



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Kukush, Schneeweiss, Wolf:

Three Estimators for the Poisson Regression Model with Measurement Errors

Sonderforschungsbereich 386, Paper 243 (2001)

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

Projektpartner



Three Estimators for the Poisson Regression Model with Measurement Errors

Alexander Kukush

Department of Mathematics, Kiev Taras Shevchenko University
Kiev, Ukraine

email: kuog@mechmat.univ.kiev.ua

Hans Schneeweiss

Department of Statistics, University of Munich
Munich, Germany

email: schneew@stat.uni-muenchen.de

Roland Wolf

Department of Statistics, University of Munich
Munich, Germany

email: wolf@stat.uni-muenchen.de

Abstract

We consider two consistent estimators for the parameters of the linear predictor in the Poisson regression model, where the covariate is measured with errors. The measurement errors are assumed to be normally distributed with known error variance σ_u^2 . The SQS estimator, based on a conditional mean-variance model, takes the distribution of the latent covariate into account, and this is here assumed to be a normal distribution. The CS estimator, based on a corrected score function, does not use the distribution of the latent

covariate. Nevertheless, for small σ_u^2 , both estimators have identical asymptotic covariance matrices up to the order of σ_u^2 . We also compare the consistent estimators to the naive estimator, which is based on replacing the latent covariate with its (erroneously) measured counterpart. The naive estimator is biased, but has a smaller covariance matrix than the consistent estimators (at least up to the order of σ_u^2).

Keywords: Poisson regression model, measurement errors, corrected score estimator, structural quasi score estimator, naive estimator

1 Introduction

In a Poisson regression model the response variable y is Poisson distributed with a parameter λ which depends on one or several covariates x_k , see, e.g., Cameron and Trivedi (1998) and Carroll et al (1995). Here we restrict our attention to the case of just one covariate x . The dependence on x is given by an exponential function: $\lambda = \lambda(x, \beta) = \exp(\beta_0 + \beta_1 x)$ with unknown parameter vector $\beta = (\beta_0, \beta_1)'$. As in a Poisson model mean and variance are equal, we have a mean-variance model, where mean and variance are both given by λ .

Given a sample of i.i.d. pairs $(x_i, y_i), i = 1, \dots, n$, the parameters β_0 and β_1 can be estimated by maximizing the log-likelihood function of β :

$$l(\beta) = \sum_1^n \{y_i \ln \lambda(x_i, \beta) - \lambda(x_i, \beta)\}, \quad (1)$$

which gives rise to the estimating equation

$$\sum_1^n \{y_i - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)\}(1, x_i)' = 0. \quad (2)$$

It can be solved by standard methods.

Here, however, we assume that the covariate x_i is contaminated with measurement errors. Instead of x_i we observe w_i such that

$$w_i = x_i + u_i,$$

where u_i is the measurement error, which is supposed to be independent of (x_i, y_i) and normally distributed:

$$u_i \sim N(0, \sigma_u^2).$$

Its variance σ_u^2 is assumed to be known.

As is usual in measurement error models, we distinguish between the functional and the structural variant of the model. Here we base our investigation on the structural variant and, in fact, we assume that

$$x_i \sim N(\mu_x, \sigma_x^2).$$

We also assume that the $(y_i, x_i, u_i), i = 1, \dots, n$, are i.i.d. Finally, for technical reasons, we assume that β is confined to an open bounded subset of \mathfrak{R}^2 . This assumption guarantees that the estimators below are well defined, at least for large n , and that they converge a.s. to certain theoretical values. In some cases it is possible to do without this assumption.

The model can be written as a mean-variance model in x . By conditioning on w a conditional mean-variance model in w is derived, from which a *structural quasi score* (SQS) estimator can be constructed, Carroll et al (1995), Thamerus (1998). The SQS estimator is consistent and asymptotically normal.

For the functional case, where no distributional assumptions for the x_i are made, we also have an estimator, the *corrected score* (CS) estimator, which is also consistent and asymptotically normal. The CS estimator is constructed by correcting the likelihood score function of the error-free model so that the measurement errors are taken into account, Stefanski (1989), Nakamura (1992). Of course, the CS estimator can be applied to the structural variant of the model, too, and then it can be compared to the SQS estimator. Stefanski and Carroll (1987) describe still another estimator, the conditional-score estimator, but we will not deal with this one; see however Carroll et al (1995), chapter 6.

The comparison is accomplished by looking at the asymptotic covariance matrices of the SQS and CS estimators. It turns out that it is difficult to compare the asymptotic covariance matrices for a large error variance σ_u^2 , but comparatively easy for small σ_u^2 . We do this by expanding the error covariance matrices up to the order of σ_u^2 omitting terms of higher order. Surprisingly, up to this order, there is no difference in the asymptotic covariance matrices of the SQS and CS estimators despite the fact that both estimators use different amounts of prior information.

We also include in our investigation the so-called naive estimator, which is found by maximizing the log-likelihood (1) with x_i replaced by w_i . The naive estimator is asymptotically biased. Nevertheless its asymptotic covariance matrix can be of some interest and is therefore compared to the other two covariance matrices, again for small σ_u^2 .

The present study shows some analogy to a similar investigation for the polynomial model with measurement errors, Kukush et al (2001). But the covariance matrices to be compared to each other are quite different. The SQS, CS, and naive estimators here correspond to the SLS, ALS, and OLS estimators, respectively, in that earlier paper.

The Poisson model with measurement errors has been extensively dealt with, e.g., in Carroll et al (1995). Thamerus (1997) studied a Poisson model in the structural case with heteroskedastic measurement errors and with a distribution of the x_i which consists of a mixture of normal distributions.

In Section 2 we introduce the naive estimator and study its asymptotic covariance matrix in particular for small σ_u^2 . We do the same for the CS and the SQS estimators in Sections 3 and 4, respectively. Section 5 contains the main result: a comparison of the three covariance matrices for small σ_u^2 . Section 6 contains some simulation results, which corroborate the theoretical findings, and Section 7 concludes.

