LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Schneeweiß:

The polynomial and the Poisson measurement error models: some further results on quasi score and corrected score estimation

Sonderforschungsbereich 386, Paper 446 (2005)

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

Projektpartner

MAX-PLANCK-GESELLSCHAFT

The polynomial and the Poisson measurement error models: some further results on quasi score and corrected score estimation

Hans Schneeweiss

Department of Statistics, University of Munich
Akademiestr. 1, 80799 Munich, Germany
schneew@stat.uni-muenchen.de

Abstract

The asymptotic covariance matrices of the corrected score, the quasi score, and the simple score estimators of a polynomial measurement error model have been derived in the literature. Here some alternative formulas are presented, which might lead to an easier computation of these matrices. In particular, new properties of the variables t_r and μ_r that constitute the estimators are derived. In addition, the term in the formula for the covariance matrix of the quasi score estimator stemming from the estimation of nuisance parameters is evaluated. The same is done for the log-linear Poisson measurement error model.

1 Introduction

Despite the many results that have been found in recent years on the estimation of regression coefficients of a polynomial model with measurement errors in the covariable, cf., e.g., Cheng and Schneeweiss (1998), Cheng and Schneeweiss (2002), Kukush *et al.* (2005b), Kukush and Schneeweiss (2005), Shklyar *et al.* (2005), some issues concerning the computation of estimators and their asymptotic covariance matrices (ACM) are still open to investigation. Although the polynomial model is the main subject of this paper, the log-linear Poisson model with measurement errors is dealt with, too. Again, despite the work of Kukush *et al.* (2004) and Shklyar and Schneeweiss (2005), there are still a few properties of the estimators of this model, which have not yet been sufficiently investigated.

The plynomial measurement error model is given by the regression equation

$$y = \zeta^\top \beta + \epsilon,$$

with $\zeta^\top = \left(1, \xi, \cdots, \xi^k\right)$, $\beta := \left(\beta_0, \beta_1, \cdots, \beta_k\right)^\top$, $\mathbb{E}\epsilon = 0$, $\mathbb{V}\epsilon = \sigma_\epsilon^2$, ϵ and ξ independent, and the measurement equation

$$x = \xi + \delta,$$

$\delta \sim N(0, \sigma_\delta^2)$ being the measurement error, which is independent of ξ and ϵ. It is assumed that σ_δ^2 is known. In addition, we here assume that $\xi \sim N(\mu_\xi, \sigma_\xi^2)$. The problem is to estimate β from an i.i.d. sample (x_i, y_i), $i = 1, \cdots, n$.

In addition to the *naive*(N) *estimator*, we consider two consistent estimators: the (structural) *quasi score* (QS) and the (functional) *corrected score* (CS) *estimator*. The first one utilizes the distribution of ξ, the latter one does not. Both methods are based on a transformation of the powers x_i^r of the data x_i into new (artificial) data, $\mu_r(x_i)$ for QS and $t_r(x_i)$ for CS.

The first issue of this paper is to explore some, up to now unknown, properties of the variables μ_r and t_r and to reveal a peculiar duality between them. Another issue is to transform the formulas for the ACMs and their small-σ_δ approximations so that they become easier to compute, possibly with the help of a matrix oriented programming language. In particular, they should be written in terms of the observable variable x instead of the unobservable ξ. An important point in this respect is the evaluation of the terms in the ACM of QS that stem from the estimation of the nuisance parameters μ_ξ and σ_ξ^2. Contrary to what one might conclude from the original form of the ACM in Kukush *et al.* (2005b), it turns out that these additional terms can be computed without any integration (although integration remains necessary to compute the main term of the ACM formula).

Shklyar *et al.* (2005) have studied a simplified version of the QS estimator, the so-called *simple score* (SS) *estimator*. Two equivalent formulas for its ACM are presented. The ACM formula has the same term originating from the estimation of the nuisance parameters as the ACM of QS.

If this term is ignored (i.e., if the nuisance parameters are taken to be known), the difference of the ACMs of the CS and SS estimators is p.s.d., cf. Shklyar *et al.* (2005). It is an open question whether this is still true if the nuisance parameters have to be estimated.

As to the log-linear Poisson measurement error model, there is no need to repeat the ACM formula for the CS estimator, which is well documented in Shklyar and Schneeweiss (2005). The ACM of the QS estimator can only be given in an implicit form (i.e., as an integral). The SS estimator of Shklyar and Schneeweiss (2005) has been derived via an ad hoc approach. There is, however a more general model from which an SS estimator can be developed. This has been done by Kukush *et al.* (2005a). Therefore the SS estimator for the Poisson model is now constructed on the basis of this last paper, and its ACM is derived. Finally, the contribution of the estimation of the nuisance parameters to the ACM of QS and SS is found. It is shown that it is the same for QS and SS and, indeed, for a general class of structural estimators, just as in the case of the polynomial measurement error model.

In Section 2, the variables μ_r and t_r are investigated. More results on derivatives of the μ_r are found in Section 3. Section 4 deals with the ACM of the QS estimator in the polynomial model and in particular with the terms resulting from estimating the nuisance parameters. Section 5 has a reformulation of the ACM of the CS estimator, and Section 6 deals with the SS estimator. Section 7 discusses efficiency problems. Some new results for the Poisson model are found in Section 8. Section 9 has some concluding remarks.

