called quasi score function or estimating function, see Carroll et al (1995), because it is the basis for estimating $\boldsymbol{\beta}$, see (10). Let us denote it by

$$\boldsymbol{\psi}(y,x,\boldsymbol{\beta}^*) = v_*^{-1}\boldsymbol{\mu}(x)(y-\boldsymbol{\mu}(x)'\boldsymbol{\beta}^*), \tag{16}$$

where we abbreviated $v_*(x,\boldsymbol{\beta}^*)$ by v_*.

In order to evaluate $\mathrm{E}\boldsymbol{\psi}(y,x,\boldsymbol{\beta}^*)$ we need to compute $E(y|x)$. We cannot use (3) anymore because x now follows a mixture law like (12), but with $\sigma^2+\sigma_\delta^2$ in place of σ^2. If we denote the difference $\mu_2-\mu_1$ by ϑ and if we set $\mu=\frac{1}{2}(\mu_1+\mu_2)$, then

$$\mu_1 = \mu - \tfrac{1}{2}\vartheta, \mu_2 = \mu + \tfrac{1}{2}\vartheta, \tag{17}$$

$$\mu_x = p_1\mu_1 + p_2\mu_2 = \mu + \tfrac{1}{2}(p_2-p_1)\vartheta \tag{18}$$

$$\sigma_x^2 = \sigma^2 + \sigma_\delta^2 + p_1 p_2 \vartheta^2. \tag{19}$$

Let us introduce the indicator variable I such that $I=j$ if ξ comes from $\mathrm{N}(\mu_j,\sigma^2), j=1,2$. We have $\mathrm{P}(I=j)=p_j$ and by Bayes' Theorem

$$\pi_j(x) := \mathrm{P}(I=j|x) = \frac{p_j\exp(-\frac{(x-\mu_j)^2}{2(\sigma^2+\sigma_\delta^2)})}{\sum_{i=1}^2 p_i\exp(-\frac{(x-\mu_i)^2}{2(\sigma^2+\sigma_\delta^2)})}. \tag{20}$$

Furthermore, the conditional distribution of ξ, given x and $I=j$, is

$$\xi|(x,I=j) \sim \mathrm{N}(\mu_j(x),\tau_j^2),$$

where in analogy to (6) and (7)

$$\mu_j(x) = \mu_j + \left(1-\frac{\sigma_\delta^2}{\sigma^2+\sigma_\delta^2}\right)(x-\mu_j) \tag{21}$$

and

$$\tau_j^2 = \sigma_\delta^2\left(1-\frac{\sigma_\delta^2}{\sigma^2+\sigma_\delta^2}\right). \tag{22}$$

We can now compute $\mathrm{E}(y|x)$:

$$\begin{aligned}\mathrm{E}(y|x) &= \sum_{j=1}^{2} \mathrm{P}(I=j|x) E(y|x, I=j) \\ &= \sum_{j=1}^{2} \pi_j(x) \boldsymbol{\mu}_j(x)'\boldsymbol{\beta}. \qquad (23)\end{aligned}$$

Here $\boldsymbol{\mu}_j(x)$ is defined in the same way as $\boldsymbol{\mu}(x)$, but with $\mu_j(x)$ and τ_j^2 in place of $\mu(x)$ and τ^2, respectively, see (6) to (9). Finally, by (16) and (23),

$$\mathrm{E}\boldsymbol{\psi}(y,x,\boldsymbol{\beta}^*) = \mathrm{E}\{\mathrm{E}[\boldsymbol{\psi}(y,x,\boldsymbol{\beta}^*)|x]\}$$
$$= \sum_{j=1}^{2} \mathrm{E}[v_*^{-1} \pi_j(x) \boldsymbol{\mu}(x) \{\boldsymbol{\mu}_j(x)'\boldsymbol{\beta} - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*\}]. \qquad (24)$$

Now, $\boldsymbol{\mu}(x)$ being dependent on $\mu(x)$ and τ^2, see (5) to (9), and the latter two being functions of σ_δ^2 and ϑ by (6), (7), (18), (19), $\boldsymbol{\mu}(x)$ is also a function of σ_δ^2 and ϑ. The same is true for $\boldsymbol{\mu}_j(x)$. Finally $\sigma_{\epsilon_*}^2$ is a function of $\boldsymbol{\beta}^*, \sigma_\delta^2, \vartheta$ by (14), and hence v_* is a function of $\boldsymbol{\beta}^*, \sigma_\delta^2$, and ϑ, too. Thus we can write

$$E\boldsymbol{\psi}(y,x,\boldsymbol{\beta}^*) = \boldsymbol{f}(\boldsymbol{\beta}^*, \sigma_\delta^2, \vartheta), \qquad (25)$$

where $\boldsymbol{f}$ is a differentiable vector-valued function in $k+3$ arguments. If we solve the equation $\boldsymbol{f}(\boldsymbol{\beta}^*, \sigma_\delta^2, \vartheta) = 0$, the solution $\boldsymbol{\beta}^*$ will be a differentiable function of σ_δ^2 and ϑ:

$$\boldsymbol{\beta}^* = \boldsymbol{\beta}^*(\sigma_\delta^2, \vartheta).$$

4 Bias for small error variances and small deviations

We want to study the dependence of $\boldsymbol{\beta}^*$ and of the bias $\boldsymbol{b} = \boldsymbol{\beta}^* - \boldsymbol{\beta}$ on σ_δ^2 and ϑ. In particular we investigate this dependence for fixed ϑ and small σ_δ^2 and for fixed σ_δ^2 and small ϑ; i.e., we consider the derivatives $\frac{\partial \boldsymbol{\beta}^*}{\partial \sigma_\delta}$ at $\sigma_\delta = 0$ and $\frac{\partial \boldsymbol{\beta}^*}{\partial \vartheta}$ at $\vartheta = 0$. These derivatives are also the derivatives of the bias $\boldsymbol{b}$ with respect to σ_δ and ϑ, respectively.

First note that $\boldsymbol{\beta}^*(0,\vartheta) = \boldsymbol{\beta}$ because SLS becomes OLS for $\sigma_\delta^2 = 0$ and OLS is consistent for $\sigma_\delta^2 = 0$. Similarly, $\boldsymbol{\beta}^*(\sigma_\delta^2, 0) = \boldsymbol{\beta}$ because SLS is consistent when the normality assumption for ξ is satisfied, i.e., when $\vartheta = 0$.