2 The naive estimator

2.1 Estimator

The naive estimator $\hat{\beta}_N$ is the solution to the estimating equation

$$\frac{1}{n} \sum_1^n \{y_i - \exp(\hat{\beta}_{0N} + \hat{\beta}_{1N} w_i)\} (1, w_i)' = 0, \quad (3)$$

which follows from (2) by replacing x_i with w_i . The solution exists and is unique for large n , Wedderburn (1976). A simple way to solve (3) is the iteratively reweighted least squares method, e.g., Fahrmeir and Tutz (1994). If β^k is the approximated value of $\hat{\beta}$ in the k 'th iteration, then $\Delta\beta^k = \beta^{k+1} - \beta^k$ is found by regressing $\eta = y/z_k - z_k$ on $\xi_1 = z_k$ and $\xi_2 = wz_k$, where $z_k = \exp\{\frac{1}{2}(\beta_0^k + \beta_1^k w)\}$. The iteration stops when $\Delta\beta^k$ is sufficiently small, and then $\hat{\beta}_N = \beta^k + \Delta\beta^k$. A starting value β^0 may be found by regressing $\log(y + \varepsilon)$, with some small ε , on 1 and w .

2.2 Bias

In order to find the bias of $\hat{\beta}_N$ we solve the estimating equation after taking expectations:

$$\mathbf{E}\{y - \exp(b_0 + b_1 w)\} (1, w)' = 0. \quad (4)$$

According to the theory of estimating equations, $\hat{\beta}_N$, which is the solution to (3), converge a.s. to the solution to (4), which we denote by $b = (b_0, b_1)'$,

i.e.:

$$\lim_{n \rightarrow \infty} \hat{\beta}_N = b.$$

To evaluate (4), and other expectations that we shall meet in the sequel, we need the following well-known and very useful relations for $w \sim N(\mu_w, \sigma_w^2)$:

$$\mathbf{E}e^{aw} = \exp(a\mu_w + \frac{1}{2}a^2\sigma_w^2) \quad (5)$$

$$\mathbf{E}(we^{aw}) = (\mu_w + a\sigma_w^2)Ee^{aw} \quad (6)$$

$$\mathbf{E}(w^2e^{aw}) = \{\sigma_w^2 + (\mu_w + a\sigma_w^2)^2\}Ee^{aw} \quad (7)$$

$$\mathbf{E}(w^3e^{aw}) = \{3(\mu_w + a\sigma_w^2)\sigma_w^2 + (\mu_w + a\sigma_w^2)^3\}Ee^{aw}. \quad (8)$$

Now,

$$\mathbf{E}y = \mathbf{E}\{\mathbf{E}(y | x)\} = \mathbf{E}\{\exp(\beta_0 + \beta_1 x)\} = \exp(\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_x^2)$$

$$\begin{aligned} \mathbf{E}(yw) &= \mathbf{E}\{\mathbf{E}(yw | x)\} = \mathbf{E}\{x \exp(\beta_0 + \beta_1 x)\} \\ &= (\mu_w + \beta_1\sigma_x^2) \exp(\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_x^2) \end{aligned}$$

$$\mathbf{E}\{\exp(b_0 + b_1 w)\} = \exp(b_0 + b_1\mu_w + \frac{1}{2}b_1^2\sigma_w^2)$$

$$\mathbf{E}\{w \exp(b_0 + b_1 w)\} = (\mu_w + b_1\sigma_w^2) \exp(b_0 + b_1\mu_w + \frac{1}{2}b_1^2\sigma_w^2),$$

where we used (5) and (6) with $a = \beta_1$ and x in place of w for the first two equations and $a = b_1$ for the last two equations. Note that $\mu_w = \mu_x$. Substituting these expected values into (4) and taking logarithms in the first component equation, we derive the following system

$$\begin{aligned}\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_x^2 &= b_0 + b_1\mu_w + \frac{1}{2}b_1^2\sigma_w^2 \\ (\mu_w + \beta_1\sigma_x^2) \exp(\beta_0 + \beta_1\mu_w + \frac{1}{2}\beta_1^2\sigma_x^2) &= (\mu_w + b_1\sigma_w^2) \exp(b_0 + b_1\mu_w + \frac{1}{2}b_1^2\sigma_w^2),\end{aligned}$$

which has the solution

$$b_1 = (1 - \frac{\sigma_u^2}{\sigma_w^2})\beta_1 \quad (9)$$

$$b_0 = \beta_0 + \beta_1\mu_w \frac{\sigma_u^2}{\sigma_w^2} + \frac{1}{2}\beta_1^2\sigma_u^2(1 - \frac{\sigma_u^2}{\sigma_w^2}), \quad (10)$$

where we used the relation $\sigma_w^2 = \sigma_x^2 + \sigma_u^2$. The bias of $\hat{\beta}_N$ now simply is $b - \beta$. We observe the well-known attenuation effect in $\hat{\beta}_{1N}$.

2.3 Variance

We shall now derive the asymptotic covariance matrix Σ_N of $\hat{\beta}_N$. To this purpose, consider the left hand side of the estimating equation (3) and call it $S_n(\hat{\beta}_N)$. A Taylor series expansion of S_n gives

$$S_n(\hat{\beta}_N) = S_n(b) + \frac{\partial S_n(b)}{\partial b'}(\hat{\beta}_N - b) + O(\|\hat{\beta}_N - b\|^2) = 0. \quad (11)$$

We shall show that, for certain symmetric nonsingular matrices $A(b)$ and $B(b)$,

$$\begin{aligned}\sqrt{n}S_n(b) &\xrightarrow{d} N(0, B(b)) \\ -\frac{\partial S_n(b)}{\partial b'} &\xrightarrow{p} A(b).\end{aligned}$$

From this and (11) it then follows that

$$\sqrt{n}(\hat{\beta}_N - b) \xrightarrow{d} N(0, \Sigma_N), \quad \Sigma_N = A^{-1}(b)B(b)A^{-1}(b), \quad (12)$$

see also Lemma 2 in Kukush and Schneeweiss (2000). We are going to derive $A(b)$ and $B(b)$. First,

$$\sqrt{n}S_n(b) = \frac{1}{\sqrt{n}} \sum_1^n \{y_i - \exp(b_0 + b_1 w_i)\}(1, w_i)'$$

