Shklyar, Schneeweiss:

A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors

Sonderforschungsbereich 386, Paper 283 (2002)

A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors

Sergiy Shklyar
Department of Mathematics, Kiev Taras Shevchenko University
Kiev, Ukraine

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

Abstract

We consider a Poisson model, where the mean depends on certain covariates in a log-linear way with unknown regression parameters. Some or all of the covariates are measured with errors. The covariates as well as the measurement errors are both jointly normally distributed, and the error covariance matrix is supposed to be known. Three consistent estimators of the parameters — the corrected score, a structural, and the quasi-score estimators — are compared to each other with regard to their relative (asymptotic) efficiencies. The paper extends an earlier result for a scalar covariate.

1 Introduction

The Poisson regression model is one of the basic models used to analyze count data, Cameron and Trivedi(1998), Winkelmann(1997). The response variable Y has a Poisson distribution with a parameter λ that depends log-linearly on a vector of covariates X: $\log \lambda = \beta_0 + \beta_x^\top X$. The regression parameters $\beta = (\beta_0, \beta_x^\top)^\top$ are to be estimated.

When working with this model, it is often assumed that the covariates are measured without errors, and then maximum likelihood (ML) leads to consistent and asymptotically efficient estimates of the regression parameters. It is, however, well-known that the presence of measurement errors U in the covariates destroys this nice picture, for the linear model see Schneeweiss and Mittag (1986), Fuller(1987), Cheng and Van Ness (1998), Wansbeek and Meijer (2000). The naive ML estimator, which does not take the errors into account and works with $W = X + U$ in place of X, is asymptotically biased.

In order to eliminate this bias, several methods have been proposed, see Cameron and Trivedi (1998), Carroll et al (1995). Most of them depend on the assumption that the error variances and covariances are known, an assumption that we also adopt.

The corrected score estimator is based on the log-likelihood function (or, alternatively, the score function) of the error-free model corrected for the measurement error. This approach has been promoted by Stefanski (1989) and Nakamura (1990); for its application to the Poisson model see Carroll et al (1995). This approach does not utilize the distribution of the covariates X. It is therefore a so-called functional method.

By contrast, structural methods work with the assumption that the distribution of X is known, possibly except for a finite number of unknown parameters. Here we assume that X is Gaussian. A well-known method within this class is based on a quasi-score function that is constructed using the conditional mean and variance of Y given W, Armstrong (1989), Caroll et al (1995), Thamerus (1998). The resulting (structural) quasi-score estimator is consistent and asymptotically normal. We here propose another, simpler, structural estimator that only uses the conditional mean of Y given W. It can be constructed either by solving an appropriate unbiased estimating equation or, equivalently, by maximizing a criterion function, both based on conditioning Y on W. The resulting structural estimator may not be efficient as compared to the quasi-score estimator, but it is much simpler, and it also serves as an intermediate type of estimator when it comes to comparing the relative efficiency of corrected score and quasi-score estimators.

The purpose of the paper is to compare the asymptotic covariance matrices of the three consistent estimators of β mentioned above: corrected score, structural, and quasi-score estimator. It turns out that the covariance matrices can be ordered according to the Loewner order relation, the corrected score estimator having the largest covariance matrix. This result holds true for any values of the error variances.

For small error variances the covariance matrices tend to become equal up the order of squared error variances. This result generalizes a corresponding result for the scalar case found in Kukush at al (2001).

The paper is an extension of the scalar case, see Shklyar and Kukush (2002), to the case of a vector valued covariate X. The elements of X need not all be measured with errors, some can be free of errors. It is an advantage of this extension that error-ridden and error-free covariates can be treated simultaneously, see also Augustin (2002).

Section 2 serves to introduce the Poisson model. In Section 3 the corrected score estimator is introduced and its asymptotic covariance matrix is determined. The same is done for the structural estimator in Section 4, and in Section 5 the two covariance matrices are compared. A further comparison with the quasi-score estimator is accomplished in Section 6. Section 7 deals with small measurement errors, and Section 8 concludes with some additional remarks.

2 The model

We consider the joint distribution of an integer valued random variable Y and a p-dimensional random vector X. X is normally distributed with mean vector μ_x and a positive definite covariance matrix Σ_x:

$$X \sim N(\mu_x, \Sigma_x).$$

The conditional distribution of Y given X is a Poisson distribution with parameter λ, which is the conditional expectation of Y given X:

$$\lambda = \mathbb{E}(Y|X).$$

The dependence of λ on X is given by

$$\lambda = \lambda(X, \beta) = \exp(\beta_0 + \beta_x^\top X),$$

where $\beta = (\beta_0, \beta_x^\top)^\top$, β_0 and β_x being the unknown parameters of interest (and $^\top$ is the transposition sign).

We assume that all or some of the components of the covariate vector X cannot be observed directly. Instead we observe the p-dimensional surrogate variable W, which is related to X by the equation

$$W = X + U,$$

where U is an unobservable measurement error vector, which is assumed to be independent of X and Y. We further assume that $U \sim N(0, \Sigma_u)$ and that Σ_u is known. If a component of X can be observed without measurement error, the corresponding component of U vanishes and the corresponding row and column of Σ_u are zero. Thus Σ_u need not be positive definite.

We observe n independent realizations of (Y, W) denoted by (Y_i, W_i), $i = 1, \ldots, n$, from which β_0 and β_x are to be estimated.

Apart from β_0 and β_x, there are also the nuisance parameters μ_x and Σ_x, which typically have to be estimated as well. This can be done easily by computing

$$\hat{\mu}_w = \overline{W} := \frac{1}{n}\sum_{i=1}^{n} W_i \quad \text{and} \quad \hat{\Sigma}_w = S_w := \frac{1}{n}\sum_{i=1}^{n}(W_i - \overline{W})(W_i - \overline{W})^\top,$$

and setting $\hat{\mu}_x = \hat{\mu}_w$ and $\hat{\Sigma}_x = \hat{\Sigma}_w - \Sigma_u$, assuming that the latter matrix is positive definite. Here, however, we suppose that μ_x and Σ_x are known. This assumption is convenient when it comes to comparing the asymptotic covariance matrices of various estimators of β. The assumption may be appropriate for cases where the distribution of X has been studied with a lot of data in advance of the main study of interest.

We suppose that the true value of β lies in the interior of a prespecified compact subset Θ of $\mathbb{R}^{p+1}$.

3 The corrected score estimator

3.1 The estimator

The log-likelihood of the error free model is given by

$$Q_L(b) = \sum_{i=1}^{n} [Y_i \ln \lambda(X_i, b) - \lambda(X_i, b)] \tag{1}$$

with $b = (b_0, b_x)^\top \in \Theta$ and $\lambda(X_i, b) = \exp(b_0 + b_x^\top X_i)$. If we replace the unobservable variables X_i by the observable surrogates W_i, we arrive at the criterion function for the so-called naive estimator, which is found by maximizing

$$Q_{\text{naive}}(b) = \sum_{i=1}^{n} [Y_i \ln \lambda(W_i, b) - \lambda(W_i, b)], \quad b \in \Theta.$$