Therefore by the differentiation rule for implicit functions

$$\left.\frac{\partial \boldsymbol{b}}{\partial \sigma_\delta}\right|_{\sigma_\delta=0} = -\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}}\right)^{-1} \left.\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta}\right|_{\substack{\sigma_\delta=0\\ \boldsymbol{\beta}^*=\boldsymbol{\beta}}} \tag{26}$$

and

$$\left.\frac{\partial \boldsymbol{b}}{\partial \vartheta}\right|_{\vartheta=0} = -\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}}\right)^{-1} \left.\frac{\partial \boldsymbol{f}}{\partial \vartheta}\right|_{\substack{\vartheta=0\\ \boldsymbol{\beta}^*=\boldsymbol{\beta}}} . \tag{27}$$

Our main result can now be expressed in the following theorem.

Theorem 4.1 *The derivatives of the bias with respect to σ_δ and ϑ are zero at $\sigma_\delta = 0$ and at $\vartheta = 0$, respectively.*

Proof: First note that if $\sigma_\delta^2 = 0$, then by (6), (7), (21), (22) $\tau_j^2 = \tau^2 = 0$, $\mu_j(x) = \mu(x) = x$, and consequently $\boldsymbol{\mu}_j(x) = \boldsymbol{\mu}(x)=(1, x, \ldots, x^k)'$. Similarly by (17) to (19), if $\vartheta = 0$, then $\tau_j^2 = \tau^2$, $\mu_j(x) = \mu(x)$, and again $\boldsymbol{\mu}_j(x) = \boldsymbol{\mu}(x)$.

In both cases $\sigma_{\epsilon^*}^2 = \sigma_\epsilon^2$ if in addition $\boldsymbol{\beta}^* = \boldsymbol{\beta}$. Indeed, for $\sigma_\delta^2 = 0$, $\boldsymbol{M}(x) = \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'$ and then, by (14) with $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, $\sigma_{\epsilon^*}^2 = \mathrm{E}\{y - (1, \xi, \ldots, \xi^k)'\boldsymbol{\beta}\} = \sigma_\epsilon^2$ by (1). For $\vartheta = 0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, according to (14), $\sigma_{\epsilon^*}^2 = \mathrm{E}[v(x, \boldsymbol{\beta}, \sigma_\epsilon^2) - \boldsymbol{\beta}'\{\boldsymbol{M}(x) - \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'\}\boldsymbol{\beta}] = \sigma_\epsilon^2$ by (4). Thus in both cases $v_* = v(x, \boldsymbol{\beta}, \sigma_\epsilon^2)$.

Note also that, according to the remark just after (23), we can write

$$\boldsymbol{\mu}(x) = \boldsymbol{m}\{\mu(x), \tau^2\} \tag{28}$$

$$\boldsymbol{\mu}_j(x) = \boldsymbol{m}\{\mu_j(x), \tau_j^2\} \tag{29}$$

with the same function $\boldsymbol{m}$ of two arguments. Now we are ready to evaluate $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}}$ according to (25) and (24):

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}} = -\sum_{j=1}^{2} E[v_*^{-1}\pi_j(x)\boldsymbol{\mu}(x)\boldsymbol{\mu}(x)']$$

$$+\sum_{j=1}^{2} E[\pi_j(x)\boldsymbol{\mu}(x)\{\boldsymbol{\mu}_j(x)'\boldsymbol{\beta} - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*\}\frac{\partial v_*^{-1}}{\partial \boldsymbol{\beta}^{*\prime}}].$$

For $\boldsymbol{\beta}^* = \boldsymbol{\beta}$ and either $\sigma_\delta = 0$ or $\vartheta = 0$, the last term vanishes because, as noted before, $\boldsymbol{\mu}_j(x) = \boldsymbol{\mu}(x)$ in these cases. Thus we have, because $\pi_1 + \pi_2 = 1$,

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}} = -\mathrm{E}[v_*^{-1}\boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'], \tag{30}$$

where all terms are taken at $\boldsymbol{\beta}^* = \boldsymbol{\beta}$ and either $\sigma_\delta = 0$ or $\vartheta = 0$. As $v_* = v(x, \boldsymbol{\beta}, \sigma_\epsilon^2) \geq \sigma_\epsilon^2 > 0$ and $\mathrm{E}\boldsymbol{\mu}(x)\boldsymbol{\mu}(x)' > 0$, the matrix $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}}$ is nonsingular, and so (26) and (27) are well-defined.

Consider (26). As $\boldsymbol{f}$ is a differentiable function of σ_δ^2, for $\sigma_\delta^2 > 0$, we have $\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta} = (\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2})(\frac{\partial \sigma_\delta^2}{\partial \sigma_\delta}) = (\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2})2\sigma_\delta$, which tends to zero for $\sigma_\delta \to 0$, because $|\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2}| < \infty$ for $\sigma_\delta \to 0$, see 7.2. This proves the first part of the theorem.

Now consider (27). With (28), (29) we get from (25), (24)

$$\begin{aligned}\frac{\partial \boldsymbol{f}}{\partial \vartheta} &= \sum_{j=1}^{2} \mathrm{E}\left[v_*^{-1}\pi_j(x)\boldsymbol{\mu}(x)\left\{\frac{\partial \boldsymbol{m}'}{\partial \mu_j(x)}\frac{\partial \mu_j(x)}{\partial \vartheta}\boldsymbol{\beta} - \frac{\partial \boldsymbol{m}'}{\partial \mu(x)}\frac{\partial \mu(x)}{\partial \vartheta}\boldsymbol{\beta}^*\right\}\right] \\ &\quad + \sum_{j=1}^{2} \mathrm{E}\left[\frac{\partial}{\partial \vartheta}\{v_*^{-1}\pi_j(x)\boldsymbol{\mu}(x)\}\{\boldsymbol{\mu}_j(x)'\boldsymbol{\beta} - \boldsymbol{\mu}(x)'\boldsymbol{\beta}^*\}\right]\end{aligned}$$

For $\boldsymbol{\beta}^* = \boldsymbol{\beta}$ and $\vartheta = 0$, the last sum vanishes and in addition $\frac{\partial \boldsymbol{m}'}{\partial \mu_j(x)} = \frac{\partial \boldsymbol{m}'}{\partial \mu(x)}$ and, by (20), $\pi_j(x) = p_j$. Thus

$$\left.\frac{\partial \boldsymbol{f}}{\partial \vartheta}\right|_{\substack{\vartheta=0 \\ \boldsymbol{\beta}^*=\boldsymbol{\beta}}} = \mathrm{E}\left[v_*^{-1}\boldsymbol{\mu}(x)\frac{\partial \boldsymbol{m}'}{\partial \mu(x)}\left\{\sum_{j=1}^{2} p_j \frac{\partial \mu_j(x)}{\partial \vartheta} - \frac{\partial \mu(x)}{\partial \vartheta}\right\}\beta\right], \tag{31}$$

where the derivatives on the right hand side are also taken at $\vartheta = 0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$. From (21) and (17)

$$\frac{\partial \mu_j(x)}{\partial \vartheta} = \frac{\sigma_\delta^2}{\sigma^2 + \sigma_\delta^2}\frac{\partial \mu_j}{\partial \vartheta} = (-1)^j \frac{1}{2}\frac{\sigma_\delta^2}{\sigma^2 + \sigma_\delta^2}. \tag{32}$$