2 QS and CS: The variables μ_r and t_r

The *QS estimator* $\hat{\beta}_Q$ of the polynomial measurement error model is based on the quasi score function

$$\psi_Q(y, x, \beta) = (y - \mu^\top \beta) v^{-1} \mu,$$

where $\mu := \mathbb{E}(\zeta|x) =: (\mu_0, \mu_1, \cdots, \mu_k)^\top$ and $v := \mathbb{V}(y|x)$. The elements of the conditional mean vector μ, $\mu_r = \mathbb{E}(\xi^r|x)$, are polynomials in x of degree r. $\mu_0 = 1$ and $\mu_1 = \mu_1(x) = \mathbb{E}(\xi|x)$ is given by

$$\mu_1 = \frac{\sigma_\delta^2}{\sigma_x^2}\mu_x + \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right) x. \tag{1}$$

The other μ_r are polynomials of μ_1 of degree r, c.f. Thamerus (1998):

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} \mu_j^* \mu_1^{r-j} \tag{2}$$

with

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

$$\tau^2 := \mathbb{V}(\xi|x) = \sigma_\delta^2 \left(1 - \frac{\sigma_\delta^2}{\sigma_x^2}\right), \tag{4}$$

where $(j-1)!!$ is short for $1 \cdot 3 \cdot 5 \cdots (j-1)$ and $(-1)!! = 1$. The conditional variance v is given by

$$v = \sigma_\epsilon^2 + \beta^\top \left(M - \mu\mu^\top\right)\beta, \tag{5}$$

where $M = M(x)$ is a $(k+1) \times (k+1)$-matrix with elements $M_{rs} = \mu_{r+s}$, $r, s = 0, \cdots, k$. Note that the $\mu_r(x_i)$ can be computed from the data x_i if the nuisance parameters μ_x and σ_x^2 are given. Typically they are unknown and must be estimated from the data x_i in the usual way.

The *CS estimator* $\widehat{\beta}_C$ is based on the corrected score function

$$\psi_C(y, x, \beta) = yt - T\beta,$$

where $t = t(x)$ is such that $\mathbb{E}(t|\xi) = \zeta$. Thus $t = (t_0, t_1, \ldots, t_k)^\top$ and $\mathbb{E}(t_r|\xi) = \xi^r$. $T = T(x)$ is a $(k+1) \times (k+1)$-matrix with elements $T_{rs} = t_{r+s}$. The t_r are polynomials in x of degree r. They can be computed via the recursion formula, cf. Stefanski (1989) and Cheng and Schneeweiss (1998),

$$t_{r+1} = t_r x - r t_{r-1}\sigma_\delta^2; \quad t_0 = 1, \quad t_{-1} = 0. \tag{6}$$

Note the duality in the definitions of μ and t:

$$\mu = \mathbb{E}(\zeta|x), \quad \mathbb{E}(t|\xi) = \zeta$$

and also in the matrices M and T:

$$M = \mathbb{E}(\zeta\zeta^T|x), \quad \mathbb{E}(T|\xi) = \zeta\zeta^\top.$$

This duality reaches farther. It turns out that, although the defining formulas (2) and (6) for μ and t, respectively, are quite different, there are other ways of computing μ and t, which very much resemble (2) and (6), but with the role of μ and t interchanged.

Proposition 1
The variables μ_r can be computed via the recursion formula

$$\mu_{r+1} = \mu_r\mu_1 + r\mu_{r-1}\tau^2, \quad \mu_0 = 1, \quad \mu_{-1} = 0. \tag{7}$$

Proof: According to (2)

$$\begin{aligned}\mu_{r+1} &= \sum_{j=0}^{r+1}\binom{r+1}{j}\mu_1^{r+1-j}\mu_j^* \\ &= \sum_{j=1}^{r+1}\binom{r}{j-1}\mu_1^{r+1-j}\mu_j^* + \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r+1-j}\mu_j^* \\ &= \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r-j}\mu_{j+1}^* + \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r+1-j}\mu_j^*.\end{aligned}$$

In the second equation we used the identity

$$\binom{r+1}{j} = \binom{r}{j-1} + \binom{r}{j}, \quad 1 \le j \le r.$$

Now again by (2), the r.h.s. of the recursion formula (7) is

$$\begin{aligned}&\sum_{j=0}^{r}\binom{r}{j}\mu_1^{r+1-j}\mu_j^* + r\sum_{j=0}^{r-1}\binom{r-1}{j}\mu_1^{r-1-j}\mu_j^*\tau^2 \\ =& \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r+1-j}\mu_j^* + \sum_{j=0}^{r-1}\binom{r}{j+1}\mu_1^{r-1-j}\mu_{j+2}^* \\ =& \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r+1-j}\mu_j^* + \sum_{j=0}^{r}\binom{r}{j}\mu_1^{r-j}\mu_{j+1}^* = \mu_{r+1}.\end{aligned}$$

In the second equation the identity,

$$(j+1)\mu_j^*\tau^2 = \mu_{j+2}^*,$$

see (3), was used and in the third equation the fact that $\mu_1^* = 0$.♦

Remark: The proof is similar to the proof of (6) as given in Cheng and Schneeweiss (1996).

Proposition 2
t_r can be computed via the closed form formula

$$t_r = \sum_{j=0}^{r}\binom{r}{j}\mu_j^+ x^{r-j} \tag{8}$$

$$\mu_j^+ := \begin{cases} 0 & \text{if} \quad j \text{ is odd} \\ (j-1)!!(-1)^{\frac{j}{2}}\sigma_\delta^j & \text{if} \quad j \text{ is even.}\end{cases}$$

Proof: If we replace μ_r, μ_1, and τ^j with t_r, x and $(-1)^{\frac{j}{2}}\sigma_\delta^j$, respectively, (7) changes to (6) and (2) changes to (8). By Proposition 1, (7) follows from (2), and so (6) follows from (8). But as (6) defines the t_r uniquely, the t_r defined by (6) must be the same as those defined by (8).♦

The great similarity in the construction of the variables μ_r and t_r can also be seen by looking at its values, e.g.:

$$\mu_1 = \mu_1, \quad \mu_2 = \mu_1^2 + \tau^2, \quad \mu_3 = \mu_1^3 + 3\tau^2\mu_1, \quad \mu_4 = \mu_1^4 + 6\tau^2\mu_1^2 + 3\tau^4$$

and

$$t_1 = x, \quad t_2 = x^2 - \sigma_\delta^2, \quad t_3 = x^3 - 3\sigma_\delta^2 x, \quad t_4 = x^4 - 6\sigma_\delta^2 x^2 + 3\sigma_\delta^4.$$

3 Derivatives of μ_r

By (2) and (3) μ_r is a function of μ_1 and τ^2. We can derive formulas for the derivatives of μ_r with respect to μ_1 and τ^2, which will be usefull later on.

