Kukush, Shklyar:

Comparison of three estimators in Poisson errors-in-variables model with one covariate

Sonderforschungsbereich 386, Paper 293 (2002)

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

Comparison of three estimators in Poisson errors–in–variables model with one covariate

Alexander Kukush and Sergiy Shklyar

National Taras Shevchenko University, Kiev, Ukraine

Abstract

A structural errors–in–variables model is investigated, where the response variable follows a Poisson distribution. Assuming the error variance to be known, we consider three consistent estimators and compare their relative efficiencies by means of their asymptotic covariance matrices. The comparison is made for arbitrary error variances. The structural quasi–likelihood (QL) estimator is based on a quasi score function, which is constructed from a conditional mean-variance model. The corrected estimator is based on an error–corrected likelihood score function. The alternative estimator is constructed to remove the asymptotic bias of the naive (i.e., ordinary maximum likelihood) estimator. It is shown that the QL estimator is strictly more efficient than the alternative estimator, and the latter one is strictly more efficient than the corrected estimator.

1 Introduction

We suppose that $Y|X$ has a Poisson distribution with parameter $\lambda = \lambda(X,\beta) = \exp(\beta_0+\beta_1 X)$. Here $X \sim N(\mu_x, \sigma_x^2)$, $\sigma_x^2 > 0$, and $\beta = (\beta_0, \beta_1)^\top$ is the parameter of interest. Let $W = X + U$ be a surrogate covariate, where U is independent of X and Y, and $U \sim N(0, \sigma_u^2)$, σ_u^2 is positive and known.

We observe independent realizations (Y_i, W_i), $i = 1, \ldots, n$. We suppose that the parameter set Θ_β is a compact set in $\mathbb{R}^2$, and the true value

β is an interior point of Θ_β; only for the naive and alternative estimator we allow the parameter set to be the entire plane $\mathbb{R}^2$.

In [4] the asymptotic covariance matrices of three estimators were compared, namely for naive estimator (which is inconsistent) and of the corrected and quasi-likelihood estimator (these two are consistent). The comparison was made for small σ_u^2 and unknown μ_x and σ_x^2. In the present paper we compare the two consistent estimators for arbitrary σ_u^2, but we suppose that μ_x and σ_x^2 are known.

We introduce an intermediate, alternative estimator, which removes the asymptotic bias of the naive estimator. The alternative estimator makes it possible to compare the two other consistent estimators.

Throughout the paper $\mathbb{E}$ denotes the mathematical expectation, and var denotes the variance of a random variable, while cov denotes the variance–covariance matrix of a random vector. Let $\partial S/\partial\beta^\top$ mean the derivative of the score function $S(b)$ at the true value of the parameter β.

In Section 2 the naive estimator is introduced and its inconsistency is shown. In section 3 the corrected estimator is defined, and its asymptotic covariance matrix is evaluated. Section 4 gives the asymptotic properties of the alternative estimator. In Sections 5 and 6 the asymptotic covariance matrices of the last two estimators are compared. Section 7 compares the quasi-likelihood estimator with the other two consistent estimators, and Section 8 concludes.

2 The naive estimator

For the naive estimator, we suppose that the parameter set is $\mathbb{R}^2$.

2.1 The estimator

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

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

with $b = (b_0, b_1)^\top \in \Theta_\beta$ and $\lambda(X_i, b) = \exp(b_0 + b_1 X_i)$. 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 $\hat{\beta}_{\text{naive}}$, which is found by maximizing

$$Q_{\text{naive}}(b) = \sum_{i=1}^{n} [Y_i \ln \lambda(W_i, b) - \lambda(W_i, b)], \quad b \in \mathbb{R}^2,$$

where $\lambda(W, b) = \exp(b_0 + b_1 W)$, (if the maximum is not attained we set $\hat{\beta}_{\text{naive}} = \infty$). The resulting estimator $\hat{\beta}_{\text{naive}}$ coincides with the ML estimator if W is measured without errors, i.e., if $W = X$, and in that case $\hat{\beta}_{\text{naive}}$ would be consistent. But as X_i has been replaced by W_i, the naive estimator is inconsistent.