The resulting estimator $\hat{\beta}_{\text{naive}}$ would be the ML estimator if W were measured without errors, i.e., if $W = X$, and in this case it would be consistent. But as X_i has been replaced by W_i, the estimator $\hat{\beta}_{\text{naive}}$ is inconsistent. To construct a consistent estimator, we have to correct for the measurement error. Let us denote a typical term of the right hand side of (1), dropping the index i, by

$$q(X, Y, b) = Y \ln \lambda(X, b) - \lambda(X, b).$$

We are looking for a "corrected" function $q_{\text{cor}}(W, Y, b)$, such that

$$\mathbb{E}\left(q_{\text{cor}}(W, Y, b) \middle| X, Y\right) = q(X, Y, b),$$

see Carroll et al (1995), Chapter 6. Such a function is given by

$$q_{\text{cor}} = Y \ln \lambda(W, b) - \exp\left(-\tfrac{1}{2} b_x^\top \Sigma_u b_x\right) \lambda(W, b) \tag{2}$$

because

$$\mathbb{E}[\ln \lambda(W, b) | X] = \mathbb{E}(b_0 + b_x^\top W | X) = b_0 + b_x^\top X = \ln \lambda(X, b)$$

and

$$\mathbb{E}[\lambda(W, b) | X] = \exp(b_0 + b_x^\top X) \mathbb{E} \exp(b_x^\top U) = \lambda(X, b) \exp\left(\tfrac{1}{2} b_x^\top \Sigma_u b_x\right),$$

see also Lemma 1 below. The corresponding corrected criterion function is

$$Q_{\mathrm{cor}}(b) = \sum_{i=1}^{n} \left[Y_i \ln \lambda(W_i, b) - \exp\left(-\tfrac{1}{2} b_x^\top \Sigma_u b_x\right) \lambda(W_i, b) \right],$$

and the estimator $\hat{\beta}_{\mathrm{cor}}$ is a measurable solution to

$$\hat{\beta}_{\mathrm{cor}} \in \arg\max_{b \in \Theta} Q_{\mathrm{cor}}(b).$$

Note that $\hat{\beta}_{\mathrm{cor}}$ is a solution to the corrected unbiased estimating equation

$$\frac{\partial}{\partial b} Q_{\mathrm{cor}}(b) = 0.$$

It is therefore called a corrected score estimator. This estimator is strongly consistent, and $\sqrt{n}(\hat{\beta}_{\mathrm{cor}} - \beta)$ converges in distribution to $N(0, \Sigma_{\mathrm{cor}})$, where Σ_{cor} can be found by the following sandwich formula.

Define the corrected score function by

$$S(W, Y, b) = \frac{\partial}{\partial b} q_{\mathrm{cor}}(W, Y, b),$$

and let $S = S(W, Y, \beta)$ and

$$A = -\,\mathbb{E} \frac{\partial S}{\partial \beta^\top}, \qquad B = \operatorname{cov} S. \tag{3}$$

Then, see Kukush et al (2001),

$$\Sigma_{\mathrm{cor}} = A^{-1} B A^{-\top}. \tag{4}$$

We are going to evaluate this matrix. We will see that A is nonsingular.

3.2 A lemma

In the sequal, we will often use the following easy to prove lemma and its corollaries

Lemma 1. Let $W \sim N(\mu_w, \Sigma_w)$ and let f be an arbitrary function for which the following expectation exists. Then, with $\lambda(W, b) = \exp(b_0 + b_x^\top W)$,

$$\mathbb{E}[f(W)\lambda(W, b)] = \exp\left(b_0 + b_x^\top \mu_w + \tfrac{1}{2} b_x^\top \Sigma_w b_x\right) \mathbb{E}[f(W + \Sigma_w b_x)]. \tag{5}$$

Proof. Let $Z \sim N(0, I)$. Then W and $\mu_w + \Sigma_w^{\frac{1}{2}} Z$ have the same distribution and therefore

$$\begin{aligned}
\mathbb{E}[f(W)\lambda(W,b)] &= \mathbb{E}\left[f(\mu_w + \Sigma_w^{\frac{1}{2}} Z)\lambda(\mu_w + \Sigma_w^{\frac{1}{2}} Z, b)\right] \\
&= (2\pi)^{-\frac{p}{2}} \int f(\mu_w + \Sigma_w^{\frac{1}{2}} z) \exp\left(b_0 + b_x^\top \mu_w + b_x^\top \Sigma_w^{\frac{1}{2}} z - \tfrac{1}{2} z^\top z\right) dz \\
&= (2\pi)^{-\frac{p}{2}} \exp\left(b_0 + b_x^\top \mu_w + \tfrac{1}{2} b_x^\top \Sigma_w b_x\right) \\
&\quad \times \int f(\mu_w + \Sigma_w^{\frac{1}{2}} z) \exp\left[-\tfrac{1}{2}(z - \Sigma_w^{\frac{1}{2}} b_x)^\top (z - \Sigma_w^{\frac{1}{2}} b_x)\right] dz \\
&= \exp\left(b_0 + b_x^\top \mu_w + \tfrac{1}{2} b_x^\top \Sigma_w b_x\right) \mathbb{E}[f(\mu_w + \Sigma_w^{\frac{1}{2}} Z + \Sigma_w b_x)],
\end{aligned}$$

which is equal to the right-hand side of (5). ⧫

Lemma 1 has two corollaries. The first one follows from Lemma 1 by applying it to the conditional distribution of W given X, which is $W|X \sim N(X, \Sigma_u)$, and by replacing μ_w and Σ_w with X and Σ_u, respectively.

Corollary 1. Let $W|X \sim N(X, \Sigma_u)$ and let f be an arbitrary function for which the following expectation exists. Then, with $\lambda(W, b) = \exp(b_0 + b_x^\top W)$,

$$\mathbb{E}[f(W)\lambda(W,b)|X] = \exp\left(\tfrac{1}{2} b_x^\top \Sigma_u b_x\right) \lambda(X,b) \mathbb{E}[f(W + \Sigma_u b_x)|X]. \quad (6)$$

For the second corollary, simply note that $\lambda^2(W, b) = \lambda(W, 2b)$.

Corollary 2. With the assumptions of Lemma 1,

$$\mathbb{E}[f(W)\lambda^2(W,b)] = \exp(2b_0 + 2b_x^\top \mu_w + 2b_x^\top \Sigma_w b_x)\mathbb{E}[f(W + 2\Sigma_w b_x)] \quad (7)$$

and with the assumptions of Corollary 1,

$$\mathbb{E}[f(W)\lambda^2(W,b)|X] = \exp(2b_x^\top \Sigma_u b_x)\lambda^2(X,b)\mathbb{E}[f(W + 2\Sigma_u b_x)|X]. \quad (8)$$

3.3 Evaluation of A

We have

$$S = \begin{pmatrix} Y - \mathrm{e}^{-\frac{1}{2}\beta_x^\top \Sigma_u \beta_x} \lambda(W,\beta) \\ YW - (W - \Sigma_u \beta_x)\, \mathrm{e}^{-\frac{1}{2}\beta_x^\top \Sigma_u \beta_x} \lambda(W,\beta) \end{pmatrix} \quad (9)$$

and

$$-\frac{\partial S}{\partial \beta^\top} = e^{-\frac{1}{2}\beta_x^\top \Sigma_u \beta_x} \lambda(W,\beta) \times \begin{pmatrix} 1 & (W-\Sigma_u\beta_x)^\top \\ W-\Sigma_u\beta_x & (W-\Sigma_u\beta_x)(W-\Sigma_u\beta_x)^\top - \Sigma_u \end{pmatrix}. \tag{10}$$

Taking the expectation of (10) and applying Lemma 1 with $b = \beta$ and noting that $\mu_w = \mu_x$ and $\Sigma_w = \Sigma_x + \Sigma_u$, we find

$$\begin{aligned} A &= e^{\beta_0+\beta_x^\top\mu_x+\frac{1}{2}\beta_x^\top\Sigma_x\beta_x} \\ &\quad \times \mathbb{E}\begin{pmatrix} 1 & (W+\Sigma_x\beta_x)^\top \\ W+\Sigma_x\beta_x & (W+\Sigma_x\beta_x)(W+\Sigma_x\beta_x)^\top - \Sigma_u \end{pmatrix} \\ &= e^{\beta_0+\beta_x^\top\mu_x+\frac{1}{2}\beta_x^\top\Sigma_x\beta_x} \\ &\quad \times \begin{pmatrix} 1 & (\mu_x+\Sigma_x\beta_x)^\top \\ \mu_x+\Sigma_x\beta_x & (\mu_x+\Sigma_x\beta_x)(\mu_x+\Sigma_x\beta_x)^\top + \Sigma_x \end{pmatrix} \end{aligned} \tag{11}$$

Note that A turns out to be symmetrical. Inverting A, we get from (11):

$$A^{-1} = e^{-(\beta_0+\beta_x^\top\mu_x+\frac{1}{2}\beta_x^\top\Sigma_x\beta_x)} \times \begin{pmatrix} (\mu_x+\Sigma_x\beta_x)^\top\Sigma_x^{-1}(\mu_x+\Sigma_x\beta_x)+1 & -(\Sigma_x^{-1}\mu_x+\beta_x)^\top \\ -(\Sigma_x^{-1}\mu_x+\beta_x) & \Sigma_x^{-1} \end{pmatrix}. \tag{12}$$

3.4 Evaluation of B

Hereafter, in symmetrical matrices, we will often write down only one of the two corresponding symmetrical entries.