Proposition 3

$$\frac{\partial \mu_r}{\partial \mu_1} = r\mu_{r-1}, \quad r \geq 1 \tag{9}$$

$$\frac{\partial \mu_r}{\partial \tau^2} = \binom{r}{2}\mu_{r-2}, \quad r \geq 2. \tag{10}$$

Proof: Instead of (9), we will prove the stronger proposition

$$\mu_r = r\int_0^{\mu_1} \mu_{r-1} d\mu_1 + \mu_r^*.$$

Indeed, by (2) the r.h.s of this equation equals

$$\begin{aligned}
& r \int_0^{\mu_1} \sum_{j=0}^{r-1} \binom{r-1}{j} \mu_j^* \mu_1^{r-1-j} d\mu_1 + \mu_r^* \\
= & r \sum_{j=0}^{r-1} \binom{r-1}{j} \mu_j^* \frac{\mu_1^{r-j}}{r-j} + \mu_r^* \\
= & \sum_{j=0}^{r-1} \binom{r}{j} \mu_j^* \mu_1^{r-j} + \mu_r^* \\
= & \sum_{j=0}^{r} \binom{r}{j} \mu_j^* \mu_1^{r-j},
\end{aligned}$$

which is equal to μ_r by (2).

To prove (10), first note that by (3), for j even, $j \geq 2$,

$$\begin{aligned}\frac{\partial \mu_j^*}{\partial \tau^2} &= (j-1)!!\frac{j}{2}\tau^{j-2} \\ &= \left(\begin{array}{c} j \\ 2 \end{array}\right)(j-3)!!\tau^{j-2} = \left(\begin{array}{c} j \\ 2 \end{array}\right)\mu_{j-2}^*.\end{aligned}$$

Now from (2) and the previous equation, for $r \geq 2$,

$$\begin{aligned}\frac{\partial \mu_r}{\partial \tau^2} &= \sum_{j=2}^{r}\left(\begin{array}{c} r \\ j \end{array}\right)\left(\begin{array}{c} j \\ 2 \end{array}\right)\mu_{j-2}^*\mu_1^{r-j} \\ &= \frac{r(r-1)}{2}\sum_{j=2}^{r}\left(\begin{array}{c} r-2 \\ j-2 \end{array}\right)\mu_{j-2}^*\mu_1^{r-j} \\ &= \left(\begin{array}{c} r \\ 2 \end{array}\right)\sum_{j=0}^{r-2}\left(\begin{array}{c} r-2 \\ j \end{array}\right)\mu_j^*\mu_1^{r-2-j} = \left(\begin{array}{c} r \\ 2 \end{array}\right)\mu_{r-2}. \blacklozenge\end{aligned}$$

By stacking the formulas (9) and (10), respectively, for $r = 0, \ldots, k$, we can now give corresponding expressions for the vector μ. We introduce the $(k+1) \times (k+1)$ triangular band matrices

$$D_1 := \left(\begin{array}{cccccc} 0 & & & & & \\ 1 & 0 & & & & \\ & 2 & 0 & & & \\ & & \cdots & & & \\ & & & \cdots & & \\ & & & & k & 0 \end{array}\right)$$

$$D_2 := \left(\begin{array}{cccccc} 0 & & & & & \\ 0 & 0 & & & & \\ \left(\begin{array}{c} 2 \\ 2 \end{array}\right) & 0 & 0 & & & \\ & \left(\begin{array}{c} 3 \\ 2 \end{array}\right) & 0 & 0 & & \\ & & & \cdots & & \\ & & & & \cdots & \\ & & & \left(\begin{array}{c} k \\ 2 \end{array}\right) & 0 & 0 \end{array}\right) \quad (11)$$

and note that

$$D_2 = \frac{1}{2}D_1^2. \quad (12)$$

Proposition 3 then, translates immediately into.

Proposition 4

$$\begin{aligned}\frac{\partial\mu}{\partial\mu_1} &= D_1\mu &(13)\\ \frac{\partial\mu}{\partial\tau^2} &= D_2\mu. &(14)\end{aligned}$$

Finally we also have

Proposition 5

$$\mu_1\frac{\partial\mu}{\partial\mu_1} = (D-\tau^2 D_1^2)\mu \qquad (15)$$

with $D := \mathrm{diag}(0,1,2,\ldots,k)$.

Proof: First note that by Proposition 1

$$\mu_1\mu = \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \vdots\\ \mu_{k+1}\end{pmatrix} - \tau^2\begin{pmatrix}0\\ \mu_0\\ 2\mu_1\\ \vdots\\ k\mu_{k-1}\end{pmatrix}.$$

The last vector equals $D_1\mu$, and the first vector on the r.h.s multiplied by D_1 equals $D\mu$. Therefore

$$\mu_1\frac{\partial\mu}{\partial\mu_1} = D_1\mu_1\mu = D\mu - \tau^2 D_1^2\mu.\blacklozenge$$

4 The ACM of QS

According to Kukush *et al.* (2005b), the ACM of $\widehat{\beta}_Q$ is given by

$$\Sigma_Q = (\mathbb{E}v^{-1}\mu\mu^\top)^{-1} + (\mathbb{E}v^{-1}\mu\mu^\top)^{-1}(\sigma_x^2 F_1F_1^\top + \frac{2}{\sigma_x^4}F_2F_2^\top)(\mathbb{E}v^{-1}\mu\mu^\top)^{-1} \qquad (16)$$

where

$$F_p = \mathbb{E}v^{-1}\mu\frac{\partial\mu^\top}{\partial\gamma_p}\beta, \quad p=1,2, \quad \gamma_1=\mu_x, \quad \gamma_2=\frac{1}{\sigma_x^2}.$$

The F-terms stem from the estimation of the nuisance parameters. The purpose of this section is to evaluate these terms so that they become computationally more accessible. It turns out that it is not necessary to compute the expected value as prescribed in the definition of F_p.