From (6) and (18), (19)

$$\begin{aligned}\frac{\partial \mu(x)}{\partial \vartheta} &= \frac{\sigma_\delta^2}{\sigma_x^2}\frac{\partial \mu_x}{\partial \vartheta} + \frac{\sigma_\delta^2(x - \mu_x)}{\sigma_x^4}\frac{\partial \sigma_x^2}{\partial \vartheta} = \\ &= \frac{\sigma_\delta^2}{\sigma_x^2}\frac{1}{2}(p_2 - p_1) + \frac{\sigma_\delta^2(x - \mu_x)}{\sigma_x^4}p_1 p_2 2\vartheta.\end{aligned}$$

For $\vartheta = 0$,

$$\left.\frac{\partial \mu(x)}{\partial \vartheta}\right|_{\sigma_\delta = 0} = \frac{1}{2}(p_2 - p_1)\frac{\sigma_\delta^2}{\sigma^2 + \sigma_\delta^2}. \tag{33}$$

Substituting $\frac{\partial \mu_j(x)}{\partial \vartheta}$ and $\frac{\partial \mu(x)}{\partial \vartheta}$ from (32) and (33) in (31), we see that (31) vanishes, which proves the second part of the theorem.

5 Simulations

To study the effect of small and also of large deviations from normality on the bias of $\hat{\beta}$ we carried out a few simulations with a quadratic regression model ($k = 2$) with $\beta = (0, 1, -0.5)'$ and $\sigma_\epsilon^2 = 1$ and a two-components mixture distribution for ξ with $\mu_1 = -\frac{\vartheta}{2}$, $\mu_2 = +\frac{\vartheta}{2}$, $\sigma^2 = 1$, and $p_1 = p_2$, where ϑ ranged from 0 to 5. The error standard deviation σ_δ ranged from 0 to 1. The sample size was taken to be $n = 500$. The bias was computed from $N = 1000$ replicated estimates of β.

In addition to the bias we also computed the standard deviations of the estimators, even though they were not discussed in the present paper. Each figure with the exception of Figure 5 is accompanied by a corresponding table. For Figure 5 Table 1 has the underlying quantities.

Figures 1 and 2 show the bias of β_0, β_1, and β_2 as depending on σ_δ when $\vartheta = 4.5$ is kept fixed, or as depending on ϑ when $\sigma_\delta^2 = 0.5$ is kept fixed, respectively. It is clearly seen that the curves approach the origin (0,0) completely flat, and so the simulation results corroborate the theoretical findings of our theorem in Section 4.

Figure 3 presents the bias of $\hat{\beta}_1$ as depending on ϑ for fixed σ_δ^2, where σ_δ^2 was fixed at the values $0.5, 0.75, 1$, and 1.5. As expected, the bias becomes larger with increasing error variance.

In Figure 4 we also present the bias for a few more estimators: the naive, the Regression Calibration (RC), and the MALS estimator. The RC estimator is computed by replacing the variable ξ in the regression model (1) with $\mu(x)$ an then applying OLS, Carroll et al (1995). MALS is a modified ALS method that reduces the small sample bias of ALS, Cheng et al (2000). The naive estimator has the largest bias, but it becomes smaller with increasing ϑ because the variance of ξ increases together with ϑ. The bias of SLS becomes larger due to misspecification, whereas

MALS and also RC hardly have any bias. Their standard deviations, shown in Table 4, are however larger than for SLS.

In Figure 5 one can see how the bias varies in dependence on the error variance instead of the error standard deviation, see also Appendix 7.2.

Figure 1: *Bias of SLS-Estimators as a function of the standard deviation of the measurement error σ_δ with $\vartheta = 4.5$*

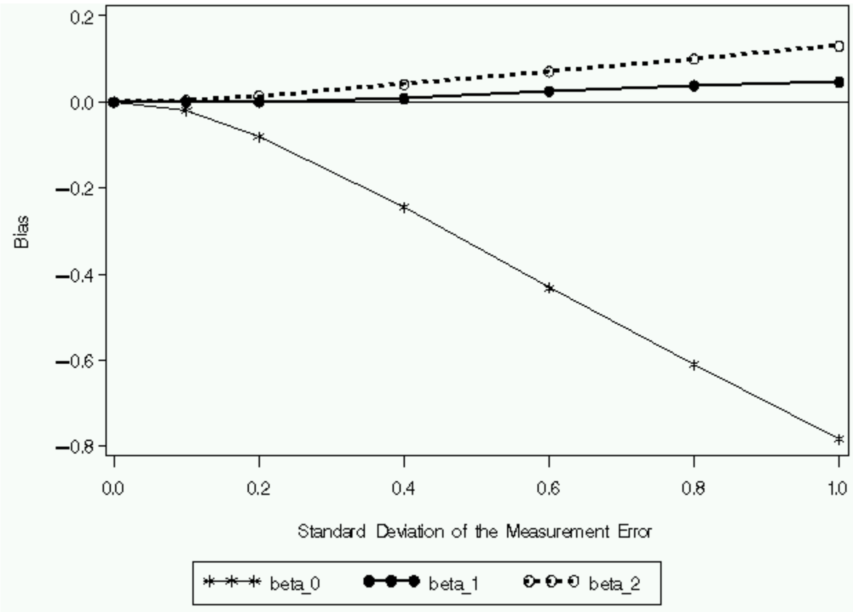


Table 1: *Bias (Standard deviation) of SLS-Estimators as a function of the standard deviation of the measurement error σ_δ with $\vartheta = 4.5$*

	σ_δ					
	0.1	0.2	0.4	0.6	0.8	1.0
β_0: Bias	0.020	-0.080	-0.244	-0.432	-0.610	-0.784
(STD)	(0.075)	(0.078)	(0.095)	(0.113)	(0.141)	(0.185)
β_1: Bias	0.000	0.000	0.008	0.025	0.039	0.047
(STD)	(0.019)	(0.022)	(0.028)	(0.036)	(0.043)	(0.051)
β_2: Bias	0.004	0.013	0.041	0.071	0.100	0.130
(STD)	(0.010)	(0.011)	(0.015)	(0.019)	(0.026)	(0.032)