We have from (9) with $\lambda = \lambda(W,\beta)$:

$$SS^\top = \begin{pmatrix} \begin{array}{c} Y^2 \\ -2Y\, e^{-\frac{1}{2}\beta_x^\top\Sigma_u\beta_x}\lambda \\ +e^{-\beta_x^\top\Sigma_u\beta_x}\lambda^2 \end{array} & \begin{array}{c} Y^2W^\top \\ -Y(2W-\Sigma_u\beta_x)^\top e^{-\frac{1}{2}\beta_x^\top\Sigma_u\beta_x}\lambda \\ +(W-\Sigma_u\beta_x)^\top e^{-\beta_x^\top\Sigma_u\beta_x}\lambda^2 \end{array} \\ \dots & \begin{array}{c} Y^2WW^\top \\ -Y[W(W-\Sigma_u\beta_x)^\top + (W-\Sigma_u\beta_x)W^\top] \\ \times e^{-\frac{1}{2}\beta_x^\top\Sigma_u\beta_x}\lambda \\ +(W-\Sigma_u\beta_x)(W-\Sigma_u\beta_x)^\top e^{-\beta_x^\top\Sigma_u\beta_x}\lambda^2 \end{array} \end{pmatrix} \tag{13}$$

We observe that $\mathbb{E}S = 0$. This follows by applying Corollary 1 with $b = \beta$ to the evaluation of $\mathbb{E}(S|X)$. As to the various parts of S in (9), we find

$$\mathbb{E}(Y|X) = \lambda(X, \beta),$$

$$\mathrm{e}^{-\frac{1}{2}\beta_x^\top \Sigma_u \beta_x}\, \mathbb{E}[\lambda(W, \beta)|X] = \lambda(X, \beta),$$

$$\mathbb{E}(YW|X) = \mathbb{E}[\mathbb{E}(YW|X, Y)|X] = \mathbb{E}(YX|X) = \lambda(X, \beta)X,$$

$$\mathrm{e}^{-\frac{1}{2}\beta_x^\top \Sigma_u \beta_x}\, \mathbb{E}[(W - \Sigma_u \beta_x)\lambda(W, \beta)|X] = \lambda(X, \beta)X$$

and then $\mathbb{E}(S|X) = 0$. Hence B can be written as

$$B = \mathbb{E}SS^\top. \tag{14}$$

Applying Corollary 1 and 2 (8) with $b = \beta$, we find from (13) with $\lambda := \lambda(X, \beta)$:

$$\mathbb{E}\left(SS^\top|X, Y\right) = \begin{pmatrix} \begin{array}{c} Y^2 - 2Y\lambda \\ + \mathrm{e}^{\beta_x^\top \Sigma_u \beta_x}\lambda^2 \end{array} & \begin{array}{c} Y^2 X^\top - Y(2X + \Sigma_u \beta_x)^\top \lambda \\ + \mathrm{e}^{\beta_x^\top \Sigma_u \beta_x}(X + \Sigma_u \beta_x)^\top \lambda^2 \end{array} \\ & \\ \cdots & \begin{array}{c} Y^2(XX^\top + \Sigma_u) \\ -Y[(X + \Sigma_u\beta_x)X^\top + X(X + \Sigma_u \beta_x)^\top]\lambda \\ -2Y\Sigma_u\lambda \\ + \mathrm{e}^{\beta_x^\top \Sigma_u \beta_x}\left[(X + \Sigma_u\beta_x)(X + \Sigma_u\beta_x)^\top + \Sigma_u\right]\lambda^2 \end{array} \end{pmatrix} \tag{15}$$

Remember that by the properties of the Poisson distribution

$$\mathbb{E}(Y^2|X) = \lambda(X, \beta) + \lambda^2(X, \beta).$$

Therefore, taking the expectation of (15) with respect to Y and using again the abbreviation $\lambda = \lambda(X, \beta)$, we get

$$
\begin{aligned}
\mathbb{E}\left[SS^\top|X\right] &= \lambda\begin{pmatrix}1 & X^\top\\ X & XX^\top+\Sigma_u\end{pmatrix}\\
&+\lambda^2\begin{pmatrix}-1+\mathrm{e}^{\beta_x^\top\Sigma_u\beta_x} & \begin{array}{r}-X^\top-\beta_x^\top\Sigma_u\\ +\mathrm{e}^{\beta_x^\top\Sigma_u\beta_x}(X+\Sigma_u\beta_x)^\top\end{array}\\ \\ \cdots & \begin{array}{r}(XX^\top+\Sigma_u)\\ -[(X+\Sigma_u\beta_x)X^\top+X(X+\Sigma_u\beta_x)^\top]-2\Sigma_u\\ +\mathrm{e}^{\beta_x^\top\Sigma_u\beta_x}\left[(X+\Sigma_u\beta_x)(X+\Sigma_u\beta_x)^\top+\Sigma_u\right]\end{array}\end{pmatrix}\\
&= \lambda\begin{pmatrix}1 & X^\top\\ X & XX^\top+\Sigma_u\end{pmatrix}\\
&+(\mathrm{e}^{\beta_x^\top\Sigma_u\beta_x}-1)\lambda^2\begin{pmatrix}1 & X^\top+\beta_x^\top\Sigma_u\\ \cdots & (X+\Sigma_u\beta_x)(X+\Sigma_u\beta_x)^\top+\Sigma_u\end{pmatrix}\\
&+\lambda^2\begin{pmatrix}0 & 0\\ 0 & \Sigma_u\beta_x\beta_x^\top\Sigma_u\end{pmatrix}.
\end{aligned}
$$

Applying again Lemma 1 and Corollary 2 (7), but now with W replaced by $X\sim N(\mu_x,\Sigma_x)$, we finally get

$$
\begin{aligned}
B &= \mathbb{E}[\mathbb{E}(SS^\top|X)] =\\
&= \exp\left(\beta_0+\beta_x^\top\mu_x+\tfrac{1}{2}\beta_x^\top\Sigma_x\beta_x\right)\\
&\quad\times\begin{pmatrix}1 & (\mu_x+\Sigma_x\beta_x)^\top\\ \cdots & (\mu_x+\Sigma_x\beta_x)(\mu_x+\Sigma_x\beta_x)^\top+\Sigma_w\end{pmatrix}\\
&\quad+\exp(2\beta_0+2\beta_x^\top\mu_x+2\beta_x^\top\Sigma_x\beta_x)(\mathrm{e}^{\beta_x^\top\Sigma_u\beta_x}-1)\\
&\quad\times\begin{pmatrix}1 & [\mu_x+(\Sigma_w+\Sigma_x)\beta_x]^\top\\ \cdots & [\mu_x+(\Sigma_w+\Sigma_x)\beta_x]\,[\mu_x+(\Sigma_w+\Sigma_x)\beta_x]^\top+\Sigma_w\end{pmatrix}\\
&\quad+\exp(2\beta_0+2\beta_x^\top\mu_x+2\beta_x^\top\Sigma_x\beta_x)\begin{pmatrix}0 & 0\\ 0 & \Sigma_u\beta_x\beta_x^\top\Sigma_u\end{pmatrix}.
\end{aligned}
\tag{16}
$$