Because of (4), $ES_n(b) = 0$ and thus by the CLT

$$\sqrt{n}S_n(b) \rightarrow N(0, B(b))$$

with

$$B(b) = \mathbf{E}\{[y - \exp(b_0 + b_1 w)]^2 W\}, \quad (13)$$

where

$$W = \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix}.$$

We shall first take the conditional expectation given w . To this purpose, we first compute, with the help of (5), the conditional expectation of y given w :

$$\begin{aligned} \mathbf{E}(y | w) =: m(w, \beta) &= \mathbf{E}\{\mathbf{E}(y | x) | w\} = \mathbf{E}\{\exp(\beta_0 + \beta_1 x) | w\}, \\ m(w, \beta) &= \exp(\beta_0 + \beta_1 \mu(w) + \frac{1}{2}\beta_1^2 \tau^2) \end{aligned} \quad (14)$$

with

$$\mu(w) = \mathbf{E}(x | w) = w - \frac{\sigma_u^2}{\sigma_w^2}(w - \mu_w) \quad (15)$$

$$\tau^2 = \mathbf{V}(x | w) = \sigma_u^2 \left(1 - \frac{\sigma_u^2}{\sigma_w^2}\right), \quad (16)$$

and similarly the conditional variance given w :

$$\begin{aligned} \mathbf{V}(y | w) &=: v(w, \beta) = \mathbf{V}\{\mathbf{E}(y | x) | w\} + \mathbf{E}\{\mathbf{V}(y | x) | w\} \\ &= \mathbf{E}[\{\mathbf{E}(y | x)\}^2 | w] - [\mathbf{E}\{\mathbf{E}(y | x) | w\}]^2 + \mathbf{E}\{\mathbf{E}(y | x) | w\} \\ v(w, \beta) &= m^2(w, \beta)(\exp(\beta_1^2 \tau^2) - 1) + m(w, \beta). \end{aligned} \quad (17)$$

See, e.g., Thamerus (1997), (1998). It then follows that

$$\mathbf{E}[\{y - \exp(b_0 + b_1 w)\}^2 | w] = v(w, \beta) + \{m(w, \beta) - \exp(b_0 + b_1 w)\}^2$$

and therefore, by (13),

$$B(b) = \mathbf{E}[(v(w, \beta) + \{m(w, \beta) - \exp(b_0 + b_1 w)\}^2)W] \quad (18)$$

Next,

$$-\frac{\partial S_n(b)}{\partial b'} = \frac{1}{n} \sum_1^n \exp(b_0 + b_1 w_i) W_i.$$

By the LLN this converges in probability to

$$A(b) = \mathbf{E}[\exp(b_0 + b_1 w)W]. \quad (19)$$

Obviously $A(b)$ is positive definite and therefore invertible, and so (12)

is proved.

It is possible to evaluate the matrices $A(b)$ and $B(b)$ by computing the expectations with the help of (5) to (8). But we shall not do so here. Instead we are looking for simpler approximations.

2.4 Small- σ_u approximation

By expanding $A(b)$ and $B(b)$ as a Taylor series in σ_u^2 we derive approximations for $A(b)$ and $B(b)$ and finally for the asymptotic covariance matrix of $\hat{\beta}_N$ up to the order of σ_u^2 for $\sigma_u^2 \rightarrow 0$. Consider $A(b)$ first. From (9) and (10) we see that

$$b_0 + b_1 w = \gamma + \sigma_u^2 \beta_1 \left(\frac{1}{2} \beta_1 + \frac{\mu_w}{\sigma_w^2} - \frac{w}{\sigma_w^2} \right) + O_p(\sigma_u^4),$$

where we introduced the linear index

$$\gamma = \gamma(w) = \beta_0 + \beta_1 w.$$

It follows that¹

$$\exp(b_0 + b_1 w) = e^\gamma + \sigma_u^2 \beta_1 \left(\frac{1}{2} \beta_1 + \frac{\mu_w}{\sigma_w^2} - \frac{w}{\sigma_w^2} \right) e^\gamma + O_p(\sigma_u^4) \quad (20)$$

and thus, according to (19),

$$A(b) = \mathbf{E}(e^\gamma W) + \sigma_u^2 \beta_1 \left\{ \left(\frac{1}{2} \beta_1 + \frac{\mu_w}{\sigma_w^2} \right) \mathbf{E}(e^\gamma W) - \frac{1}{\sigma_w^2} \mathbf{E}(e^\gamma w W) \right\} + O(\sigma_u^4).$$

¹We use the relation $e^{a\sigma_u^2} = 1 + a\sigma_u^2 + O(\sigma_u^4)$.

Now $\mathbf{E}(e^\gamma W) = A(\beta)$, which we denote by A . With the help of (5) to (7) we obtain

$$A = \mathbf{E}(e^\gamma W) = \mathbf{E} \left\{ e^\gamma \begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} \right\} = e^d \begin{pmatrix} 1 & g \\ g & g^2 + \sigma_w^2 \end{pmatrix} \quad (21)$$

with

$$d = \beta_o + \beta_1 \mu_w + \frac{1}{2} \beta_1^2 \sigma_w^2 \quad (22)$$

$$g = \mu_w + \beta_1 \sigma_w^2 \quad (23)$$

Again using (5) to (7) and also (8), we find

$$\mathbf{E}(e^\gamma w W) = gA + \sigma_w^2 e^d \begin{pmatrix} 0 & 1 \\ 1 & 2g \end{pmatrix}. \quad (24)$$

Putting all this together, we finally get

$$A(b) = A - \sigma_u^2 C + O(\sigma_u^4) \quad (25)$$

with

$$C = \frac{1}{2} \beta_1^2 A + \beta_1 G, \quad (26)$$

$$G = e^d \begin{pmatrix} 0 & 1 \\ 1 & 2g \end{pmatrix}.$$

Let us now turn to $B(b)$. We shall need approximations for $m(w, \beta)$ and $v(w, \beta)$. Substituting (15) and (16) into (14) and taking (20) into account, we obtain

$$m(w, \beta) = \exp(b_0 + b_1 w) + O_p(\sigma_u^4). \quad (27)$$

In addition, see (17) and (16),

$$v(w, \beta) = m(w, \beta) + \sigma_u^2 \beta_1^2 m^2(w, \beta) + O_p(\sigma_u^4),$$

and as $m(w, \beta) = e^\gamma + O_p(\sigma_u^2)$ because of (27) and (20),

$$v(w, \beta) = m(w, \beta) + \sigma_u^2 \beta_1^2 e^{2\gamma} + O_p(\sigma_u^4). \quad (28)$$