Proposition 6
The ACM of $\widehat{\beta}_Q$ equals

$$\Sigma_Q = (\mathbb{E}v^{-1}\mu\mu^\top)^{-1} + F, \tag{17}$$

where

$$\begin{aligned} F &= \sigma_\delta^4(G_1^\top\beta\beta^\top G_1 + 2G_2^\top\beta\beta^\top G_2) \\ G_1 &= \frac{1}{\sigma_x}D_1 \\ G_2 &= \frac{1}{\sigma_x^2 - \sigma_\delta^2}(\mu_X D_1 - D + \tau^2 D_2). \end{aligned}$$

Proof: As μ is a function of μ_1 and τ^2, we have

$$\frac{\partial\mu}{\partial\gamma_p} = \frac{\partial\mu}{\partial\mu_1}\frac{\partial\mu_1}{\partial\gamma_p} + \frac{\partial\mu}{\partial\tau^2}\frac{\partial\tau^2}{\partial\gamma_p}, \quad p = 1, 2.$$

For $p = 1$ and $p = 2$, we find because of (1) and (4)

$$\begin{aligned} \frac{\partial\mu}{\partial\gamma_1} &= \frac{\partial\mu}{\partial\mu_1}\frac{\sigma_\delta^2}{\sigma_x^2} \\ \frac{\partial\mu}{\partial\gamma_2} &= \left[\frac{\partial\mu}{\partial\mu_1}(\mu_x - x) - \frac{\partial\mu}{\partial\tau^2}\sigma_\delta^2\right]\sigma_\delta^2. \end{aligned}$$

With

$$\mu_x - x = \frac{\sigma_x^2}{\sigma_x^2 - \sigma_\delta^2}(\mu_x - \mu_1),$$

which follows from (1), the latter becomes

$$\frac{\partial\mu}{\partial\gamma_2} = \sigma_\delta^2\left[\frac{\partial\mu}{\partial\mu_1}\frac{\sigma_x^2}{\sigma_x^2 - \sigma_\delta^2}(\mu_x - \mu_1) - \frac{\partial\mu}{\partial\tau^2}\sigma_\delta^2\right].$$

Finally, by (13) to (15),

$$\begin{aligned} \frac{\partial\mu}{\partial\gamma_1} &= \frac{\sigma_\delta^2}{\sigma_x^2}D_1\mu = \frac{\sigma_\delta^2}{\sigma_x}G_1\mu \\ \frac{\partial\mu}{\partial\gamma_2} &= \sigma_\delta^2\left[\frac{\sigma_x^2}{\sigma_x^2 - \sigma_\delta^2}(\mu_x D_1 - D + \tau^2 D_1^2)\mu - \sigma_\delta^2 D_2\mu\right]. \end{aligned}$$

Because of (12) and (4), the latter becomes

$$\frac{\partial \mu}{\partial \gamma_2} = \sigma_\delta^2 \frac{\sigma_x^2}{\sigma_x^2 - \sigma_\delta^2}(\mu_x D_1 - D + \tau^2 D_2)\mu = \sigma_\delta^2 \sigma_x^2 G_2 \mu.$$

We thus have

$$\begin{aligned} F_1 &= \frac{\sigma_\delta^2}{\sigma_x} \mathbb{E} v^{-1} \mu\mu^\top G_1^\top \beta \\ F_2 &= \sigma_\delta^2 \sigma_x^2 \mathbb{E} v^{-1} \mu\mu^\top G_2^\top \beta. \end{aligned}$$

By substituting F_1 and F_2 in (16) we finally obtain (17).♦

For $k = 2$ the two matrices G_1 and G_2 are, respectively,

$$G_1 = \frac{1}{\sigma_x} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix} \tag{18}$$

$$G_2 = \frac{1}{\sigma_x^2 - \sigma_\delta^2} \begin{pmatrix} 0 & 0 & 0 \\ \mu_x & -1 & 0 \\ \tau^2 & 2\mu_x & -2 \end{pmatrix}. \tag{19}$$

For small σ_δ^2 an approximation to Σ_Q can be derived. The general formula in Kukush and Schneeweiss (2005) can be specialized to the polynomial case and yields

$$\begin{aligned} \Sigma_Q &= \sigma_\epsilon^2 (\mathbb{E}Z)^{-1} \\ &\quad + \sigma_\delta^2 (\mathbb{E}Z)^{-1} \mathbb{E} \left\{ \left(\frac{\partial z^\top}{\partial x} \beta \right)^2 Z + \sigma_\epsilon^2 \left(\frac{1}{2} \frac{\partial^2 Z}{\partial x^2} + \frac{\partial z}{\partial x} \frac{\partial z^\top}{\partial x} \right) \right\} (\mathbb{E}Z)^{-1} \\ &\quad + O(\sigma_\delta^4), \end{aligned}$$

where $z := (1, x, \ldots, x^k)^\top$ and $Z := zz^\top$. By noting that

$$\frac{\partial z}{\partial x} = D_1 z, \quad \frac{\partial^2 Z}{\partial x^2} = D_1^2 Z + 2 D_1 Z D_1^\top + Z D_1^{\top 2},$$

this can be written as

$$\begin{aligned} \Sigma_Q &= \sigma_\epsilon^2 (\mathbb{E}Z)^{-1} \\ &+ \sigma_\delta^2 (\mathbb{E}Z)^{-1} \mathbb{E}\{(\beta^\top D_1 Z D_1^\top \beta) Z + \sigma_\epsilon^2 (D_2 Z + Z D_2^\top + 2 D_1 Z D_1^\top)\} (\mathbb{E}Z)^{-1} \\ &+ O(\sigma_\delta^4). \end{aligned} \tag{20}$$

It may be noted that, contrary to (17), the expectations involved simply yield moments of x and are therefore easy to compute.