3.5 Change of basis

In order to simplify the expressions for A and B, see (11) and (16), we introduce

$$g := \mu_x+\Sigma_x\beta_x$$

and

$$R := \begin{pmatrix} 1 & g^\top \\ 0 & I \end{pmatrix}$$

Then

$$A = R^\top A_1 R \tag{17}$$

with

$$A_1 = \exp\left(\beta_0 + \beta_x^\top \mu_x + \tfrac{1}{2}\beta_x^\top \Sigma_x \beta_x\right) \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_x \end{pmatrix} \tag{18}$$

and

$$B = R^\top B_1 R \tag{19}$$

with

$$\begin{aligned} B_1 &= \exp\left(\beta_0 + \beta_x^\top \mu_x + \tfrac{1}{2}\beta_x^\top \Sigma_x \beta_x\right) \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_w \end{pmatrix} \\ &+ \exp(2\beta_0 + 2\beta_x^\top \mu_x + 2\beta_x^\top \Sigma_x \beta_x)(\mathrm{e}^{\beta_x^\top \Sigma_u \beta_x} - 1) \\ &\times \begin{pmatrix} 1 & \beta_x^\top \Sigma_w \\ \Sigma_w \beta_x & \Sigma_w \beta_x \beta_x^\top \Sigma_w + \Sigma_w \end{pmatrix} \\ &+ \exp(2\beta_0 + 2\beta_x^\top \mu_x + 2\beta_x^\top \Sigma_x \beta_x) \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_u \beta_x \beta_x^\top \Sigma_u \end{pmatrix}. \end{aligned} \tag{20}$$

Here we used the identity

$$\begin{pmatrix} 1 & 0 \\ g & I \end{pmatrix} \begin{pmatrix} 1 & h^\top \\ h & hh^\top + H \end{pmatrix} \begin{pmatrix} 1 & g^\top \\ 0 & I \end{pmatrix} = \begin{pmatrix} 1 & (g+h)^\top \\ g+h & (g+h)(g+h)^\top + H \end{pmatrix}. \tag{21}$$

3.6 Final expression for Σ_{cor}

From (4), (17), and (19) we have

$$R\Sigma_{\mathrm{cor}}R^\top = A_1^{-1} B_1 A_1^{-1}$$

and hence, by (18) and (20),

$$\begin{aligned} R\Sigma_{\mathrm{cor}}R^\top &= \mathrm{e}^{-(\beta_0 + \beta_x^\top \mu_x + \frac{1}{2}\beta_x^\top \Sigma_x \beta_x)} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_x^{-1}\Sigma_w\Sigma_x^{-1} \end{pmatrix} \\ &+ \left(\mathrm{e}^{\beta_x^\top \Sigma_w \beta_x} - \mathrm{e}^{\beta_x^\top \Sigma_x \beta_x}\right) \begin{pmatrix} 1 & \beta_x^\top \Sigma_w \Sigma_x^{-1} \\ \cdots & \Sigma_x^{-1}\Sigma_w \beta_x \beta_x^\top \Sigma_w \Sigma_x^{-1} + \Sigma_x^{-1}\Sigma_w\Sigma_x^{-1} \end{pmatrix} \\ &+ \mathrm{e}^{\beta_x^\top \Sigma_x \beta_x} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_x^{-1}\Sigma_u \beta_x \beta_x^\top \Sigma_u \Sigma_x^{-1} \end{pmatrix}. \end{aligned} \tag{22}$$

4 A simple structural estimator

4.1 The estimator

The corrected score estimator is constructed without using the distribution of X. (In the previous section we used the distribution of X only in order to evaluate the asymptotic covariance matrix of the corrected score estimator $\hat{\beta}_{\text{cor}}$.) There is, however, a completely different approach to the construction of consistent estimators, which utilizes the distribution of X, here specifically the fact that $X \sim N(\mu_x, \Sigma_x)$. The idea is to set up unbiased estimating equations with the help of the conditional mean and possibly also the conditional variance of Y given X. We call estimators originating as the solution to such estimating equations structural estimators because, in the theory of measurement error models, a model with a well-specified distribution for the variable X is often called a structural model.

A simple structural estimator can be defined via the following criterion function. Denote the conditional expectation of Y given W by

$$\mathbb{E}(Y|W) =: m(W, \beta) \tag{23}$$

and replace $\lambda(X_i, b)$ in (1) with $m(W_i, b)$, then

$$Q_{\text{s}}(b) = \sum_{i=1}^{n} [Y_i \ln m(W_i, b) - m(W_i, b)] \tag{24}$$

can be used as a criterion function, which yields a consistent structural estimator as a measurable solution to

$$\hat{\beta}_{\text{s}} \in \arg\max_{b \in \Theta} Q_{\text{s}}(b).$$

As by assumption β is an interior point of Θ, the minimum is eventually (i.e., for sufficiently large n) found by solving the equation

$$\frac{\partial Q_{\text{s}}(b)}{\partial b} = \sum_{i=1}^{n} \frac{Y_i - m(W_i, b)}{m(W_i, b)} \frac{\partial m(W_i, b)}{\partial b} = 0. \tag{25}$$

This is an unbiased estimating equation. Indeed, owing to (23),

$$\mathbb{E}\left(\frac{\partial Q_{\text{s}}(b)}{\partial b} \middle| W_1, \ldots, W_n \right) = 0$$

for $b = \beta$. Consistency of $\hat{\beta}_{\mathrm{s}}$ can be inferred from the general theory of unbiased estimating equations, see, e.g., Heyde (1997). However, a simpler proof can be given via the criterion function (24) along similar lines as the conventional consistency proof for the ML estimator in an error-free model, see also Shklyar and Kukush (2002).

The structural estimator is also asymptotically normal:

$$\sqrt{n}(\hat{\beta}_{\mathrm{s}} - \beta) \longrightarrow N(0, \Sigma_{\mathrm{s}})$$

with an asymptotic covariance matrix which can be computed by a sandwich formula similar to (4). To this purpose, we denote a typical term of (24) by

$$q_{\mathrm{s}}(W, Y, b) = Y \ln m(W, b) - m(W, b), \tag{26}$$

where the index i has been dropped, and define the structural estimating function for $\hat{\beta}_{\mathrm{s}}$ by

$$S_{\mathrm{s}}(W, Y, b) = \frac{Y - m(W, b)}{m(W, b)} \frac{\partial m(W, b)}{\partial b}. \tag{27}$$

Let $S_{\mathrm{s}} := S_{\mathrm{s}}(W, Y, \beta)$ and let

$$A_{\mathrm{s}} = -\mathbb{E}\frac{\partial S_{\mathrm{s}}}{\partial \beta^{\top}}, \qquad B_{\mathrm{s}} = \operatorname{cov} S_{\mathrm{s}} \tag{28}$$

Then A_{s} is nonsingular and

$$\Sigma_{\mathrm{s}} = A_{\mathrm{s}}^{-1} B_{\mathrm{s}} A_{\mathrm{s}}^{\top}. \tag{29}$$

Before we are going to evaluate this matrix we have to determine $m(W, \beta)$.

4.2 The conditional mean

As X and U are Gaussian, the conditional distribution of X given $W = X + U$ is also Gaussian:

$$X|W \sim N(\mu(W), T),$$

where

$$T = \Sigma_x - \Sigma_x \Sigma_w^{-1} \Sigma_x = \Sigma_u - \Sigma_u \Sigma_w^{-1} \Sigma_u \tag{30}$$

and

$$\mu(W) = \Sigma_u \Sigma_w^{-1} \mu_x + \Sigma_x \Sigma_w^{-1} W. \tag{31}$$

Obviously, $\mu(W)$ is a normal random vector:

$$\mu(W) \sim N(\mu_x, \Sigma_x \Sigma_w^{-1} \Sigma_x). \tag{32}$$

Now we consider the conditional mean of Y given X. We first have

$$\mathbb{E}(Y|W) = \mathbb{E}[\mathbb{E}(Y|W,X)|W] = \mathbb{E}[\mathbb{E}(Y|X)|W] = \mathbb{E}[\lambda(X,\beta)|W].$$

Applying Lemma 1 with $b = \beta$ to $X|W$ in place of W, we finally get

$$m(W,\beta) = \exp\left(\beta_0 + \beta_x^\top \mu(W) + \tfrac{1}{2}\beta_x^\top T \beta_x\right). \tag{33}$$

For future reference, we also compute the conditional variance of Y given W, denoted by $v(W,\beta)$, in a similar way.