Now, by (18), (27) and (28),

$$\begin{aligned} B(b) &= \mathbf{E}[v(w, \beta)W] + O(\sigma_u^4) \\ &= \mathbf{E}[m(w, \beta)W] + \sigma_u^2 \beta_1^2 \mathbf{E}(e^{2\gamma}W) + O(\sigma_u^4) \end{aligned} \quad (29)$$

and by (27) and (19)

$$\begin{aligned} B(b) &= \mathbf{E}[\exp(b_0 + b_1 w)W] + \sigma_u^2 \beta_1^2 \mathbf{E}(e^{2\gamma}W) + O(\sigma_u^4) \\ &= A(b) + \sigma_u^2 \beta_1^2 \tilde{A} + O(\sigma_u^4), \end{aligned}$$

where we introduced the matrix

$$\tilde{A} = \mathbf{E}(e^{2\gamma}W), \quad (30)$$

which is constructed in the same way as A , see (21) to (23), except that β is replaced by 2β . Thus

$$\begin{aligned} \tilde{A} &= e^{\tilde{d}} \begin{pmatrix} 1 & \tilde{g} \\ \tilde{g} & \tilde{g} + \sigma_w^2 \end{pmatrix} \\ \tilde{d} &= 2(\beta_0 + \beta_1 \mu_w + \beta_1^2 \sigma_w^2) \\ \tilde{g} &= \mu_w + 2\beta_1 \sigma_w^2. \end{aligned}$$

As a last step we use (25) and get

$$B(b) = A + \sigma_u^2(-C + \beta_1^2 \tilde{A}) + O(\sigma_u^4) =: B + O(\sigma_u^4). \quad (31)$$

Note that

$$B = A + O(\sigma_u^2). \quad (32)$$

Now, with the help of (25) and (31), the asymptotic covariance matrix of $\hat{\beta}_N$, see (12), can be expanded as follows, where the last equation follows from (32):

$$\begin{aligned} \Sigma_N &= (A - \sigma_u^2 C)^{-1} B (A - \sigma_u^2 C)^{-1} + O(\sigma_u^4) \\ &= (A^{-1} + \sigma_u^2 A^{-1} C A^{-1}) B (A^{-1} + \sigma_u^2 A^{-1} C A^{-1}) + O(\sigma_u^4) \\ &= A^{-1} B A^{-1} + 2\sigma_u^2 A^{-1} C A^{-1} + O(\sigma_u^4) \end{aligned}$$

and finally, again by (31),

$$\Sigma_N = A^{-1} + \sigma_u^2 A^{-1} (C + \beta_1^2 \tilde{A}) A^{-1} + O(\sigma_u^4). \quad (33)$$

3 The corrected score estimator

3.1 Estimator

The corrected score estimator $\hat{\beta}_{CS}$ is found as the solution to a ‘‘corrected’’ estimating equation. Starting from the likelihood score function of the model, see (2),

$$\psi(\beta; y, x) = \{y - \exp(\beta_0 + \beta_1 x)\}(1, x)',$$

we are looking for a “corrected” quasi score function $\psi_c(\beta; y, w)$ such that

$$\mathbf{E}[\psi_c(\beta; y, w) \mid y, x] = \psi(\beta; y, x). \quad (34)$$

Such a function is given by

$$\psi_c(\beta; y, w) = (y - e^c)(1, w)' + \sigma_u^2 e^c(0, \beta_1)' \quad (35)$$

with the abbreviation

$$c = \beta_0 + \beta_1 w - \frac{1}{2} \beta_1^2 \sigma_u^2 = \gamma - \frac{1}{2} \beta_1^2 \sigma_u^2.$$

(34) is verified with the help of (5) and (6) with w replaced by u . See also Carroll et al (1995), Chapter 6.

The corrected estimating equation is now

$$\sum_{i=1}^n \psi_c(\hat{\beta}; y_i, w_i) = 0$$

or

$$\sum_1^n y_i(1, w_i)' = \sum_1^n \exp(\hat{\beta}_0 + \hat{\beta}_1 w_i - \frac{1}{2} \hat{\beta}_1^2 \sigma_u^2)(1, w_i - \hat{\beta}_1 \sigma_u^2)'$$

It can be solved by a Newton algorithm. Let $\beta^k = (\beta_0^k, \beta_1^k)'$ be the approximate value of $\hat{\beta}$ at the k 'th iteration and define $c_i^k = \beta_0^k + \beta_1^k w_i - \frac{1}{2} (\beta_1^k)^2 \sigma_u^2$. Then $\Delta \beta^k := \beta^{k+1} - \beta^k$ is computed from the equation

$$\Delta \beta^k = \left[\sum_1^n e^{c_i^k} \begin{pmatrix} 1 & w_i - \beta_1^k \sigma_u^2 \\ w_i - \beta_1^k \sigma_u^2 & (w_i - \beta_1^k \sigma_u^2)^2 - \sigma_u^2 \end{pmatrix} \right]^{-1} * \sum_1^n \left[y_i \begin{pmatrix} 1 \\ w_i \end{pmatrix} - e^{c_i^k} \begin{pmatrix} 1 \\ w_i - \beta_1^k \sigma_u^2 \end{pmatrix} \right].$$

The algorithm stops when $\Delta\beta^k$ is sufficiently small and then $\hat{\beta}_{CS} = \beta^k + \Delta\beta^k$.

Note: It is also possible, and even simpler, to derive the corrected score function ψ_c by first correcting the log-likelihood function (1) via

$$l_c(\beta) = \sum_1^n \{y_i(\beta_0 + \beta_1 w_i) - \exp(\beta_0 + \beta_1 w_i - \frac{1}{2}\beta_1^2 \sigma_u^2)\}$$

and then taking derivatives with respect to β . It might also be advantageous to maximize l_c directly rather than to solve the corrected estimating equation.