From Kukush *et al.* (2005b) a similar formula can be derived, which however is stated in terms of ξ rather than x. Both formulas differ in value but the difference is of the order σ_δ^4.

5 The ACM of CS

In Kukush *et al.* (2005b) a formula for the ACM of $\widehat{\beta}_C$ has been derived:

$$\Sigma_C = (\mathbb{E}\zeta\zeta^\top)^{-1}\{\sigma_\epsilon^2\mathbb{E}tt^\top + \mathbb{E}(T - t\zeta^\top)\beta\beta^\top(T - \zeta t^\top)\}(\mathbb{E}\zeta\zeta^\top)^{-1}. \qquad (21)$$

This is a hybrid formula in so far as t and T are functions of x, whereas ζ is a function of ξ. With (5) and with the help of the identity

$$\begin{aligned} & \mathbb{E}[(T - t\zeta^\top)\beta\beta^\top(T - \zeta t^\top)|x] \\ = \; & T\beta\beta^\top T - t\mu^\top\beta\beta^\top T - T\beta\beta^\top\mu t^\top + t\beta^\top M\beta t^\top \\ = \; & (T - t\mu^\top)\beta\beta^\top(T - \mu t^\top) + t\beta^\top(M - \mu\mu^\top)\beta t^\top, \end{aligned}$$

(21) can be written as

$$\Sigma_C = (\mathbb{E}T)^{-1}\mathbb{E}\{(T - t\mu^\top)\beta\beta^\top(T - \mu t^\top) + vtt^\top\}(\mathbb{E}T)^{-1}. \qquad (22)$$

Again only moments of x are needed in order to compute the ACM of $\widehat{\beta}_C$. We have several options to evaluate $\mathbb{E}T$ because, cf. Shklyar *et al.* (2005),

$$\mathbb{E}T = \mathbb{E}M = \mathbb{E}t\mu^\top = \mathbb{E}\zeta\zeta^\top.$$

In passing, it might be worthwile to mention the ACM of the *naive* (N) *estimator* $\hat{\beta}_N := (\sum_1^n z_i z_i^\top)^{-1}\sum_1^n z_i y_i$. A hybrid formula for its ACM is given in Kukush *et al.* (2005b). It can be "improved" to a formula that is based on the observed variables x_i solely:

$$\Sigma_N = (\mathbb{E}Z)^{-1}\mathbb{E}vZ(\mathbb{E}Z)^{-1}.$$

6 SS and its ACM

Another structural estimator can be constructed as a simplified version of QS. It is called *simple score* (SS) *estimator* and is based on the simplified score function

$$\psi_S(y,x,\beta) = (y - \mu^\top\beta)t.$$

An equivalent score function for SS is

$$\psi_S^*(y,x,\beta) = (y - \mu^\top\beta)\mu,$$

cf. Shklyar *et al.* (2005), which differs from ψ_Q just by the omission of the factor v^{-1}.

The merit of the SS estimator is that it is much simpler to compute than the QS estimator. It is, however, (slightly) less efficient than the latter, but it is still more efficient than the CS estimator as long as μ_ξ and σ_ξ^2 are known and need not be estimated, see Section 7. It serves as an intermediate estimator between QS and CS and is useful if one wants to compare the relative efficiencies of the latter two.

The ACM of the SS estimator is given by two equivalent formulas depending on whether it is derived from ψ_S or ψ_S^*:

$$\begin{aligned}\Sigma_S &= (\mathbb{E}T)^{-1}\mathbb{E}vtt^\top(\mathbb{E}T)^{-1} + F \\ &= (\mathbb{E}\mu\mu^\top)^{-1}Ev\mu\mu^\top(\mathbb{E}\mu\mu^\top)^{-1} + F,\end{aligned} \tag{23}$$

where F is the same as in (17).

The first formula (23) is implicitly given in Shklyar *et al.* (2005), the second one follows in a similar way from ψ_S^*. Their equivalence can be directly seen by noting that $t = K\mu$ with some nonsingular matrix K and that $\mathbb{E}(\mu t^\top) = \mathbb{E}\zeta\zeta^\top = \mathbb{E}T$, cf. Shklyar *et al.* (2005).

7 Efficiency comparison

One can show that $\Sigma_Q \le \Sigma_S$, cf. Shklyar *et al.*(2005). Indeed, since the term F in (17) and (23) is the same, one needs only to compare the first terms in (17) and (23), respectively, and for this comparison one can use the Cauchy-Schwartz inequality.

These arguments do not hold for an efficiency comparison of CS and SS. The difference of their ACMs is

$$\Sigma_C - \Sigma_S = (\mathbb{E}T)^{-1}\mathbb{E}(T - t\mu^\top)\beta\beta^\top(T - \mu t^\top)(\mathbb{E}T)^{-1} - F. \tag{24}$$

It is not clear at the outset whether this difference is always ≥ 0. (It is, of course, ≥ 0 and, indeed, even > 0 if the last term vanishes, which occurs when the nuisance parameters need not be estimated: $\Sigma_C > \Sigma_S$ if μ_ξ and σ_ξ^2 are both known, cf. Shklyar *et al.* (2005)).

There are cases where $\Sigma_C - \Sigma_S$ is singular if nuisance parameters are present. Consider a quadratic model with $\beta_1 = 0$ and $\mu_\xi = 0$ (β_0 plays no role). Then a detailed algebraic calculation shows that $\det(\Sigma_C - \Sigma_S) = 0$. On the other hand, all the diagonal elements of $\Sigma_C - \Sigma_S$ have positive leading terms and thus tend to ∞ for $\sigma_\delta^2 \to \infty$.