2.2 Uniqueness of solution

The function $b \mapsto Y_i \ln \lambda(W_i, b) - \lambda(W_i, b) = Y_i(b_0 + b_1 W_i) - \exp(b_0 + b_1 W_i)$ is bounded from above and concave because it is a composition of the bounded from above strictly concave function $Y_i t - e^t$ and the linear functional $b_0 + b_1 W_i$. Hence $Q_{\text{naive}}(b)$ is a bounded from above concave function, so to find zeroes of

$$S_{\text{naive}} = \frac{\partial Q_{\text{naive}}}{\partial b} = \sum_{i=1}^{n} (Y_i - \lambda(b, W_i)) \begin{pmatrix} 1 \\ W \end{pmatrix}$$

is an equivalent way to define $\hat{\beta}_{\text{naive}}$.

Denote by n_1 the least positive integer such that $Y_{n_1} > 0$ and by n_2 the least integer such that $n_1 > n_2$ and $Y_{n_2} > 0$. We have $n_1 < n_2 < \infty$ a.s.

Statement 1. If the random event $n_2 \le n$ occurs, then almost surely Q_{naive} has a unique maximum point. It is the unique zero of S_{naive}.

Proof. Almost surely $W_{n_1} \ne W_{n_2}$. In this case, it is easy to prove that the sum of the two terms of Q_{naive} corresponding to n_1-th and n_2-th observations is strictly concave and has a unique maximum point. We obtain Q_{naive} by adding the bounded from above and concave function, so Q_{naive} is strictly concave. In the considered case $Y_{n_1} > 0$, $Y_{n_2} > 0$, $W_{n_1} \ne W_{n_2}$, and $n \ge n_2$ hold, and the maximum point of Q_{naive} exists because $\sup_{\|b\| \ge R} Q_{\text{naive}}(b) \to -\infty$, as $R \to \infty$ (indeed, we have this for the sum of the two summands, and the remaining summands of Q_{naive} are bounded from above). But the maximum point of Q_{naive} is unique because Q_{naive} is strictly concave. The second part of Statement 1 holds because Q_{naive} is concave.

Definition 1. Let Z be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $Z_1,\ Z_2, \dots$ be independent observations on Z. A sequence of statements $A_1(Z_1),\ A_2(Z_1, Z_2), \dots$ *holds eventually* iff

$$\exists \Omega_0,\ \mathbb{P}(\Omega_0) = 1\ \forall \omega \in \Omega_0\ \exists N(\omega)\ \forall n \geq N(\omega):\ A_n(Z_1, \dots, Z_n)$$

holds true.

Because of Statement 1, the estimator $\hat{\beta}_{\text{naive}}$ is eventually finite, and it satisfies eventually $S_{\text{naive}}(\hat{\beta}_{\text{naive}}) = 0$.

2.3 Inconsistency of the naive estimator

For all b, almost surely, when the number of observations $n \longrightarrow \infty$,

$$\frac{1}{n} Q_{\text{naive}}(b) \longrightarrow Q^{\infty}_{\text{naive}}(b) \tag{2}$$

uniformly on any compact set, where

$$\begin{aligned} Q^{\infty}_{\text{naive}}(b) &= \mathbb{E}\left[Y(b_0 + b_1 W) - \lambda(b, W)\right] = \mathbb{E}\{\mathbb{E}[Y_i(b_0 + b_1 W)|X]\} - \mathbb{E}\lambda(b, W) \\ &= \mathbb{E}\left[\lambda(\beta, X)(b_0 + b_1 X)\right] - \mathbb{E}\lambda(b, W). \end{aligned} \tag{3}$$

Now, remember that $W \sim N(\mu_x, \sigma_w^2)$ and consider, for arbitrary function f, the following expectation, assuming its existence:

$$\begin{aligned} \mathbb{E}[f(W)\lambda(W, b)] &= \int_{-\infty}^{\infty} f(\mu_x + \tau)\, \mathrm{e}^{b_0 + b_1 \mu_x + b_1 \tau}\, \mathrm{e}^{-\frac{\tau^2}{2\sigma_w^2}}\, \frac{1}{\sqrt{2\pi}\sigma_w}\, \mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} f(\mu_x + \tau)\, \mathrm{e}^{b_0 + b_1 \mu_x + \frac{1}{2} b_1^2 \sigma_w^2}\, \mathrm{e}^{-\frac{(\tau - b_1 \sigma_w^2)^2}{2\sigma_w^2}}\, \frac{1}{\sqrt{2\pi}\sigma_w}\, \mathrm{d}\tau \\ &= e^{b_0 + b_1 \mu_x + \frac{1}{2} b_1^2 \sigma_w^2}\, \mathbb{E} f(W + b_1 \sigma_w^2). \end{aligned} \tag{4}$$