$$v(W,\beta) = \mathbb{E}[Y^2|W] - m^2(W,\beta) = \mathbb{E}[\lambda(X,\beta) + \lambda^2(X,\beta)|W] - m^2(W,\beta),$$

and, applying again Lemma 1 and in addition Corollary 2 (7) to $X|W$ in place of W, we get

$$v(W,\beta) = m(W,\beta) + \left(e^{\beta_x^\top T \beta_x} - 1\right) m^2(W,\beta). \tag{34}$$

4.3 Evaluation of A_{s}

By (27) and (33) we have

$$S_{\mathrm{s}} = (Y - m(W,\beta)) \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix} \tag{35}$$

and

$$-\frac{\partial S_{\mathrm{s}}}{\partial \beta^\top} = -(Y - m(W,\beta)) \frac{\partial}{\partial \beta^\top} \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix} + \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix} \frac{\partial m(W,\beta)}{\partial \beta^\top}.$$

Because of (23) the first term vanishes when taking the conditional expectation given W, and, using again (33), we get

$$\mathbb{E}\left(-\frac{\partial S_{\mathrm{s}}}{\partial \beta^\top}\middle| W\right) = m(W,\beta) \times \begin{pmatrix} 1 & (\mu(W) + T\beta_x)^\top \\ \mu(W) + T\beta_x & (\mu(W) + T\beta_x)(\mu(W) + T\beta_x)^\top \end{pmatrix}. \tag{36}$$

In order to compute the expected value of (36), we need a further corollary of Lemma 1.

Corollary 3. With the assumptions of Lemma 1 and with $\mu(W)$ as in (32) and $m(W,\beta)$ as in (33),

$$\begin{aligned}\mathbb{E}[m(W,\beta)f\{\mu(W)\}] &= \exp\left(\beta_0+\beta_x^\top\mu_x+\tfrac{1}{2}\beta_x^\top\Sigma_x\beta_x\right) \\ &\quad\times\mathbb{E}\{f[\mu(W)+\Sigma_x\Sigma_w^{-1}\Sigma_x\beta_x]\}\end{aligned} \tag{37}$$

and

$$\begin{aligned}&\mathbb{E}[m(W,\beta)^2 f\{\mu(W)\}] = \\ &\exp[2\beta_0+2\beta_x^\top\mu_x+\beta_x^\top(\Sigma_x+2\Sigma_x\Sigma_w^{-1}\Sigma_x)\beta_x] \\ &\times\mathbb{E}\{f[\mu(W)+2\Sigma_x\Sigma_w^{-1}\Sigma_x\beta_x]\}\end{aligned} \tag{38}$$

Proof. Apply Lemma 1 with $\mu(W)$ in place of W and with $b_0=\beta_0+\frac{1}{2}\beta_x^\top T\beta_x$ and $b_x=\beta_x$ for (37) and with $b_0=2\beta_0+\beta_x^\top T\beta_x$ and $b_x=2\beta_x$ for (38), respectively. Finally substitute T from (30). ♦

Now we can take the expectation of (36) and get, because of (28) and again using (30) and (31),

$$\begin{aligned}A_\mathrm{s} &= e^{\beta_0+\beta_x^\top\mu_x+\frac{1}{2}\beta_x^\top\Sigma_x\beta_x} \\ &\times\begin{pmatrix}1 & (\mu_x+\Sigma_x\beta_x)^\top \\ \mu_x+\Sigma_x\beta_x & (\mu_x+\Sigma_x\beta_x)(\mu_x+\Sigma_x\beta_x)^\top+\Sigma_x\Sigma_w^{-1}\Sigma_x\end{pmatrix}.\end{aligned} \tag{39}$$

4.4 Evaluation of B_s

By (35) we have

$$S_\mathrm{s}S_\mathrm{s}^\top=[Y-m(W,\beta)]^2\begin{pmatrix}1 & (\mu(W)+T\beta_x)^\top \\ \mu(W)+T\beta_x & (\mu(W)+T\beta_x)(\mu(W)+T\beta_x)^\top\end{pmatrix}.$$

Because $\mathbb{E}\{[Y-m(W,\beta)]^2|W\}=v(W,\beta)$, we get with (34)

$$\begin{aligned}\mathbb{E}[S_\mathrm{s}S_\mathrm{s}^\top|W] &= \left[m(W,\beta)+\left(e^{\beta_x^\top T\beta_x}-1\right)m^2(W,\beta)\right] \\ &\times\begin{pmatrix}1 & (\mu(W)+T\beta_x)^\top \\ \mu(W)+T\beta_x & (\mu(W)+T\beta_x)(\mu(W)+T\beta_x)^\top\end{pmatrix},\end{aligned}$$

where the term with $m(W,\beta)$ is the same as the right-hand side of (36). Then, using (37) and (38) of Corollary 3, we get by (28) and (30)

$$B_{\mathrm{s}} = A_{\mathrm{s}} + \left(1 - e^{-\beta_x^\top T \beta_x}\right) e^{2\beta_0 + 2\beta_x^\top \mu_x + 2\beta_x^\top \Sigma_x \beta_x} \begin{pmatrix} 1 & z^\top \\ z & Z \end{pmatrix} \tag{40}$$

with $z = \mu_x + (\Sigma_x + \Sigma_x \Sigma_w^{-1} \Sigma_x)\beta_x$ and $Z = zz^\top + \Sigma_x \Sigma_w^{-1} \Sigma_x$.

4.5 Change of basis

We use the same matrix R as in Section 3.5 in order to simplify A_{s} and B_{s} from (39) and (40), respectively. We have, see (21),

$$A_{\mathrm{s}} = R^\top A_2 R \tag{41}$$

with

$$A_2 = e^{\beta_0 + \beta_x^\top \mu_x + \frac{1}{2}\beta_x^\top \Sigma_x \beta_x} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_x \Sigma_w^{-1} \Sigma_x \end{pmatrix}, \tag{42}$$

and

$$B_{\mathrm{s}} = R^\top B_2 R \tag{43}$$

with

$$\begin{aligned} B_2 = A_2 &+ \left(1 - e^{-\beta_x^\top T \beta_x}\right) e^{2\beta_0 + 2\beta_x^\top \mu_x + 2\beta_x^\top \Sigma_x \beta_x} \\ &\times \begin{pmatrix} 1 & \beta_x^\top \Sigma_x \Sigma_w^{-1} \Sigma_x \\ \Sigma_x \Sigma_w^{-1} \Sigma_x \beta_x & \Sigma_x \Sigma_w^{-1} \Sigma_x \beta_x \beta_x^\top \Sigma_x \Sigma_w^{-1} \Sigma_x + \Sigma_x \Sigma_w^{-1} \Sigma_x \end{pmatrix}. \end{aligned} \tag{44}$$

4.6 Final expression for Σ_{s}

From the sandwich formula (29) and from (41) and (43), we have $R\Sigma_{\mathrm{s}}R^\top = A_2^{-1} B_2 A_2^{-1}$ and hence, by (42) and (44),

$$\begin{aligned} R\Sigma_{\mathrm{s}}R^\top &= e^{-(\beta_0 + \beta_x^\top \mu_x + \frac{1}{2}\beta_x^\top \Sigma_x \beta_x)} \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_x^{-1} \Sigma_w \Sigma_x^{-1} \end{pmatrix} \\ &\quad + e^{\beta_x^\top \Sigma_x \beta_x} \left(1 - e^{-\beta_x^\top T \beta_x}\right) \begin{pmatrix} 1 & \beta_x^\top \\ \beta_x & \beta_x \beta_x^\top + \Sigma_x^{-1} \Sigma_w \Sigma_x^{-1} \end{pmatrix}. \end{aligned} \tag{45}$$