Figure 2: *Bias of SLS-Estimators as a function of ϑ with $\sigma_\delta^2 = 0.5$*

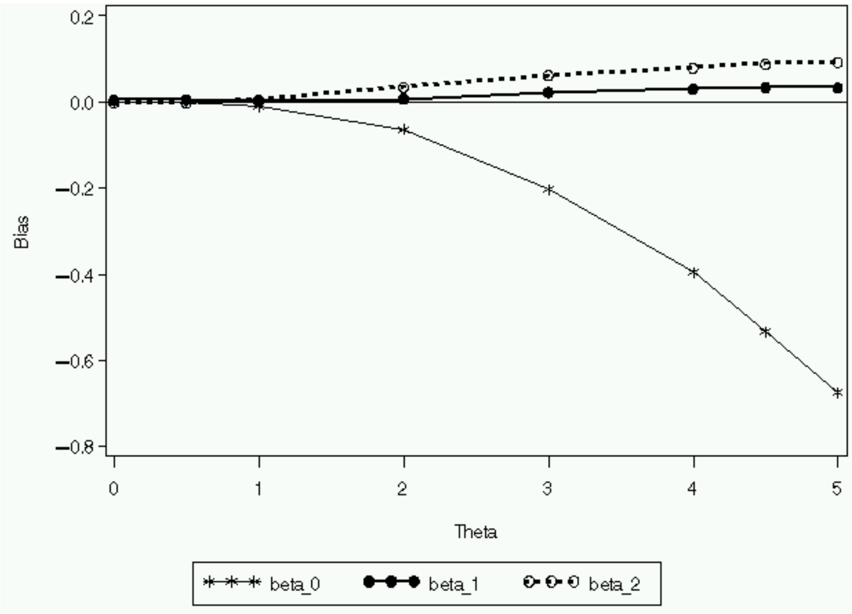


Table 2: *Bias (Standard deviation) of SLS-Estimators as a function of ϑ with $\sigma_\delta^2 = 0.5$.*

	ϑ							
	0.0	0.5	1.0	2.0	3.0	4.0	4.5	5.0
β_0: Bias	-0.001	0.000	-0.010	-0.064	-0.203	-0.395	-0.533	-0.674
(STD)	(0.084)	(0.084)	(0.084)	(0.086)	(0.095)	(0.116)	(0.133)	(0.152)
β_1: Bias	0.004	0.006	0.004	0.007	0.023	0.030	0.034	0.033
(STD)	(0.085)	(0.085)	(0.077)	(0.059)	(0.049)	(0.044)	(0.040)	(0.039)
β_2: Bias	-0.002	-0.003	0.005	0.034	0.062	0.079	0.088	0.092
(STD)	(0.075)	(0.071)	(0.063)	(0.042)	(0.029)	(0.024)	(0.023)	(0.022)

Figure 3: *Bias of SLS-Estimator of* β_1 *as a function of* ϑ *for various values of* σ_δ^2

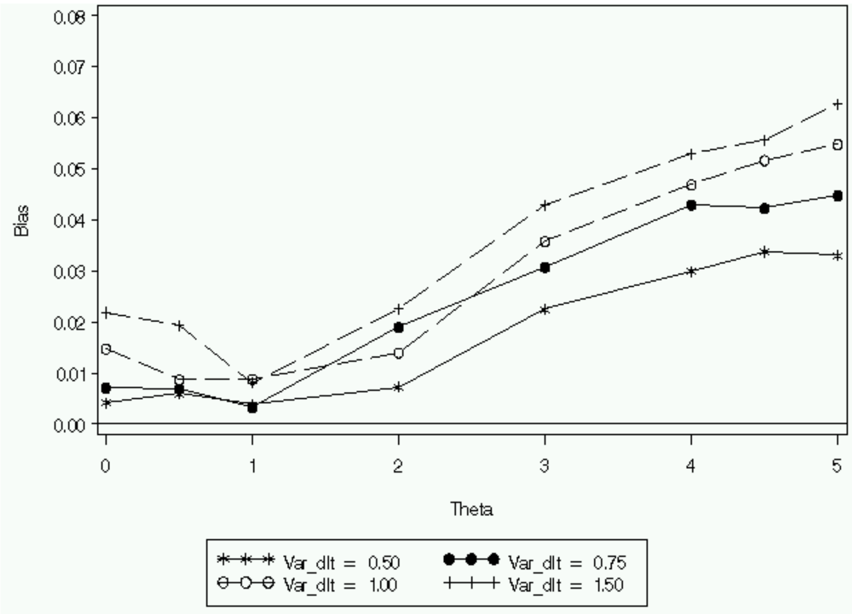


Table 3: *Bias (Standard deviation) of SLS-Estimators* β_1 *as a function of* ϑ *for various values of* σ_δ^2.

	ϑ							
	0	0.5	1.0	2.0	3.0	4.0	4.5	5.0
$\sigma_\delta^2 = 0.5$: Bias	0.004	0.006	0.004	0.007	0.023	0.030	0.034	0.033
(STD)	(0.085)	(0.085)	(0.077)	(0.059)	(0.049)	(0.044)	(0.040)	(0.039)
$\sigma_\delta^2 = 0.75$: Bias	0.007	0.007	0.003	0.019	0.031	0.043	0.042	0.045
(STD)	(0.107)	(0.102)	(0.094)	(0.076)	(0.060)	(0.051)	(0.048)	(0.045)
$\sigma_\delta^2 = 1.0$: Bias	0.015	0.009	0.009	0.014	0.036	0.047	0.052	0.055
(STD)	(0.124)	(0.123)	(0.109)	(0.082)	(0.067)	(0.056)	(0.053)	(0.051)
$\sigma_\delta^2 = 1.5$: Bias	0.022	0.019	0.008	0.023	0.043	0.053	0.056	0.063
(STD)	(0.166)	(0.155)	(0.136)	(0.103)	(0.083)	(0.067)	(0.064)	(0.060)

Figure 4: *Bias of different estimators of* β_1 *as a function of* ϑ *with* $\sigma_\delta^2 = 0.5$