8 The Poisson model

8.1 CS, QS and SS

The log-linear Poisson model with measurement errors is given by a response variable y which is Poisson distributed with a parameter λ that is a log-linear function of a random vector ξ (cf. Shklyar and Schneeweiss, 2005).

$$\begin{aligned} y|\xi &\sim Po(\lambda) \\ \lambda &= \exp(\beta_0 + \beta_1^\top \xi), \end{aligned}$$

$\xi = (\xi_1, \ldots, \xi_p)^\top$. The compound vector $\beta := (\beta_0, \beta_1^\top)^\top$ is the parameter of interest. In addition, there are nuisance parameters γ characterizing the dirstribution of ξ. Here it is assumed that $\xi \sim N(\mu_\xi, \Sigma_\xi)$. Finally, as in the polynomial model, ξ is latent. Instead $x = (x_1, \ldots, x_p)^\top$ is observed with a measurement error vector δ:

$$x = \xi + \delta,$$

where $\delta \sim N(0, \Sigma_\delta)$, δ independent of ξ and y, and Σ_δ is assumed to be known.

The likelihood score function for β in the error free model is given by

$$\psi^*(y, \xi, \beta) = (y - \lambda)(1, \xi^\top)^\top.$$

The corrected score function $\psi_C(y, x, \beta)$, which is the basis for the *corrected score* (CS) *estimator*, is constructed such that $\mathbb{E}(\psi_C|y, \xi) = \psi^*$ and is given by

$$\psi_C(y, x, \beta) = \begin{pmatrix} y - e \\ yx - e(x - \Sigma_\delta \beta_1) \end{pmatrix},$$

where $e = \exp(\beta_0 + \beta_1^\top x - \frac{1}{2}\beta_1^\top \Sigma_\delta \beta_1)$. The CS estimator constructed from an i.i.d. sample (y_i, x_i), $\quad i = 1, \ldots, n$, is the solution to

$$\sum_{i=1}^{n} \psi_C(y_i, x_i, \widehat{\beta}_C) = 0.$$

The ACM of $\widehat{\beta}_C$ can be found in Shklyar and Schneeweiss (2005), equation (22),albeit with a different notation.

For the *quasi score* (QS) *estimator* $\widehat{\beta}_Q$ we need to know the conditional expectation and variance of y given ξ. The conditional distribution of x given ξ is

$$x|\xi \sim N(\mu(x), \top)$$

with

$$\begin{aligned} \mu(x) &= \Sigma_\delta \Sigma_x^{-1}(\mu_x - x) + x && (25)\\ \top &= \Sigma_\delta - \Sigma_\delta \Sigma_x^{-1} \Sigma_\delta \quad = \quad \Sigma_\xi - \Sigma_\xi \Sigma_x^{-1} \Sigma_\xi, && (26) \end{aligned}$$

cf. Shklyar and Schneeweiss (2005). (Note that the $\top$ of this section is different from the T of the preceding sections and is not to be mixed up with the transposition sign; $\top = \tau^2$ if $p = 1$. Similarly the vector $\mu(x)$ should not be confused with the vector μ of the preceding sections; it is equal to $\mu_1 = \mu_1(x)$ if $p = 1$, see(1)). Therefore,

$$\begin{aligned} \mathbb{E}(y|x) &=: m(x, \beta) = \exp\{\beta_0 + \beta_1^\top \mu(x) + \frac{1}{2}\beta_1^\top \top \beta_1\} && (27)\\ V(y|x) &=: v(x, \beta) = m(x, \beta) + \{\exp(\beta_1^\top \top \beta_1) - 1\} m^2(x, \beta). \end{aligned}$$

The quasi score function then is

$$\psi_Q(y, x, \beta) = (y - m) v^{-1} \frac{\partial m}{\partial \beta},$$

where

$$\frac{\partial m}{\partial \beta} = m \begin{pmatrix} 1 \\ \mu(x) + \top \beta_1 \end{pmatrix} =: mg, \qquad (28)$$

and the QS estimator is the solution to

$$\sum_{i=1}^{n} \psi_Q(y_i, x_i, \widehat{\beta}_Q) = 0.$$

The ACM of $\widehat{\beta}_Q$ is given by

$$\Sigma_Q = (\mathbb{E}v^{-1}m^2 gg^\top)^{-1}. \tag{29}$$

Here it is assumed that the nuisance parameters μ_x and Σ_x are given and known to the statistician. The case of unknown nuisance parameters is treated in the next subsection.

One can also construct a *simplified score* (SS) *estimator* $\widehat{\beta}_S$ which is based on the simplified quasi score function.

$$\psi_S(y, x, \beta) = (y - m)(1, x)^\top.$$

This score function is derived from Kukush *et al.* (2005a). It differs from (but is actually equivalent to) another simplified score function, which is given in Shklyar and Schneeweiss (2005, equation (27)). Under known nuisance parameters, the ACM of $\widehat{\beta}_S$ can be computed from the sandwich formula

$$\Sigma_S = A_S^{-1} B_S A_S^{-\top}, \tag{30}$$

where

$$\begin{aligned} A_S &= -\mathbb{E}\frac{\partial \psi_S}{\partial \beta^\top} = \mathbb{E}m(1, x^\top)^\top g^\top \\ B_S &= \mathbb{E}\psi_S \psi_S^\top = \mathbb{E}v(1, x^\top)^\top (1, x^\top). \end{aligned}$$

By arguments similar to those of Shklyar and Schneeweiss (2005)[1] one can evaluate A_S and B_S and thus Σ_S.