5 Comparison of corrected score estimator and structural estimator

After having derived explicit expressions for the asymptotic covariance matrices of $\hat{\beta}_{\text{cor}}$ and $\hat{\beta}_{\text{s}}$, we can now compare the relative (asymptotic) efficiencies of these two estimators. We have from (22) and (45)

$$\begin{aligned}
&R(\Sigma_{\text{cor}}-\Sigma_{\text{s}})R^\top \\
&= e^{\beta_x^\top \Sigma_x \beta_x}\left(e^{\beta_x^\top \Sigma_u \beta_x}-1\right)\begin{pmatrix} 1 & \beta_x^\top \Sigma_w \Sigma_x^{-1} \\ \Sigma_x^{-1}\Sigma_w\beta_x & \Sigma_x^{-1}\Sigma_w\beta_x\beta_x^\top\Sigma_w\Sigma_x^{-1}+\Sigma_x^{-1}\Sigma_w\Sigma_x^{-1}\end{pmatrix} \\
&+ e^{\beta_x^\top \Sigma_x \beta_x}\begin{pmatrix} 0 & 0 \\ 0 & \Sigma_x^{-1}\Sigma_u\beta_x\beta_x^\top\Sigma_u\Sigma_x^{-1}\end{pmatrix} \\
&- e^{\beta_x^\top \Sigma_x \beta_x}\left(1-e^{-\beta_x^\top T\beta_x}\right)\begin{pmatrix} 1 & \beta_x^\top \\ \beta_x & \beta_x\beta_x^\top+\Sigma_x^{-1}\Sigma_w\Sigma_x^{-1}\end{pmatrix}.
\end{aligned}$$

If $\Sigma_u\beta_x = 0$, then $\Sigma_{\text{cor}} = \Sigma_{\text{s}}$. We shall prove that otherwise $\Sigma_{\text{cor}} > \Sigma_{\text{s}}$. We change the basis once more. Let

$$D = \begin{pmatrix} 1 & \beta_x^\top \\ 0 & I \end{pmatrix}.$$

Then

$$R(\Sigma_{\text{cor}} - \Sigma_{\text{s}})R^\top = e^{\beta_x^\top \Sigma_x \beta_x} D^\top F D, \tag{46}$$

where

$$\begin{aligned}
F &= \left(e^{\beta_x^\top \Sigma_u \beta_x}-1\right)\begin{pmatrix} 1 & \beta_x^\top \Sigma_u \Sigma_x^{-1} \\ \Sigma_x^{-1}\Sigma_u\beta_x & \Sigma_x^{-1}\Sigma_u\beta_x\beta_x^\top\Sigma_u\Sigma_x^{-1}+\Sigma_x^{-1}\Sigma_w\Sigma_x^{-1}\end{pmatrix} \\
&+ \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_x^{-1}\Sigma_u\beta_x\beta_x^\top\Sigma_u\Sigma_x^{-1}\end{pmatrix} \\
&- \left(1-e^{-\beta_x^\top T\beta_x}\right)\begin{pmatrix} 1 & 0 \\ 0 & \Sigma_x^{-1}\Sigma_w\Sigma_x^{-1}\end{pmatrix}.
\end{aligned}$$

To derive this formula, one may use (21) with g replaced with β_x. Rearranging terms we get

$$\begin{aligned} F &= \begin{pmatrix} e^{\beta_x^\top \Sigma_u \beta_x} - 2 + e^{-\beta_x^\top \Sigma_u \beta_x} & \left(e^{\beta_x^\top \Sigma_u \beta_x} - 1\right) \beta_x^\top \Sigma_u \Sigma_x^{-1} \\ \left(e^{\beta_x^\top \Sigma_u \beta_x} - 1\right) \Sigma_x^{-1} \Sigma_u \beta_x & e^{\beta_x^\top \Sigma_u \beta_x} \Sigma_x^{-1} \Sigma_u \beta_x \beta_x^\top \Sigma_u \Sigma_x^{-1} \end{pmatrix} \\ &+ \begin{pmatrix} e^{-\beta_x^\top T \beta_x} - e^{-\beta_x^\top \Sigma_u \beta_x} & 0 \\ 0 & \left(e^{\beta_x^\top \Sigma_u \beta_x} - 2 + e^{-\beta_x^\top T \beta_x}\right) \Sigma_x^{-1} \Sigma_w \Sigma_x^{-1} \end{pmatrix} \\ &=: F_1 + F_2. \end{aligned} \tag{47}$$

Let us first consider F_1. As

$$F_1 = e^{\beta_x^\top \Sigma_u \beta_x} \begin{pmatrix} 1 - e^{-\beta_x^\top \Sigma_u \beta_x} \\ \Sigma_x^{-1} \Sigma_u \beta_x \end{pmatrix} \left(1 - e^{-\beta_x^\top \Sigma_u \beta_x}, \quad \beta_x^\top \Sigma_u \Sigma_x^{-1}\right),$$

F_1 is positive semidefinite, and $F_1 = 0$ if $\Sigma_u \beta_x = 0$. As to F_2, let us first note that because of (30)

$$\Sigma_u - T = \Sigma_u \Sigma_w^{-1} \Sigma_u.$$

Consider two cases: If $\Sigma_u \beta_x = 0$, then $\beta_x^\top \Sigma_u \beta_x = \beta_x^\top T \beta_x = 0$ and $F_2 = 0$. If $\Sigma_u \beta_x \neq 0$, then $\beta_x^\top \Sigma_u \beta_x > \beta_x^\top T \beta_x$ and hence

$$e^{-\beta_x^\top T \beta_x} - e^{-\beta_x^\top \Sigma_u \beta_x} > 0.$$

Also, by the property that $e^x > 1 + x$ for $x \neq 0$,

$$e^{\beta_x^\top \Sigma_u \beta_x} - 2 + e^{-\beta_x^\top T \beta_x} > \beta_x^\top \Sigma_u \beta_x - \beta_x^\top T \beta_x > 0.$$

Therefore F_2 is positive definite in this case, and so are F and $\Sigma_{\text{cor}} - \Sigma_{\text{s}}$. We thus have proved the following main result of the paper.

Theorem 1. Let Σ_u, μ_x, and $\Sigma_x > 0$ be known. If $\Sigma_u \beta_x = 0$, then $\Sigma_{\text{cor}} = \Sigma_{\text{s}}$, otherwise, if $\Sigma_u \beta_x \neq 0$, then $\Sigma_{\text{cor}} - \Sigma_{\text{s}}$ is a positive definite matrix.

Note. In general, $\Sigma_{\text{cor}} - \Sigma_{\text{s}}$ is positive semidefinite. If all elements of the vector X are error-prone and the errors are linearly independent a.s., then $\Sigma_{\text{cor}} - \Sigma_{\text{s}}$ is positive definite if, and only if, $\beta_x \neq 0$. More generally, if some elements of X are free of measurement errors and the errors of the remaining elements are linearly independent a.s., then $\Sigma_{\text{cor}} - \Sigma_{\text{s}}$ is positive definite if, and only if, for at least one error-prone element of X the regression coefficient does not vanish. In this sense $\hat{\beta}_{\text{s}}$ is more efficient than $\hat{\beta}_{\text{cor}}$.

6 The (structural) quasi-score estimator

The structural estimator $\hat{\beta}_{\text{s}}$ defined in Section 5 is a member of a wider class of linear structural estimators, viz. those which are given as the solution to an unbiased estimating equation that is linear in the Y_i. Indeed, the estimating function (27) for $\hat{\beta}_{\text{s}}$ is linear in Y. Within this class, an optimal estimating function can be constructed using not only the conditional mean function $m(W,b)$ as in (27), but also the conditional variance function $v(W,b)$ of (34), see Heyde (1997). It is given by

$$S_{\text{qs}}(W,Y,b) = \frac{Y - m(W,b)}{v(W,b)} \frac{\partial m(W,b)}{\partial b} \tag{48}$$

and is called (structural) quasi-score function. The corresponding quasi-score estimator $\hat{\beta}_{\text{qs}}$ is a measurable solution to

$$\sum_{i=1}^{n} S_{\text{qs}}(W_i, Y_i, b) = 0, \quad b \in \Theta.$$

Note that $\hat{\beta}_{\text{qs}}$ is not defined via a criterion function. Nevertheless one can show that $\hat{\beta}_{\text{qs}}$ is consistent and asymptotically normal, see Kukush et al (2002).

$$\sqrt{n}(\hat{\beta}_{\text{qs}} - \beta) \longrightarrow N(0, \Sigma_{\text{qs}})$$

with an asymptotic covariance matrix which is given again by a sandwich formula. Let $S_{\text{qs}} := S_{\text{qs}}(W,Y,\beta)$ and

$$A_{\text{qs}} = -\mathbb{E}\frac{\partial S_{\text{qs}}}{\partial \beta^{\top}}, \qquad B_{\text{qs}} = \mathbb{E} S_{\text{qs}} S_{\text{qs}}^{\top},$$

then

$$\Sigma_{\text{qs}} = A_{\text{qs}}^{-1} B_{\text{qs}} A_{\text{qs}}^{-\top}.$$

However, for a quasi-score estimator this reduces to

$$\Sigma_{\mathrm{qs}} = B_{\mathrm{qs}}^{-1} \tag{49}$$

because

$$A_{\mathrm{qs}} = B_{\mathrm{qs}} = \mathbb{E}\left(\frac{1}{v(W,\beta)}\frac{\partial m(W,\beta)}{\partial\beta}\frac{m(W,\beta)}{\partial\beta^{\top}}\right), \tag{50}$$

as can be easily seen from (48).