3.2 Variance and small- σ_u approximation

According to the theory of quasi score estimators, see Heyde (1997),

$$\sqrt{n}(\hat{\beta}_{CS} - \beta) \rightarrow N(0, \Sigma_{CS})$$

with an asymptotic covariance matrix given by

$$\begin{aligned} \Sigma_{CS} &= A_c^{-1} B_c A_c^{-1}, \\ A_c &= -\mathbf{E}\left\{\frac{\partial}{\partial \beta'} \psi_c(\beta; y, w)\right\} \\ B_c &= \mathbf{E}\{\psi_c(\beta; y, w) \psi_c'(\beta; y, w)\}. \end{aligned} \tag{36}$$

From (35) we get after some algebra

$$-\frac{\partial \psi_c}{\partial \beta'} = e^c \left[\begin{pmatrix} 1 & w \\ w & w^2 \end{pmatrix} - \sigma_u^2 \begin{pmatrix} 0 & \beta_1 \\ \beta_1 & 1 + 2\beta_1 w - \beta_1^2 \sigma_u^2 \end{pmatrix} \right].$$

Taking expectations and remembering (21), we obtain

$$\begin{aligned}
A_c &= \exp(-\tfrac{1}{2}\beta_1^2\sigma_u^2) \left[A - \sigma_u^2 \mathbf{E} \left\{ e^\gamma \begin{pmatrix} 0 & \beta_1 \\ \beta_1 & 1 + 2\beta_1 w - \beta_1^2\sigma_u^2 \end{pmatrix} \right\} \right] \\
A_c &= \exp(-\tfrac{1}{2}\beta_1^2\sigma_u^2) \{ A - \sigma_u^2 F \} + O(\sigma_u^4), \tag{37}
\end{aligned}$$

where

$$F = \mathbf{E} \left\{ e^\gamma \begin{pmatrix} 0 & \beta_1 \\ \beta_1 & 1 + 2\beta_1 w \end{pmatrix} \right\}.$$

This expectation can be evaluated by borrowing from (21):

$$F = e^d \begin{pmatrix} 0 & \beta_1 \\ \beta_1 & 1 + 2\beta_1 g \end{pmatrix}; \tag{38}$$

for d and g see (22) and (23).

As for the term B_c , we first obtain from (35)

$$\begin{aligned}
B_c &= \mathbf{E}\{(y - e^c)^2 W + \sigma_u^2 (y - e^c) \beta_1 e^c \tilde{W}\} + O(\sigma_u^4), \\
\tilde{W} &= \begin{pmatrix} 0 & 1 \\ 1 & 2w \end{pmatrix}.
\end{aligned}$$

The expectation is evaluated by first taking conditional expectations given w and then expectations with respect to w . Thus

$$\begin{aligned}
B_c &= \mathbf{E}\{[v(w, \beta) + (m(w, \beta) - e^c)^2] W\} \\
&+ \sigma_u^2 \mathbf{E}\{(m(w, \beta) - e^c) \beta_1 e^c \tilde{W}\} + O(\sigma_u^4).
\end{aligned}$$

Due to the definition of c , see (35), we have $m - e^c = m - e^\gamma + O_p(\sigma_u^2) = O_p(\sigma_u^2)$ because of (27) and (20). Hence $B_c = \mathbf{E}(vW) + O(\sigma_u^4)$, and according to (29) this equals $B(b) + O(\sigma_u^4)$. Thus due to (31)

$$B_c = B + O(\sigma_u^4). \quad (39)$$

Now we are ready to derive an approximation for Σ_{CS} , see (36). With (37) and (39) we obtain

$$\begin{aligned} \Sigma_{CS} &= \exp(\beta_1^2 \sigma_u^2) (A - \sigma_u^2 F)^{-1} B (A - \sigma_u^2 F)^{-1} + O(\sigma_u^4) \\ &= \exp(\beta_1^2 \sigma_u^2) (A^{-1} + \sigma_u^2 A^{-1} F A^{-1}) B (A^{-1} + \sigma_u^2 A^{-1} F A^{-1}) + O(\sigma_u^4) \\ &= \exp(\beta_1^2 \sigma_u^2) (A^{-1} B A^{-1} + 2\sigma_u^2 A^{-1} F A^{-1}) + O(\sigma_u^4) \\ &= A^{-1} B A^{-1} + \beta_1^2 \sigma_u^2 A^{-1} + 2\sigma_u^2 A^{-1} F A^{-1} + O(\sigma_u^4), \end{aligned} \quad (40)$$

where (32) was used each time for the derivation of the last two equations.

With the help of (31) we get the more extensive expression

$$\Sigma_{CS} = A^{-1} + \sigma_u^2 A^{-1} (\beta_1^2 A + 2F - C + \beta_1^2 \tilde{A}) A^{-1} + O(\sigma_u^4). \quad (41)$$

4 The structural quasi score estimator

4.1 Estimator

We can write the Poisson regression model as a mean-variance model:

$$\begin{aligned} \mathbf{E}(y \mid x) &= \exp(\beta_0 + \beta_1 x) \\ \mathbf{V}(y \mid x) &= \exp(\beta_0 + \beta_1 x). \end{aligned}$$

By conditioning on the observable variable w we derive a new, “observable”, mean-variance model. It is given by (14) and (17), which we repeat for the convenience of the reader.

$$\begin{aligned}\mathbf{E}(y | w) = m(w, \beta) &= \exp(\beta_0 + \beta_1 \mu(w) + \frac{1}{2} \beta_1^2 \tau^2) \\ \mathbf{V}(y | w) = v(w, \beta) &= m^2(w, \beta)(\exp(\beta_1^2 \tau^2) - 1) + m(w, \beta).\end{aligned}$$

The conventional quasi score function for this model is, see Carroll et al (1995):

$$\begin{aligned}\psi_s(\beta; y, w) &= \{y - m(w, \beta)\} v^{-1}(w, \beta) \frac{\partial m(w, \beta)}{\partial \beta} \\ &= \{y - m(w, \beta)\} v^{-1}(w, \beta) m(w, \beta) (1, \mu(w) + \beta_1 \tau^2)'.\end{aligned}\tag{42}$$