Proposition 7
If μ_x and Σ_x are known, the ACM of the SS estimator is given by (30) with

$$A_S = q \begin{pmatrix} 1 & b^\top \\ b & bb^\top + \Sigma_\xi \end{pmatrix} \tag{31}$$

$$B_S = q \begin{pmatrix} 1 & b^\top \\ b & bb^\top + \Sigma_x \end{pmatrix} + q^* \begin{pmatrix} 1 & b^{*\top} \\ b^* & b^* b^{*\top} + \Sigma_x \end{pmatrix}, \tag{32}$$

where

$$\begin{aligned} q &:= \exp(\beta_0 + \beta_1^\top \mu_x + \frac{1}{2}\beta_1^\top \Sigma_\xi \beta_1) \\ q^* &:= [1 - \exp(-\beta_1^\top \top \beta_1)] \exp[2(\beta_0 + \beta_1^\top \mu_x + \beta_1^\top \Sigma_\xi \beta_1)] \\ b &:= \mu_x + \Sigma_\xi \beta_1 \\ b^* &:= \mu_x + 2\Sigma_\xi \beta_1. \end{aligned}$$

[1]There is a mistake in Corollary 3 of that paper: In (38), the term $(\Sigma_x + 2\Sigma_x \Sigma_w^{-1} \Sigma_x)$ must be replaced with $(\Sigma_x + \Sigma_x \Sigma_w^{-1} \Sigma_x)$.

It can be shown, cf. Shklyar and Schneeweiss (2005) and Kukush *et al.* (2005a), that

$$\Sigma_Q \leq \Sigma_S \leq \Sigma_C.$$

Note, however, that the equations for Σ_Q and Σ_S are only valid under the assumption of known nuisance parameters.

8.2 Nuisance parameters

The "structural" estimators $\widehat{\beta}_Q$ and $\widehat{\beta}_S$ of the previous section have been constructed assuming the (nuisance) parameters γ characterizing the distribution of ξ to be known. We now drop this assumption. Instead we assume that γ can be estimated from the observed data $x_i,\ 1,\ldots,n$, solely, without the necessity to resort to the model and to the data y_i. Under our assumption that $\xi \sim N(\mu_\xi, \Sigma_\xi)$ and consequently $x \sim N(\mu_x, \Sigma_x)$, μ_ξ and Σ_ξ or, equivalently, μ_x and Σ_x can be easily estimated by the corresponding sample moments. We take as γ the vector composed of μ_x and $w := vech(\Sigma_x^{-1})$, i.e., $\gamma = (\mu_x, w^\top)^\top$, which is just a reparameterization of (μ_x, Σ_x).

The regression parameter vector β is then estimated by using a (structural) score function like ψ_Q or ψ_S, where the nuisance parameter vector γ has been substituted by its estimate $\widehat{\gamma}$. The resulting estimator is still consistent. But the formula for its ACM has to be augmented by a term stemming from the estimation of γ.

From a general point of view, assume that β is estimated on the basis of some general estimating function $\psi := \psi(y, x, \beta, \gamma)$, where the nuisance parameter γ has been estimated in advance from the data $x_i,\ i = 1,\ldots,n$. Then the ACM of $\widehat{\beta}$ is given by, cf. Shklyar *et al.* (2005),

$$\Sigma = A^{-1}BA^{-\top} + A^{-1}A_\gamma \Sigma_\gamma A_\gamma^{-\top} A^{-\top} =: \Sigma^\circ + F, \tag{33}$$

where $A = -\mathbb{E}\frac{\partial\psi}{\partial\beta^\top}$, $\quad B = \mathbb{E}\psi\psi^\top$, $\quad A_\gamma = -\mathbb{E}\frac{\partial\psi}{\partial\gamma^\top}$ and Σ_γ is the ACM of $\widehat{\gamma}$. $\Sigma^\circ := A^{-1}BA^{-\top}$ is the ACM of β if γ is known, just as in the previous section. The matrix F, which is due to the estimation of the nuisance parameters, corresponds to the matrix F of the polynomial model, see (17) and (23), but is different from this F.

To be more specific, let the estimating function ψ be of the form

$$\psi(y, x, \beta, \gamma) = (y - m)a, \tag{34}$$

where $a := a(x, \beta, \gamma)$ is a known vector-valued function that specifies the estimation procedure. For QS, $a = v^{-1}\frac{\partial m}{\partial \beta}$, and for SS, $a = (1, x^\top)^\top$. Note that m is now also a function of γ, i.e., $m := m(y, x, \beta, \gamma)$.

Proposition 8
For the Poisson model, F is independent of a and thus independent of the estimation procedure chosen. In particular, F is the same for QS and SS.

This property has been proved for the polynomial measurement error model in Shklyar *et al.* (2005) – see also (17) and (23) – but not for the Poisson model, where F takes a different form.

Proof: To evaluate F for the Poisson model, first note that with the estimation function (34)

$$A = \mathbb{E}a\frac{\partial m}{\partial \beta^\top} = \mathbb{E}mag^\top,$$

where g comes from (28). Similarly,

$$A_\gamma = \mathbb{E}a\frac{\partial m}{\partial \gamma^\top}.$$

Now, by (27) and (25),

$$\frac{\partial m}{\partial \mu_x} = m\Sigma_x^{-1}\Sigma_\delta\beta_1 = m\Sigma_x^{-1}d, \tag{35}$$

where $d := \Sigma_\delta\beta_1$, and by (25) to (27),

$$\frac{\partial m}{\partial w} = m\frac{\partial vec^\top(\Sigma_x^{-1})}{\partial w}[vec\{d(\mu_x - x)^\top\} - \frac{1}{2}vec(dd^\top)],$$

see Dhyrmes (1984, Prop. 100) for the differentiation rule employed. With the abbreviation $D_w := \partial vec^\top(\Sigma_x^{-1})/\partial w$ and using the rule $vec(ab^\top) = b \otimes a$, the last expression can also be written as

$$\frac{\partial m}{\partial w} = mD_w[\{\mu_x - x - \frac{1}{2}d\} \otimes d].$$

D_w is a matrix of ones and zeros such that $D_w^\top vech(A) = vec(A)$ for any symmetric matrix A.