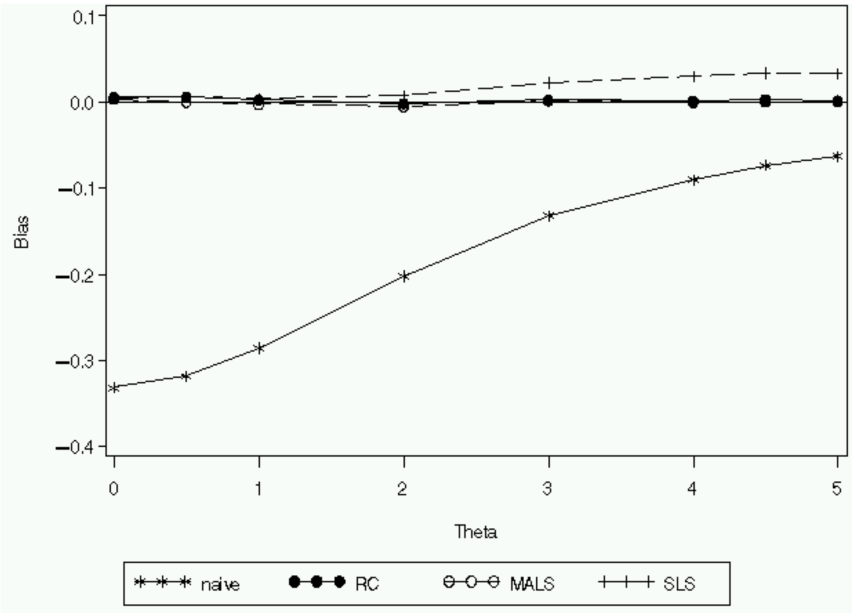
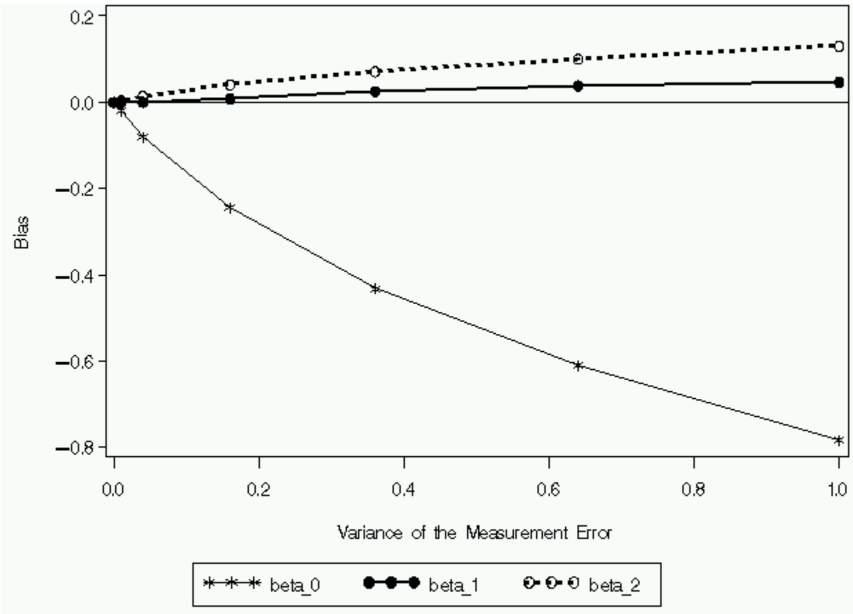


Table 4: *Bias (Standard deviation) of different estimators of* β_1 *as a function of* ϑ *with* $\sigma_\delta^2 = 0.5$.

	ϑ							
	0	0.5	1.0	2.0	3.0	4.0	4.5	5.0
naive: Bias	-0.331	-0.318	-0.285	-0.202	-0.132	-0.090	-0.073	-0.063
(STD)	(0.053)	(0.053)	(0.051)	(0.046)	(0.043)	(0.040)	(0.037)	(0.037)
MALS: Bias	0.004	-0.000	-0.003	-0.005	0.002	-0.001	0.002	0.002
(STD)	(0.110)	(0.108)	(0.095)	(0.075)	(0.061)	(0.054)	(0.049)	(0.048)
RC: Bias	0.006	0.006	0.003	-0.002	0.003	0.001	0.003	0.002
(STD)	(0.087)	(0.087)	(0.0880)	(0.061)	(0.051)	(0.045)	(0.040)	(0.040)
SLS: Bias	0.004	0.006	0.004	0.007	0.023	0.030	0.034	0.033
(STD)	(0.085)	(0.085)	(0.077)	(0.059)	(0.049)	(0.044)	(0.040)	(0.039)

Figure 5: *Bias of SLS-Estimators as a function of the Variance of the measurement error* σ_δ^2 *with* $\vartheta = 4.5$



6 Conclusion

We investigated the bias of the structural least squares estimation method for a polynomial regression with measurement errors under misspecification of the distribution of the latent covariate ξ. We studied only a very specific kind of misspecification, but it seems plausible that the results found should also apply to other kinds of misspecification.

Whenever the true distribution of ξ differs from the assumed normal distribution by a small amount ϑ, measured in some appropriate way, the bias will be of order ϑ^2. Here ϑ was taken to be the distance between the two means of a two-normal-components mixture distribution with equal variances. In a simulation study we also experimented with a mixture distribution where the two variances depended on ϑ in such a way that the overall variance of ξ was kept constant. The same result was found: $\frac{\partial \boldsymbol{b}}{\partial \vartheta} = 0$ or, equivalently, $\boldsymbol{b} = O(\vartheta^2)$, and this is true for any value of σ_δ^2.

A similar result exists with respect to the dependence of the bias on the measurement error variance: $\boldsymbol{b} = O(\sigma_\delta^2)$, whatever the amount of deviation ϑ.

Two caveats should be mentioned. First, there are other consistent estimation methods, notably Adjusted Least Squares (ALS), which do not depend on a specific distribution of ξ and are therefore robust against a distributional misspecification. This robustness however has to be payed for: ALS is usually less efficient than SLS when ξ is in fact normal, even though the small sample variance of ALS can be greatly reduced by a modified ALS method, Cheng et al (2000).

Second, one can modify SLS if a mixture of normals is the underlying latent distribution by taking this mixture into account. The ingredients for this modification have been given in this paper, see (20) to (23), see also Thamerus (1998, 2003), Augustin (2002). This modification should render SLS much more flexible, but this also has its price. It is often difficult to estimate the components of the mixture especially if the number of components is unknown. This difficulty will add to the variance of the SLS estimator of β making it less efficient.

A general recommendation therefore might be to use the simple SLS method whenever one has good reason to suppose that the true latent distribution comes close to a normal one and to use a modified SLS or ALS (or better still MALS) when the latent distribution differs severely from the normal one.