According to Heyde (1997), S_{qs} is optimal within the class of linear (in Y) estimating functions. As S_{s} belongs to this class, the difference $\Sigma_{\mathrm{s}} - \Sigma_{\mathrm{qs}}$ is positive semidefinite. But we can say more:

Theorem 2. Let Σ_u, μ_x, and $\Sigma_x > 0$ be known. If $\Sigma_u\beta_x = 0$, then $\Sigma_{\mathrm{s}} = \Sigma_{\mathrm{qs}}$; otherwise, if $\Sigma_u\beta_x \neq 0$, then $\Sigma_{\mathrm{s}} - \Sigma_{\mathrm{qs}}$ is a positive definite matrix.

Proof: If $\Sigma_u\beta_x = 0$, then also $T\beta_x = 0$ and, according to (34), $v(W,\beta) = m(W,\beta)$, so that, by (27) and (48), $S_{\mathrm{qs}} = S_{\mathrm{s}}$ and hence $\Sigma_{\mathrm{qs}} = \Sigma_{\mathrm{s}}$.

Now suppose $\Sigma_u\beta_x \neq 0$. According to (29) and (49) we have to prove

$$A_{\mathrm{s}}^{-1}B_{\mathrm{s}}A_{\mathrm{s}}^{-1} \quad > \quad B_{\mathrm{qs}}^{-1} \tag{51}$$

where, by (27) and (28),

$$A_{\mathrm{s}} = \mathbb{E}\left(\frac{1}{m}\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^{\top}}\right) \tag{52}$$

$$B_{\mathrm{s}} = \mathbb{E}\left(\frac{v}{m^2}\frac{\partial m}{\partial\beta}\frac{\partial m}{\partial\beta^{\top}}\right) \tag{53}$$

and B_{qs} is given by (50). Here and in the sequel we abbreviate $m(W,\beta)$ by m and $v(W,\beta)$ by v. (51) is equivalent to

$$A_{\mathrm{s}}^{-\frac{1}{2}}B_{\mathrm{s}}A_{\mathrm{s}}^{-\frac{1}{2}} > (A_{\mathrm{s}}^{-\frac{1}{2}}B_{\mathrm{qs}}A_{\mathrm{s}}^{-\frac{1}{2}})^{-1} \tag{54}$$

Let

$$w = A_{\mathrm{s}}^{-\frac{1}{2}} \frac{1}{\sqrt{m}} \frac{\partial m}{\partial \beta}$$

and

$$v_0 = \frac{v}{m}.$$

Then, by (50), (52), and (53), inequality (54) is equivalent to

$$\mathbb{E}(v_0\, w\, w^\top) > \left[\mathbb{E}\Big(\frac{1}{v_0} w w^\top\Big)\right]^{-1} \tag{55}$$

with $v_0 > 0$ and $\mathbb{E}(w\, w^\top) = \mathrm{I}$.

According to the Matrix Inequality Lemma of the appendix, (55) is true if we can show for any two vectors x and y that if $y^\top w = v_0\, x^\top w$ a.s. then $x = 0$.

From (33) we get

$$\frac{\partial m}{\partial \beta} = m \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix}$$

and therefore

$$w = A_{\mathrm{s}}^{-\frac{1}{2}} \;\; \sqrt{m} \;\; \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix}. \tag{56}$$

From (34) it follows that

$$v_0 = 1 + \left(\mathrm{e}^{\beta_x^\top T \beta_x} - 1\right) m. \tag{57}$$

Now, by the definition of T, (30), $T\beta_x = \Sigma_x \Sigma_w^{-1} \Sigma_u \beta_x$ and, as Σ_x is nonsingular, the assumption $\Sigma_u \beta_x \neq 0$ implies $T\beta_x \neq 0$. Because T is positive semidefinite, it follows that

$$\mathrm{e}^{\beta_x^\top T \beta_x} - 1 > 0.$$

From the definition of m, (33), we therefore get

$$v_0 = 1 + c\, \mathrm{e}^{\beta_x^\top \mu(W)}$$

with a constant $c > 0$.

Now suppose that for some vectors x and y

$$y^\top w = v_0 \, x^\top w \qquad a.s.$$

By (56) and (57) this can be written as

$$y_0^\top \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix} = \left(1 + c\, \mathrm{e}^{\beta_x^\top \mu(W)}\right) x_0^\top \begin{pmatrix} 1 \\ \mu(W) + T\beta_x \end{pmatrix} \qquad a.s.$$

with $y_0 = A_\mathrm{s}^{-\frac{1}{2}}\sqrt{m}y$ and $x_0 = A_\mathrm{s}^{-\frac{1}{2}}\sqrt{m}x$. As $\mu(W)$ has a density in $\mathbb{R}^p$, see (32), this equality holds true for (Lebesgue measure) almost all $\mu \in \mathbb{R}^p$ in place of $\mu(W)$ and by continuity for all $\mu \in \mathbb{R}^p$; i.e., we have, rearranging terms,

$$(y_0 - x_0)^\top \begin{pmatrix} 1 \\ \mu + T\beta_x \end{pmatrix} = c\, \mathrm{e}^{\beta_x' \mu}\, x_0^\top \begin{pmatrix} 1 \\ \mu + T\beta_x \end{pmatrix} \tag{58}$$

for all $\mu \in \mathbb{R}^p$. But since the left hand side of (58) is linear in μ, whereas the right hand side is exponential, (58) can only hold true if $x_0 = 0$ and thus $x = 0$.

This shows that the condition of the Matrix Inequality Lemma is satisfied, which proves the theorem.

7 Comparison under small errors

Although, according to Theorem 2, the quasi-score estimator $\hat{\beta}_{\mathrm{qs}}$ is more efficient than the corrected score estimator $\hat{\beta}_{\mathrm{cor}}$, it can be shown that their asymptotic covariance matrices are approximatly equal if the measurement errors are small. To be more precise, Kukush at al (2001) showed for the scalar case, where X was a real-valued variable, that $\Sigma_{\mathrm{cor}} = \Sigma_{\mathrm{qs}} + O(\sigma_u^4)$ for $\sigma_u^2 \to 0$. This can be generalized to the vector case of the present paper. The question then is whether this equality also holds true up to a higher order of σ_u^2 or whether the difference of Σ_{cor} and Σ_{qs} shows up already at the order of σ_u^4. It will be shown that the latter is the case. We can also give an explicit formula for the difference of Σ_{cor} and Σ_{s} up to this order.