Now from (25),

$$\mu_x - x = \Sigma_x\Sigma_\xi^{-1}\{\mu_x - \mu(x)\}$$

and thus

$$\frac{\partial m}{\partial w} = mD_w[\{\Sigma_x\Sigma_\xi^{-1}\mu_x - \frac{1}{2}d - \Sigma_x\Sigma_\xi^{-1}\mu(x)\} \otimes d].$$

Together with (35) we thus have

$$\frac{\partial m}{\partial \gamma} = mh$$

with

$$h := \begin{pmatrix} \Sigma_x^{-1}d \\ D_w[\{\Sigma_x\Sigma_\xi^{-1}\mu_x - \frac{1}{2}d - \Sigma_x\Sigma_\xi^{-1}\mu(x)\} \otimes d] \end{pmatrix}.$$

Obviously, g and h are linearly related:

$$h = Gg$$

with a non-stochastic matrix G. With the help of the identity $\Sigma_x\Sigma_\xi^{-1}{}^\top\beta_1 = d$, see (26), one can verify that G is given by

$$G = \begin{pmatrix} \Sigma_x^{-1}d & 0 \\ D_w[(\Sigma_x\Sigma_\xi^{-1}\mu_x + \frac{1}{2}d) \otimes d] & -D_w[(\Sigma_x\Sigma_\xi^{-1}) \otimes d] \end{pmatrix}.$$

It follows that

$$A_\gamma = \mathbb{E}mah^\top = \mathbb{E}mag^\top G^\top = AG^\top$$

and, by (33),

$$F = G^\top\Sigma_\gamma G.$$

As G is independent of the estimation procedure a, the proposition is proved.♦

As a consequence of Proposition 8, $\Sigma_Q \leq \Sigma_S$ also holds true when nuisance parameters are present.

In the univariate case, where β_1 is a scalar – a case, which has been dealt with in Kukush *et al.* (2004) – the matrix G has a rather simple form. Note that in this case $w = \sigma_x^{-2}$ and thus $D_w = 1$. We have

$$G = \sigma_\delta^2\beta_1 \begin{pmatrix} \frac{1}{\sigma_x^2} & 0 \\ \frac{\sigma_x^2}{\sigma_x^2-\sigma_\delta^2}(\mu_x + \frac{1}{2}\tau^2\beta_1) & -\frac{\sigma_x^2}{\sigma_x^2-\sigma_\delta^2} \end{pmatrix}.$$

With

$$\Sigma_\gamma = \mathrm{diag}(\sigma_x^2, 2\sigma_x^{-4}),$$

we finally obtain

$$F = \sigma_\delta^4\beta_1^2 \left[\frac{1}{\sigma_x^2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^\top + \frac{2}{(\sigma_x^2 - \sigma_\delta^2)^2} \begin{pmatrix} \mu_x + \frac{1}{2}\beta_1\tau^2 \\ -1 \end{pmatrix} \begin{pmatrix} \mu_x + \frac{1}{2}\beta_1\tau^2 \\ -1 \end{pmatrix}^\top \right].$$

9 Conclusion

The ACMs of three estimators (CS, QS, and SS) have been studied for the polynomial as well as for the Poisson measurement error model. Some alternative formulas that are based solely on the observable variables have been presented. The ACMs of QS and SS (and also of other structural estimators) have a term that stems from the estimation of the nuisance parameters. This term has been evaluated for both models. In particular for the Poisson model, this term is the same for a large class of structural estimators, a result which has been found previously for the polynomial model, too.

The presence of this term in the ACMs of the QS and SS estimators diminishes the efficiency of QS and SS, which would be greater if the nuisance parameters were known. In particular for a polynomial model, the efficiency of SS is so much reduced that, in some cases and for some parameter combinations, it is not strongly higher than the efficiency of CS anymore (as it would be if the nuisance parameters were known).

In the polynomial model, the CS and QS estimators are constructed with the help of transformed variables $t_r(x_i)$ and $\mu_r(x_i)$, respectively. New formulas for the computation of these variables have been derived.

10 References

1. Cheng, C.-L. and Schneeweiss, H. (1996), *The Polynomial Regression with Errors in the Variables*, Discussion Paper 42, Sonderforschungsbereich 386, University of Munich.

2. Cheng, C.-L. and Schneeweiss, H. (1998), Polynomial regression with errors in the variables, *J. Roy. Stat. Soc.* B 60, 189-199.

3. 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.

4. Dhrymes, P. J. (1984), *Mathematics for Econometrics.* 2nd ed., Springer, New York.

5. Kukush, A. and Schneeweiss, H. (2005), Comparing different estimators in a nonlinear measurement error model I and II, *Mathematical Methods of Statistics* 14, 53-79 and 203-223.

6. Kukush, A., Schneeweiss, H., Wolf, R. (2004), Three estimators for the Poisson regression model with measurement errors, *Statistical Papers* 45, 351-368.

7. Kukush, A., Schneeweiß, H., Shklyar, S. (2005a): *Quasi Score is More Efficient than Corrected Score in a General Nonlinear Measurement Error Model*, Discussion Paper 451, Sonderforschungsbereich 386, University of Munich.

8. Kukush, A., Schneeweiss, H., Wolf, R. (2005b), Relative efficiency of three estimators in a polynomial regression with measurement errors, *Journ. Statistical Planning and Inference* 127,179-203.

9. Shklyar, S. and Schneeweiss, H. (2005), A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors, *Journal of Multivariate Analysis* 94, 250-270.

10. Shklyar, S., Schneeweiss, H., Kukush, A. (2005), *Quasi Score is more Efficient than Corrected Score in a Polynomial Measurement Error Model*, Discussion Paper 445, Sonderforschungsbereich 386, University of Munich.

11. 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.

12. 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.