7 Appendix

7.1 Proof of (3) and (4)

We introduce the vector $\boldsymbol{\xi} = (1, \xi, \cdots, \xi^k)'$ so that (1) becomes $y = \boldsymbol{\xi}'\boldsymbol{\beta} + \varepsilon$ with the index i being suppressed. Then

$$
\begin{aligned}
\mathrm{E}(y|x) &= \mathrm{E}[\mathrm{E}(y|\xi, x)|x] = \mathrm{E}[\mathrm{E}(y|\xi)|x] = \mathrm{E}(\boldsymbol{\xi}|x)'\boldsymbol{\beta} \\
&= \boldsymbol{\mu}(x)'\boldsymbol{\beta}. \\
\mathrm{V}(y|x) &= \mathrm{E}[(y - \boldsymbol{\mu}(x)'\boldsymbol{\beta})^2|x] = \mathrm{E}[\{(\boldsymbol{\xi} - \boldsymbol{\mu}(x))'\boldsymbol{\beta} + \varepsilon\}^2|x] \\
&= \sigma_\varepsilon^2 + \boldsymbol{\beta}'\mathrm{E}[(\boldsymbol{\xi} - \boldsymbol{\mu}(x))(\boldsymbol{\xi} - \boldsymbol{\mu}(x))'|x]\boldsymbol{\beta} \\
&= \sigma_\varepsilon^2 \boldsymbol{\beta}'\{\mathrm{E}[\boldsymbol{\xi}\boldsymbol{\xi}'|x] - \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'\}\boldsymbol{\beta} \\
&= \sigma_\varepsilon^2 \boldsymbol{\beta}'\{\mathbf{M}(x) - \boldsymbol{\mu}(x)\boldsymbol{\mu}(x)'\}\boldsymbol{\beta}.
\end{aligned}
$$

7.2 The second derivative of $\boldsymbol{b}$ with respect to σ_δ

Let us denote the partial function of the bias $\boldsymbol{b}$, for fixed ϑ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, as a function of σ_δ by $\boldsymbol{b}(\sigma_\delta)$ and as a function of σ_δ^2 by $\tilde{\boldsymbol{b}}(\sigma_\delta^2)$. Then

$$
\begin{aligned}
\frac{d\boldsymbol{b}}{d\sigma_\delta} &= \frac{d\tilde{\boldsymbol{b}}}{d\sigma_\delta^2} 2\sigma_\delta \\
\frac{d^2\boldsymbol{b}}{(d\sigma_\delta)^2} &= \frac{d\tilde{\boldsymbol{b}}}{d\sigma_\delta^2} 2 + \frac{d^2\tilde{\boldsymbol{b}}}{(d\sigma_\delta^2)^2} 4\sigma_\delta^2.
\end{aligned}
$$

For $\sigma_\delta^2 = 0$, therefore

$$
\frac{d^2\boldsymbol{b}}{(d\sigma_\delta)^2} = 2\frac{d\tilde{\boldsymbol{b}}}{d\sigma_\delta^2}.
$$

The Taylor series of $\boldsymbol{b}$ up to the second order term in σ_δ^2 is therefore

$$
\begin{aligned}
\boldsymbol{b} = \boldsymbol{b}(0) + \frac{d\boldsymbol{b}}{d\sigma_\delta}(0)\sigma_\delta + \frac{1}{2}\frac{d^2\boldsymbol{b}}{(d\sigma_\delta)^2}(0)\sigma_\delta^2 + O(\sigma_\delta^3) \\
= \frac{d\tilde{\boldsymbol{b}}}{d\sigma_\delta^2}(0)\sigma_\delta^2 + O(\sigma_\delta^3)
\end{aligned}
$$

In analogy to (26), we can compute $d\tilde{\boldsymbol{b}}/d\sigma_\delta^2$ at $\sigma_\delta^2 = 0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$ as

$$
\frac{d\tilde{\boldsymbol{b}}}{d\sigma_\delta^2} = -\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\beta}^{*\prime}}\right)^{-1} \frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2},
$$

where all derivatives here and in the sequel are taken at $\sigma_\delta^2 = 0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}$ and where the first factor is given by (30) and the second by

$$\begin{aligned}
\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2} &= \sum_{j=1}^{2} \mathrm{E}\left[v_*^{-1} \pi_j(x) \boldsymbol{\mu}(x) \left\{ \frac{\partial \mathbf{m}'}{\partial \mu_j(x)} \frac{\partial \mu_j(x)}{\partial \sigma_\delta^2} - \frac{\partial \mathbf{m}'}{\partial \mu(x)} \frac{\partial \mu(x)}{\partial \sigma_\delta^2} \right\} \boldsymbol{\beta} \right] \\
&= \mathrm{E}\left[v_*^{-1} \boldsymbol{\mu}(x) \frac{\partial \boldsymbol{m}'}{\partial \mu(x)} \left(\frac{x - \sum p_j \mu_j}{\sigma^2 + p_1 p_2 \vartheta^2} - \frac{x - \Sigma \pi_j(x) \mu_j}{\sigma^2} \right) \boldsymbol{\beta} \right]
\end{aligned}$$

because by (28), (29) as well as (21), (22) and (6), (7)

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}_j(x)}{\partial \sigma_\delta^2} &= \frac{\partial \boldsymbol{m}(x)}{\partial \mu_j(x)} \frac{\partial \mu_j(x)}{\partial \sigma_\delta^2} + \frac{\partial \boldsymbol{m}(x)}{\partial \tau_j^2} \frac{\partial \tau_j^2}{\partial \sigma_\delta^2} \\
&= -\frac{\partial \boldsymbol{m}(x)}{\partial \mu_j(x)} \frac{x - \mu_j}{\sigma^2} + \frac{\partial \boldsymbol{m}(x)}{\partial \tau_j^2} \\
\frac{\partial \boldsymbol{\mu}(x)}{\partial \sigma_\delta^2} &= \frac{\partial \boldsymbol{m}(x)}{\partial \mu(x)} \frac{\partial \mu(x)}{\partial \sigma_\delta^2} + \frac{\partial \boldsymbol{m}(x)}{\partial \tau^2} \frac{\partial \tau^2}{\partial \sigma_\delta^2} \\
&= -\frac{\partial \boldsymbol{m}(x)}{\partial \mu(x)} \frac{x - \mu_x}{\sigma_\xi^2} + \frac{\partial \boldsymbol{m}(x)}{\partial \tau^2}
\end{aligned}$$

with $\sigma_\xi^2 = \sigma^2 + p_1 p_2 \vartheta^2$, see (19), and because $\mu_j(x) = \mu(x)$, $\tau_j^2 = \tau^2$ for $\sigma_\delta^2 = 0$. Obviously $\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2}$ is not zero in general. More importantly, $\frac{\partial \boldsymbol{f}}{\partial \sigma_\delta^2}$ stays finite when σ_δ tends to 0. See also Figure 5.