In order to be able to deal with the vector case, we split a common factor σ^2 from Σ_u writing

$$\Sigma_u = \sigma^2 \Omega_u$$

and let σ^2 tend to zero keeping Ω_u fixed. (The factor σ^2 could be, e.g., $\frac{1}{n}\operatorname{tr}\Sigma_u$). To simplify the notation, we introduce the abbreviations

$$\gamma := \beta_x^\top \Omega_u \beta_x, \quad \delta := \Sigma_x^{-1}\Omega_u\beta_x, \quad \phi := \beta_x^\top \Omega_u \Sigma_x^{-1}\Omega_u\beta_x$$

and note that , due to (30) and because $\Sigma_w^{-1} = \Sigma_x^{-1} + O(\sigma^2)$,

$$\beta_x^\top T \beta_x = \sigma^2\gamma - \sigma^4\phi + O(\sigma^6).$$

We then find from (47)

$$F = F_1 + F_2 = \begin{pmatrix} e^{\sigma^2\gamma} - 2 + e^{-\sigma^2\gamma} & \sigma^2\left(e^{\sigma^2\gamma} - 1\right)\delta^\top \\ \sigma^2\left(e^{\sigma^2\gamma} - 1\right)\delta & \sigma^4 e^{\sigma^2\gamma}\delta\delta^\top \end{pmatrix} + \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

with

$$\begin{aligned} f_{11} &= e^{-\sigma^2\gamma + \sigma^4\phi + O(\sigma^6)} - e^{-\sigma^2\gamma} \\ f_{21} &= f_{12}^\top = 0 \\ f_{22} &= \left(e^{\sigma^2\gamma} - 2 + e^{-\sigma^2\gamma + \sigma^4\phi + O(\sigma^6)}\right)(\Sigma_x^{-1} + \sigma^2\Sigma_x^{-1}\Omega_u\Sigma_x^{-1}). \end{aligned}$$

Using the expansion

$$e^{a\sigma^2} = 1 + a\sigma^2 + \tfrac{1}{2}a^2\sigma^4 + O(\sigma^6),$$

we finally get with some algebra

$$\begin{aligned} F &= \sigma^4\left[\begin{pmatrix}\gamma \\ \delta\end{pmatrix}\begin{pmatrix}\gamma \\ \delta\end{pmatrix}^\top + \begin{pmatrix}\phi & 0 \\ 0 & (\phi + \gamma^2)\Sigma_x^{-1}\end{pmatrix}\right] + O(\sigma^6). \\ &=: \sigma^4 F_0 + O(\sigma^6). \end{aligned}$$

Under the assumption $\Sigma_u\beta_x \neq 0$, we have $\phi > 0$, and hence F_0 is positive definite. In the following theorem the relative (asymptotic) efficiencies of $\hat\beta_{\text{cor}}$ and $\hat\beta_{\text{s}}$ are compared to each other for the case of small error variances. Let $G_0 = e^{\beta_x^\top \Sigma_x \beta_x} R^{-1} D^\top F_0 D R^{-\top}$,where R and D are defined in Sections 3.5 and 5, respectively.

Theorem 3. Let $\Sigma_u = \sigma^2\Omega_u$. Then, when $\sigma^2 \to 0$ with Ω_u fixed,

$$\Sigma_{\text{cor}} - \Sigma_{\text{s}} = \sigma^4 G_0 + O(\sigma^6)$$

and G_0 is positive definite if $\Sigma_u\beta_x \neq 0$. (Otherwise $G_0 = 0$). Also

$$\Sigma_{\text{cor}} - \Sigma_{\text{qs}} = \sigma^4 G_1 + O(\sigma^6)$$

with a positive semidefinite matrix G_1, which is positive definite if $\Sigma_u\beta_x \neq 0$.

8 Conclusion

We compared three consistent estimators of the parameters of a Poisson regression model with measurement errors. The asymptotic covariance matrices of the estimators (but not the estimators themselves) are equal if, and only if, $\Sigma_u\beta_x = 0$. In the typical case, where the error variables are linearly independent, this condition means that the regression coefficients corresponding to error-prone covariates are all zero. Otherwise, if at least one error-prone variable has a non-vanishing regression coefficient, the covariance matrices are strongly ordered with regard to the Loewner ordering such that

$$\Sigma_{\text{cor}} > \Sigma_{\text{s}} > \Sigma_{\text{qs}}.$$

The corrected score estimator $\hat{\beta}_{\text{cor}}$ is constructed without regard to the distribution of the regressor variable X. It is therefore robust against any misspecification of that distribution. On the other hand, both $\hat{\beta}_{\text{s}}$ and $\hat{\beta}_{\text{qs}}$ depend on the distribution of X. If X is not Gaussian, these estimators will be asymptotically biased, just as the naive estimator. It is only when the asumption of normality for X is correct that $\hat{\beta}_{\text{s}}$ and $\hat{\beta}_{\text{qs}}$ are consistent. In that case they are more efficient than $\hat{\beta}_{\text{cor}}$, and, in fact, $\hat{\beta}_{\text{qs}}$ is the most efficient one. Still $\hat{\beta}_{\text{cor}}$ might be the preferred estimator in all cases where one cannot be sure about the distribution of X.

9 Appendix: A matrix inequality

Lemma Let v be a positive random variable and w a random column vector in $\mathbb{R}^m$ with $\mathbb{E}(ww^\top) = I_m$. Assume $\mathbb{E}(\frac{1}{v}w^\top w) < \infty$ and $\mathbb{E}(vw^\top w) < \infty$, then (in the Loewner oder)

$$\mathbb{E}(vww^\top) \geq \left[\mathbb{E}\Big(\frac{1}{v}ww^\top\Big)\right]^{-1}. \qquad (*)$$

Assume further that, for any two vectors $x, y \in \mathbb{R}^m$, the equality $y^\top w = vx^\top w$ a.s. implies $x = 0$ (and therefore also $y = 0$), then the $\geq$ sign in $(*)$ can be replaced by the $>$ sign.

Proof: First note that $\mathbb{E}(\frac{1}{v}ww^\top)$ is p.d. and therefore invertible. Indeed, $x^\top\mathbb{E}(\frac{1}{v}ww^\top)x \geq 0$ for any $x \in \mathbb{R}^m$, and $x^\top\mathbb{E}(\frac{1}{v}ww^\top)x = 0$ implies $w^\top x = 0$ a.s., but then $\mathbb{E}(x^\top ww^\top x) = x^\top x = 0$ and thus $x = 0$. Now let

$$q := \left[\mathbb{E}\Big(\frac{1}{v}ww^\top\Big)\right]^{-1}\frac{w}{\sqrt{v}} - \sqrt{v}w.$$

Then

$$\mathbb{E}(qq^\top) = \mathbb{E}(vww^\top) - \left[\mathbb{E}\Big(\frac{1}{v}ww^\top\Big)\right]^{-1},$$

which is p.s.d..

Now suppose there is an $x \in \mathbb{R}^m$ such that

$$x^\top\mathbb{E}(vww^\top)x = x^\top\left[\mathbb{E}\Big(\frac{1}{v}ww^\top\Big)\right]^{-1}x.$$

Then $x^\top\mathbb{E}(qq^\top)x = 0$ and consequently $x^\top q = 0$ a.s. or equivalently

$$x^\top\left[\mathbb{E}\Big(\frac{1}{v}ww^\top\Big)\right]^{-1}w = v\,x^\top w \qquad a.s.$$

By assumption this implies $x = 0$ and thus the $\geq$ sign in $(*)$ can be replaced with $>$.

Acknowledgement

We thank Alexander Kukush for helpful discussions. Support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

- Armstrong, B. (1995), *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.

- 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.

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

- Fuller, W.A. (1987), *Measurement Error Models*, Wiley, New York.

- Heyde, C.C. (1997), *Quasi-Likelihood And Its Application*, Springer, New York.

- Kukush, A., Schneeweiss, H. and Wolf, R. (2001), *Three Estimators for the Poisson Regression Model with Measurement Errors*, Discussion Paper 243, SFB 386, Ludwig-Maximilians-Universität München.

- 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.

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

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

- Shklyar, S. and Kukush, A. (2002): *Comparison of three estimators in Poisson errors–in–variables model with one covariate*, Discussion Paper 293, Sonderforschungsbereich 386, University of Munich.

- 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.
- 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.
- Wansbeek, T. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics.* Elsevier, Amsterdam et al.
- Winkelmann, R. (1997), *Econometric Analysis of Count Data. (2nd ed.)* Springer, Berlin