The structural quasi score estimator $\hat{\beta}_{SQS}$ is the solution to the estimating equation

$$\sum_{i=1}^n \psi_s(\hat{\beta}; y_i, w_i) = 0.$$

It can be solved by the iteratively reweighted least squares method in a similar way as in 2.1. Here $\Delta \beta^k$ is found from a least squares regression of $\eta = \{y - m(w, \beta^k)\} / \sqrt{v(w, \beta^k)}$ on $\xi_1 = m(w, \beta^k) / \sqrt{v(w, \beta^k)}$ and $\xi_2 = m(w, \beta^k)(\mu(w) + \beta_1^k \tau^2) / \sqrt{v(w, \beta^k)}$. See also Thamerus (1997). The terms $\mu(w)$ and τ^2 , which are also needed to compute m and v , have to be estimated by replacing μ_w and σ_w^2 in (15) and (16) by their usual estimates: $\hat{\mu}_w = \bar{w}$, $\hat{\sigma}_w^2 = \overline{(w - \bar{w})^2}$.

4.2 Variance

If the (nuisance) parameters μ_w and σ_w^2 were known, the asymptotic covariance matrix of $\hat{\beta}_{SQS}$ (more precisely: of $\sqrt{n}(\hat{\beta}_{SQS} - \beta)$) would simply be given by the inverse of the covariance matrix of the quasi score function, see Carroll et al (1995). The latter is, see (42),

$$\begin{aligned}
\Phi &= \mathbf{E}\{\psi_s(\beta; y, w)\psi_s'(\beta; y, w)\} \\
&= \mathbf{E}\{\mathbf{E}[\psi_s(\beta; y, w)\psi_s'(\beta; y, w) \mid w]\} \\
&= \mathbf{E}\{m^2(w, \beta)v^{-1}(w, \beta)(1, \mu(w) + \beta_1\tau^2)'(1, \mu(w) + \beta_1\tau^2)\} \\
&= \mathbf{E}\left\{\frac{m(w, \beta)}{m(w, \beta)(\exp(\beta_1^2\tau^2) - 1) + 1}(1, \mu(w) + \beta_1\tau^2)'(1, \mu(w) + \beta_1\tau^2)\right\}.
\end{aligned} \tag{43}$$

But as μ_w and σ_w^2 have to be estimated, too, the asymptotic covariance matrix of $\hat{\beta}_{SQS}$ is, in fact, given by

$$\Sigma_{SQS} = \Phi^{-1} + \Phi^{-1}(\sigma_w^2 F_1 F_1' + \frac{2}{\sigma_w^4} F_2 F_2')\Phi^{-1},$$

where $F_1 = \mathbf{E}(\partial\psi_s/\partial\mu_w)$ and $F_2 = \mathbf{E}(\partial\psi_s/\partial\sigma_w^{-2})$, see Kukush and Schneeweiss (2000), Theorem 2.

However, for the approximations to be consider below, the additional terms $F_1 F_1'$ and $F_2 F_2'$ can be dropped. The reason for this is that ψ_s depends on μ_w and σ_w^2 only through $\mu(w)$ and τ^2 and that

$$\begin{aligned}
\frac{\partial\mu(w)}{\partial\mu_w} &= O_p(\sigma_w^2), & \frac{\partial\mu(w)}{\partial\sigma_w^{-2}} &= O_p(\sigma_w^2) \\
\frac{\partial\tau^2}{\partial\mu_w} &= 0, & \frac{\partial\tau^2}{\partial\sigma_w^{-2}} &= -\sigma_w^4,
\end{aligned}$$

see (15) and (16), and therefore $F_i = O(\sigma_u^2)$, $i = 1, 2$. Hence Σ_{SQS} can be approximated by Φ^{-1} :

$$\Sigma_{SQS} = \Phi^{-1} + O(\sigma_u^4). \quad (44)$$

4.3 Small- σ_u approximation

We want to derive an approximation to (43) and (44) for small σ_u^2 . We start with the factor in front of the matrix in (43). Let us abbreviate $m(w, \beta)$ by m . By (16)

$$\begin{aligned} \frac{m}{m(\exp(\beta_1^2 \tau^2) - 1) + 1} &= \frac{m}{m\beta_1^2 \sigma_u^2 + 1} + O_p(\sigma_u^4) \\ &= m - \sigma_u^2 \beta_1^2 m^2 + O_p(\sigma_u^4). \end{aligned} \quad (45)$$

Now for the matrix in (43). By (15), (16) and (23)

$$\mu(w) + \beta_1 \tau^2 = w + \sigma_u^2 (g - w) / \sigma_w^2 + O_p(\sigma_u^4)$$

and so

$$\begin{aligned} &(1, \mu(w) + \beta_1 \tau^2)' (1, \mu(w) + \beta_1 \tau^2) \\ &= W + \frac{\sigma_u^2}{\sigma_w^2} \left\{ (g - w) \begin{pmatrix} 0 & 1 \\ 1 & 2w \end{pmatrix} \right\} + O_p(\sigma_u^4). \end{aligned} \quad (46)$$