7.3 The linear regression ($k = 1$)

In the linear case we have

$$\boldsymbol{\mu}(x) = (1, \mu(x))', \quad \boldsymbol{\mu}_j(x) = (1, \mu_j(x))', \quad j = 1, 2,$$

where according to (6) and (21), respectively,

$$\begin{aligned}
\mu(x) &= \mu_x + (1 - e)(x - \mu_x) = (1 - e)x + e\mu_x && (34) \\
\mu_j(x) &= (1 - d)x + d\mu_j && (35) \\
\text{with} \quad e &= \sigma_\delta^2 / \sigma_x^2 \quad \text{and} \quad d = \sigma_\delta^2 / (\sigma^2 + \sigma_\delta^2). && (36)
\end{aligned}$$

It follows that

$$\begin{aligned}
\boldsymbol{M}(x) - \boldsymbol{\mu}(x)\boldsymbol{\mu}'(x) &= \begin{pmatrix} 1 & \mu(x) \\ \mu(x) & \mu^2(x) \end{pmatrix} - \begin{pmatrix} 1 & \mu(x) \\ \mu(x) & \mu(x)^2 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & \tau^2 \end{pmatrix}
\end{aligned}$$

because, by (8) and (9), $\mu^2(x) = \mu(x)^2 + \tau^2$. Hence

$$v(x, \boldsymbol{\beta}, \sigma_\varepsilon^2) = \sigma_\varepsilon^2 + \beta_1^2 \tau^2$$

is a constant.
We can therefore solve the estimating equation (10) without regard to (11) by simply deleting $v(x, \hat{\beta}, \hat{\sigma}_\varepsilon)$. Equation (10) then reduces to

$$\begin{aligned}
\hat{\beta}_0 + \overline{\mu(x)}\hat{\beta}_1 &= \bar{y} \\
\overline{\mu(x)}\hat{\beta}_0 + \overline{\mu(x)^2}\hat{\beta}_1 &= \overline{y\mu(x)},
\end{aligned} \tag{37}$$

where the bar denotes averaging over $i = 1, \ldots, n$. After replacing the parameters μ_x and σ_x^2 in (34) by their estimates $\bar{x}$ and s_x^2, respectively, we get the identities

$$\overline{\mu(x)} = \bar{x}, \quad \overline{\mu(x)^2} = \bar{x}^2 + (1-e)^2 s_x^2, \quad \overline{y\mu(x)} = \bar{y}\bar{x} + (1-e)s_{xy}.$$

The estimating equations (37) then become

$$\begin{aligned}
\hat{\beta}_0 \quad &+ \quad \bar{x}\hat{\beta}_1 = \bar{y} \\
\bar{x}\hat{\beta}_0 \quad &+ \quad (\bar{x}^2 + (1-e)^2 s_x^2)\hat{\beta}_1 = \bar{y}\bar{x} + (1-e)s_{xy},
\end{aligned}$$

from which we derive the following equation for $\hat{\beta}_1$:

$$(1-e)s_x^2 \hat{\beta}_1 = s_{xy}$$

with the familiar solution

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2 - \sigma_\delta^2}.$$

This estimator has been derived without any distributional assumption for ξ or x. It is a simple method of moments estimator and is consistent whatever the distribution of ξ. This can be shown by simply computing

plim $\hat{\beta}_1$. (The same is true for $\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$).
The consistency of $\hat{\boldsymbol{b}}$ in the case $k = 1$, irrespective of how ξ is distributed, should imply that $\mathrm{E}\psi(y, x, \boldsymbol{\beta}^*) = 0$ for $\boldsymbol{\beta}^* = \boldsymbol{\beta}$. It is illuminating to show this directly from (24). This will provide a check for (24). As v_* is a constant, we only need to show that

$$\mathrm{E}\left[\sum_{j=1}^{2} \pi_j(x)\boldsymbol{\mu}(x)\{\boldsymbol{\mu}_j(x) - \boldsymbol{\mu}(x)\}'\right] \quad = \quad 0$$

or

$$\mathrm{E}[\boldsymbol{\mu}(x)\sum_{j=1}^{2} \pi_j(x)\boldsymbol{\mu}_j(x)'] \quad = \quad \mathrm{E}[\boldsymbol{\mu}(x)\boldsymbol{\mu}(x)']$$

or

$$\mathrm{E}\begin{pmatrix} 1 & \Sigma\pi_j(x)\mu_j(x) \\ \mu(x) & \Sigma\pi_j(x)\mu_j(x)\mu(x) \end{pmatrix} \quad = \quad \mathrm{E}\begin{pmatrix} 1 & \mu(x) \\ \mu(x) & \mu(x)^2 \end{pmatrix}.$$

Now, from (34),

$$\mathrm{E}[\mu(x)] = \mu_x, \quad \mathrm{E}[\mu(x)^2] = \mu_x^2 + (1-e)^2\sigma_x^2.$$

So we have to show that

(a) $\mathrm{E}\left[\sum_{j=1}^{2} \pi_j(x)\mu_j(x)\right] = \mu_x$

(b) $\mathrm{E}\left[\sum_{j=1}^{2} \pi_j(x)\mu_j(x)\mu(x)\right] = \mu_x^2 + (1-e)^2\sigma_x^2$.

As to (a): We have with (35) and (17)

$$\begin{aligned} \mathrm{E}\Sigma\pi_j(x)\mu_j(x) \quad &= \quad (1-d)\mathrm{E}x + d\mathrm{E}\Sigma\pi_j(x)\mu_j \\ &= \quad (1-d)\mu_x + d\mu + \tfrac{1}{2}d\vartheta\mathrm{E}[\pi_2(x) - \pi_1(x)]. \end{aligned} \tag{38}$$

But

$$\begin{aligned} \mathrm{E}[\pi_2(x) - \pi_1(x)] \quad &= \quad p_1\mathrm{E}[\pi_2(x)|I=1] + p_2\mathrm{E}[\pi_2(x)|I=2] \\ &- \quad p_1\mathrm{E}[\pi_1(x)|I=1] - p_2\mathrm{E}[\pi_1(x)|I=2], \end{aligned}$$

and, by (20),

$$p_i \mathrm{E}[\pi_j(x)|I = i] = p_i p_j \int \frac{\varphi_i(x)\varphi_j(x)}{p_i\varphi_1(x) + p_2\varphi_2(x)} dx,$$

where $\varphi_j(x)$ is the density of a $N(\mu_j, \sigma^2 + \sigma_\delta^2)$ distribution. Thus

$$\begin{aligned} \mathrm{E}[\pi_2(x) - \pi_1(x)] &= \int \frac{p_2^2\varphi_2(x)^2 - p_1^2\varphi_1(x)^2}{p_i\varphi_1(x) + p_2\varphi_2(x)} dx \\ &= \int [p_2\varphi_2(x) - p_1\varphi_1(x)]dx = p_2 - p_1. \qquad (39) \end{aligned}$$

Substituting this result in (38) we finally get the desired result:

$$\mathrm{E}[\Sigma\pi_j(x)\mu_j(x)] = \mu_x + d(\mu - \mu_x) + \tfrac{1}{2}d\vartheta(p_2 - p_1) = \mu_x$$

because by (18)

$$\mu - \mu_x = \tfrac{1}{2}\vartheta(p_1 - p_2). \qquad (40)$$

As to (b): Using (35), we first have

$$\begin{aligned} & \mathrm{E}[\Sigma\pi_j(x)\mu_j(x)\mu(x)] = (1-d)\mathrm{E}[x\mu(x)] + d\mathrm{E}[\Sigma\pi_j(x)\mu_j\mu(x)] \\ = \; & (1-d)\{\mu_x^2 + (1-e)\sigma_x^2\} + d\mu\mathrm{E}\mu(x) + \tfrac{1}{2}d\vartheta\mathrm{E}[\{\pi_2(x) - \pi_1(x)\}\mu(x)] \\ & [\text{ by (34) and (17)}] \\ = \; & (1-d)\mu_x^2 + (1-d)(1-e)\sigma_x^2 + d\mu\mu_x \\ & + \tfrac{1}{2}d\vartheta\{(1-e)\mathrm{E}[x\pi_2(x) - x\pi_1(x)] + e\mu_x\mathrm{E}[\pi_2(x) - \pi_1(x)]\} \end{aligned}$$

by (34). Now a derivation analogous to the one that led to (39) gives the result

$$\mathrm{E}[x\pi_2(x) - x\pi_1(x)] = p_2\mu_2 - p_1\mu_1.$$

Hence $\mathrm{E}[\Sigma\pi_j(x)\mu_j(x)\mu(x)]$ can be further simplified to

$$
\begin{aligned}
& \mu_x^2+(1-d)(1-e)\sigma_x^2+d(\mu-\mu_x)\mu_x \\
& +\tfrac{1}{2}d\vartheta\{(1-e)(p_2\mu_2-p_1\mu_1)+e\mu_x(p_2-p_1)\} \\
= \;& \mu_x^2+(1-e)^2\sigma_x^2-(1-e)(d-e)\sigma_x^2+\tfrac{1}{2}d\vartheta(p_1-p_2)\mu_x \\
& -\tfrac{1}{2}de\vartheta(p_1-p_2)\mu_x+\tfrac{1}{2}d(1-e)\vartheta(p_2\mu_2-p_1\mu_1) \quad [\text{ by } (40)] \\
= \;& \mu_x^2+(1-e)^2\sigma_x^2-(1-e)d(\sigma_x^2-\sigma^2-\sigma_\delta^2) \\
& +\tfrac{1}{2}d(1-e)\vartheta\{(p_1-p_2)(p_1\mu_1+p_2\mu_2)+p_2\mu_2-p_1\mu_1\} \\
& [\text{ by } (36) \text{ and } (18)] \\
= \;& \mu_x^2+(1-e)^2\sigma_x^2-(1-e)d\vartheta^2p_1p_2+\tfrac{1}{2}d(1-e)\vartheta^2 2p_1p_2 \quad [\text{ by } (19)] \\
= \;& \mu_x^2+(1-e)^2\sigma_x^2,
\end{aligned}
$$

which is the desired result.
A much simpler way to verify $\mathrm{E}\psi(y,x,\boldsymbol{\beta})=0$ is to go back to (13). One can show that (13) is satisfied for $\boldsymbol{\beta}^*=\boldsymbol{\beta}$, in the linear case, without the need to assume any particular distribution of ξ. We note that $\mathrm{E}y=\beta_0+\beta_1\mu_x$ and $\mathrm{E}(yx)=\beta_0\mu_x+\beta_1\sigma_\xi+\beta_1\mu_x^2$. The rest then follows with some algebra when we remember that v_* is a constant.

Acknowledgements: We thank T. Augustin for some valuable comments on an earlier version of the paper. Support by the Deutsche Forschungsgemeinschaft, the Deutsche Akademische Austauschdienst, and the National Science Council of the Republic of China, Taiwan is gratefully acknowledged.

8 References

Armstrong, B. (1985), *Measurement error in the generalized linear model*, Comm. in Stat. - Simul. and Comp. 14, 529-544.

Augustin, T. (2002), *Some basic results on the extension of quasi-likelihood based measurement error correction to multivariate and flexible structural models.* In W. Gaul and G. Ritter (eds.): Classification, Automation, and New Media. Springer, Heidelberg, 29-36.

Card, D. (1999), *The causal effect of education on earnings.* In O. Ashenfelter and D. Card (Eds.): Handbook of Labor Economics, Elsevier, Amsterdam, 1801-1863.

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

Cheng, C.-L. and Schneeweiss, H. (1998), *Polynomial regression with errors in the variables*, J.R. Statist. Soc. B60, 189-199.

Cheng, C.-L. and Schneeweiss, H. (2002), *On the Polynomial Measurement Error Model*. In S. van Huffel and P. Lemmerling (Eds.): Total Least Squares and Errors-in-Variables Modeling. Kluwer, Dordrecht, 131-143.

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

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.

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

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

Kukush, A. und 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.

Kukush, A., Schneeweiss, H. and Wolf, R. (2002), *Comparing Different Estimators in a Nonlinear Measurement Error Model*, Discussion Paper 244, Sonderforschungsbereich 386, University of Munich.

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

Schneeweiss, H. and Nittner, T. (2001), *Estimating a polynomial regression with measurement errors in the structural and in the functional case - a comparison*. In M. Sadeh (Ed.): Data Analysis from Statistical Foundations, A Festschrift in Honour of the 75th Birthday of D.A.S. Fraser, Nova Science, New York, 195-205.

Thamerus, M. (1998), *Nichtlineare Regressionsmodelle mit heteroskedastischen Messfehlern*, Logos Verlag, Berlin.

Thamerus, M. (2003), *Fitting a mixture distribution to a variable subject to heteroscedastic measurement errors*, to be published in Computational Statistics.

Wansbeek, T., Meijer, E. (2000): *Measurement Error and Latent Variables in Econometrics*, Elsevier, Amsterdam.