Taking the product of (45) and (46), we see that Φ can be approximated as follows:

$$\Phi = \mathbf{E} \left[mW - \sigma_u^2 (m^2 \beta_1^2 W - \frac{m}{\sigma_w^2} \mathbf{V}) \right] + O(\sigma_u^4)$$

$$\mathbf{V} = (g - w) \begin{pmatrix} 0 & 1 \\ 1 & 2w \end{pmatrix}.$$

We know from (27) and (20) that

$$m = \exp(b_0 + b_1 w) + O_p(\sigma_u^4) = e^\gamma + O_p(\sigma_u^2).$$

Therefore by (19) and (25)

$$\mathbf{E}(mW) = A(b) + O(\sigma_u^4) = A - \sigma_u^2 C + O(\sigma_u^4).$$

Borrowing from (21), we further obtain

$$\begin{aligned} \mathbf{E}\left(\frac{m}{\sigma_w^2} \mathbf{V}\right) &= \frac{1}{\sigma_w^2} \mathbf{E}(e^\gamma \mathbf{V}) + O(\sigma_u^2) \\ &= \frac{1}{\sigma_w^2} e^d \left\{ g \begin{pmatrix} 0 & 1 \\ 1 & 2g \end{pmatrix} - \begin{pmatrix} 0 & g \\ g & 2(g^2 + \sigma_w^2) \end{pmatrix} \right\} + O(\sigma_u^2) \\ &= -D + O(\sigma_u^2) \end{aligned}$$

with

$$D = 2e^d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (47)$$

Finally, because $m^2 = e^{2\gamma} + O_p(\sigma_u^2)$ and because of (30),

$$\mathbf{E}(m^2 \beta_1^2 W) = \beta_1^2 \tilde{A} + O(\sigma_u^2).$$

Collecting all these results, we obtain

$$\Phi = A - \sigma_u^2 \{C + D + \beta_1^2 \tilde{A}\} + O(\sigma_u^4) \quad (48)$$

and thus, according to (44),

$$\Sigma_{SQS} = A^{-1} + \sigma_u^2 A^{-1}(C + D + \beta_1^2 \tilde{A})A^{-1} + O(\sigma_u^4). \quad (49)$$

5 Comparing covariance matrices for small measurement error variance σ_u^2

We first compare the asymptotic covariance matrices of the two consistent estimators. They turn out to be equal up to the order of σ_u^2 .

Proposition 1

$$\Sigma_{SQS} = \Sigma_{CS} + O(\sigma_u^4).$$

Proof: Consider the difference of (49) and (41)

$$A(\Sigma_{SQS} - \Sigma_{CS})A = \sigma_u^2(2C + D - \beta_1^2 A - 2F) + O(\sigma_u^4).$$

But according to the definitions (26), (38) and (47) the matrix in brackets is zero. (Note that $F = \beta_1 G + \frac{1}{2}D$).

A convenient formula for both Σ_{SQS} and Σ_{CS} up to the order of σ_u^2 is

$$\Sigma_{SQS} = A^{-1} + \sigma_u^2 \left\{ \frac{1}{2} \beta_1^2 A^{-1} + A^{-1}(\beta_1 G + D + \beta_1^2 \tilde{A})A^{-1} \right\} + O(\sigma_u^4) \quad (50)$$

with A, G, D , and \tilde{A} from (21), (26), (47), and (30), respectively, where d

and g come from (22) and (23). The inverse of A is given by

$$A^{-1} = \frac{1}{\sigma_w^2} e^{-d} \begin{pmatrix} g^2 + \sigma_w^2 & -g \\ -g & 1 \end{pmatrix}.$$

Now let us compare the covariance matrices of $\hat{\beta}_N$ and of either $\hat{\beta}_{SQS}$ or $\hat{\beta}_{CS}$.

Proposition 2

$$A(\Sigma_{SQS} - \Sigma_N)A = \sigma_u^2 D + O(\sigma_u^4).$$

Proof: The result follows immediately from the difference of (33) and (49).

We can give an explicit expression for the difference of the covariance matrices:

$$\Sigma_{SQS} - \Sigma_N = \frac{2\sigma_u^2}{\sigma_w^4} e^{-d} \begin{pmatrix} g^2 & -g \\ -g & 1 \end{pmatrix} + O(\sigma_u^4) \quad (51)$$

with d and g as in (22) and (23), respectively. The difference matrix is positive semidefinite up to the order of σ_u^2 .

6 Simulation

The following small simulation study was designed to test the theoretical findings of the previous sections, especially Propositions 1 and 2. We designed a Poisson regression model with $\beta_0 = -1$, $\beta_1 = 0.5$, $x \sim N(5, 1)$, and $u \sim N(0, 0.05)$. Samples of size $n = 500$ were taken, and, for each

sample, estimates of β_0 and β_1 were computed by the various estimation methods (Naive, CS, SQS). This was done $N = 1000$ times. Bias (= average - true parameter value), standard deviation, and the covariance matrix multiplied by n for each estimator were evaluated from the 1000 replications. For ease of comparison the differences of the covariance matrices multiplied by n are also presented.

In addition to the estimators discussed in this paper, we also studied two more estimators: an artificial “Benchmark” estimator, which is simply the ML estimator derived from the Poisson regression model with the “true” variable x in place of the observed variable w , and the Regression Calibration (CAL) estimator, which comes from a Poisson regression model where the variable x is replaced with $\mathbf{E}(x | w) = \mu(w)$, see Carroll et al (1995).

The sample size is large enough so that one may expect the results to correspond to asymptotic theory. We also experimented with $n = 100$ and found results similar to those for $n = 500$. The error variance, σ_u^2 , is pretty small so that the approximations of our small- σ_u theory are brought out most clearly. On the other hand, it is big enough so that there is still a marked bias in the naive estimator. We also tried out larger error covariances, e.g., $\sigma_u^2 = 0.1$ and found similar results. In order to be able to compare the simulation results to those of small- σ_u theory the latter have been presented in the following table under the heading “theoretical”. The bias has been evaluated by (9) and (10), the covariance

matrices Σ_N and Σ_{SQS} or Σ_{CS} by (33) and (49), respectively, and the difference $\Sigma_{SQS} - \Sigma_N$ by (51). It is seen that the theoretical results are rather well corroborated by the simulation results.

Table 1: Comparison of different estimators ($n = 500$)

	Bench	Naive		SQS		CS	CAL	
		simul.	simul.	theor.	simul.	theor.	simul.	simul.
Bias	β_0	0.0001	0.1242	0.1253	-0.0001	0	-0.0012	0.0056
	β_1	-0.0002	-0.0238	-0.0238	-0.0001	0	-0.0000	-0.0001
Std	β_0	0.1108	0.1151	0.1131	0.1190	0.1178	0.1193	0.1185
	β_1	0.0200	0.0210	0.0204	0.0215	0.0212	0.0214	0.0214
	$n\sigma_{\beta_0}^2$	6.1419	6.6238	6.3929	7.0779	6.9347	7.1220	7.0257
	$n\sigma_{\beta_1}^2$	0.2001	0.2214	0.2071	0.2316	0.2248	0.2297	0.2282
	$n\sigma_{\beta_0\beta_1}$	-1.0868	-1.1874	-1.1313	-1.2564	-1.2294	-1.2598	-1.2469

Table 2: Comparison of n -fold Covariance Matrices ($n = 500$)

SQS - Naive (simulated):	$\begin{pmatrix} .454 & -.069 \\ -.069 & .010 \end{pmatrix}$
CS - Naive (simulated):	$\begin{pmatrix} .498 & -.073 \\ -.073 & .008 \end{pmatrix}$
SQS or CS - Naive (theoretical):	$\begin{pmatrix} .540 & -.098 \\ -.098 & .018 \end{pmatrix}$
SQS - CS (simulated):	$\begin{pmatrix} -.044 & .003 \\ .003 & .002 \end{pmatrix}$

The regression calibration estimator, $\hat{\beta}_{CAL}$, fares almost as good as the truly consistent estimators. $\hat{\beta}_{CAL}$ seems to have a small bias for β_0 .

We should like to mention that the iterative algorithms of Sections 3.1 and 4.1 to compute $\hat{\beta}_{CS}$ and $\hat{\beta}_{SQS}$, respectively, do not always converge. Especially for larger σ_u^2 and smaller n than those used in the present simulation study, the algorithm may diverge. Taking $\hat{\beta}_{CAL}$ as a starting value for the iterations helped only in a few cases. Searching for “good” starting values might succeed in bringing the algorithms to convergence,

but seems to be a difficult job. Another possibility, which, however, we did not try, is to bound the estimates of β a priori.

The algorithms were programmed in GAUSS (Edlefsen and Jones, 1995), a mathematical programming language on IBM-compatible PC.

7 Conclusion and discussion

We studied three estimators for the parameters of a Poisson regression model: the naive estimator $\hat{\beta}_N$, which is inconsistent, but for which the bias can be evaluated, see (9) and (10), and two consistent estimators, which make use of the knowledge of the measurement error variance, here assumed to be known. The corrected score estimator $\hat{\beta}_{CS}$ is constructed on the basis of the functional variant of the measurement error model, while the structural quasi score estimator $\hat{\beta}_{SQS}$ is based on the structural variant.

It should be noted that, as expected, all three estimators become identical when there is no measurement error, see (3), (35) and (42).

We computed the asymptotic covariance matrices for all three estimators and approximated them by rather simple expressions up to the order of σ_u^2 , see (33), (41) and (49) or (50). Up to this order $\hat{\beta}_{CS}$ and $\hat{\beta}_{SQS}$ have the same covariance matrix, which is greater than the covariance matrix of $\hat{\beta}_N$. These formulas can, in principle, be used to compute the asymptotic covariance matrices of the estimators at least approximately

for small σ_u^2 .

According to our simulation study they agree pretty well to the variances of the estimates that result from finite samples from models with reasonably small measurement error variances.

We do not maintain that $\hat{\beta}_{CS}$ and $\hat{\beta}_{SQS}$ are equally good estimators. In fact, it seems that $\hat{\beta}_{SQS}$ performs better than $\hat{\beta}_{CS}$ if the error variance becomes larger in the sense that $\hat{\beta}_{SQS}$ has smaller variances for its two components than $\hat{\beta}_{CS}$.

On the other hand, $\hat{\beta}_{SQS}$ is biased when the assumption of a normal distribution for the latent covariate, an assumption which is the basis for the construction of the estimator, is not true. The bias depends on the size of the deviation from normality and can be appreciable. The corrected score estimator $\hat{\beta}_{CS}$ does not depend on the distribution of the covariate and is always consistent. It is thus more robust than $\hat{\beta}_{SQS}$.

It is good to know that the more robust estimator $\hat{\beta}_{CS}$, although it is inferior to the more sensitive estimator $\hat{\beta}_{SQS}$ in case the normality assumption holds true, is approximately almost as efficient as $\hat{\beta}_{SQS}$ for small error variances. This seems to point to $\hat{\beta}_{CS}$ as the estimator of one's choice in general unless one is sure that the normality assumption is correct and the error variance is large. In this case $\hat{\beta}_{SQS}$ seems to be the better estimator.

Finally we would like to mention that the theory developed in the present paper can be extended to the case where the distribution of the

latent covariate is not just a normal distribution but rather a finite mixture of normal distributions. This would rend the structural quasi score estimator more flexible.

References

- Cameron, A.C.and Trivedi, P.K. (1998), *Regression Analysis of Count Data*, Cambridge University Press.
- Carroll, R.J., Ruppert, D., and Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models*, Chapman and Hall, London.
- Edlefsen, L.E.and Jones, S.D. (1995), *GAUSS version 3.2.6*, Kent, Washington: Aptech Systems, Inc.
- Fahrmeir, L. and Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, Heidelberg.
- Heyde, C.C. (1997), *Quasi-Likelihood And Its Application*, Springer, New York
- 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, SFB 386, Ludwig-Maximilians-Universität München.
- Kukush, A., Schneeweiss, H. and Wolf, R. (2001), *Comparison of Three Estimators in a Polynomial Regression with Measurement Errors*,

Discussion Paper 233, SFB 386, Ludwig-Maximilians-Universität München.

Nakamura, T. (1992), *Corrected score functions for errors-in-variables models: methodology and applications to generalized linear models*, Biometrika 77, 127-137.

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

Stefanski, L.A. and Carroll, R.J. (1987), *Conditional scores and optimal scores in generalized linear measurement error models*, Biometrika 74, 703-716.

Thamerus, M. (1997), *Modelling Count Data with Heteroscedastic Measurement Error in the Covariates*, Discussion Paper 58, SFB 386, Ludwig-Maximilians-Universität München.

Thamerus, M. (1998), *Different nonlinear regression models with incorrectly observed covariates*. In: Galata, R. and H. Küchenhoff (eds.), *Econometrics in Theory and Practice*, Physica, Heidelberg.

Wedderburn, R.W.M. (1976), *On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models*, Biometrika 63, 27-32.