Analogously, remembering $X \sim N(\mu_x, \sigma_x^2)$, we can prove that

$$\mathbb{E}[f(X)\lambda(X, \beta)] = e^{\beta_0 + \beta_1 \mu_x + \frac{1}{2}\beta_1^2 \sigma_x^2}\, \mathbb{E} f(X + \beta_1 \sigma_x^2). \tag{5}$$

Applying (4) and (5) to (3) we get

$$Q^{\infty}_{\text{naive}}(b) = e^{\beta_0 + \beta_1 \mu_x + \frac{1}{2}\beta_1^2 \sigma_x^2}(b_0 + b_1 \mu_x + b_1 \beta_1 \sigma_x^2) - e^{b_0 + b_1 \mu_x + \frac{1}{2} b_1^2 \sigma_w^2}.$$

$Q^{\infty}_{\text{naive}}(b)$ is a concave function. So it attains its maximum at the point where its derivative is equal to zero. We have

$$\frac{\partial Q^{\infty}_{\text{naive}}(b)}{\partial b} = e^{\beta_0+\beta_1\mu_x+\frac{1}{2}\beta_1^2\sigma_x^2}\begin{pmatrix}1\\ \mu_x+\beta_1\sigma_x^2\end{pmatrix} - e^{b_0+b_1\mu_x+\frac{1}{2}b_1^2\sigma_w^2}\begin{pmatrix}1\\ \mu_x+b_1\sigma_w^2\end{pmatrix},$$

and $\frac{\partial Q^{\infty}_{\text{naive}}(b)}{\partial b}=0$ holds if and only if

$$\begin{cases} \beta_0+\beta_1\mu_x+\frac{1}{2}\beta_1^2\sigma_x^2 = b_0+b_1\mu_x+\frac{1}{2}b_1^2\sigma_x^2 \\ \mu_x+\beta_1\sigma_x^2=\mu_x+b_1\sigma_w^2 \end{cases}.$$

This system has the only solution $\beta^*_{\text{naive}} = \left(\beta_0+\beta_1\frac{\sigma_u^2}{\sigma_w^2}+\beta_1^2\frac{\sigma_u^2\sigma_x^2}{2\sigma_w^2},\ \beta_1\frac{\sigma_x^2}{\sigma_w^2}\right)^{\top}$. So $Q^{\infty}_{\text{naive}}$ has unique maximum point β^*_{naive}.

Let $\epsilon>0$. Denote by K the circumference of radius ϵ and centre at β^*_{naive}. We have $\sup\limits_K Q^{\infty}_{\text{naive}} < Q^{\infty}_{\text{naive}}(\beta^*_{\text{naive}})$. Because of (2) $\sup\limits_K Q_{\text{naive}} < Q_{\text{naive}}(\beta^*_{\text{naive}})$ eventually. Because of concavity of Q_{naive} the maximum point $\hat\beta_{\text{naive}}$ of Q_{naive} eventually lies inside K. Therefore $\hat\beta_{\text{naive}}$ tends to β^*_{naive} almost surely.

Statement 2. The naive estimator is convergent, as $n\to\infty$,

$$\hat\beta_{\text{naive}} \longrightarrow \begin{pmatrix}\beta_0+\beta_1\frac{\sigma_u^2}{\sigma_x^2}+\beta_1^2\frac{\sigma_u^2\sigma_x^2}{2\sigma_w^2}\\ \beta_1\frac{\sigma_x^2}{\sigma_w^2}\end{pmatrix} \quad \text{(a.s.)}$$

3 The corrected estimator

In this Section and hereafter we assume that the parameter set Θ_β is a compact set in $\mathbb{R}^2$, and the true value of β is an interior point of Θ_β.

3.1 Definition

To construct a consistent estimator, we have to correct the log-likelihood function (1) for the measurement error. Let us denote a typical term of (1) by

$$q_{\text{naive}}(X,Y,b) = Y\ln\lambda(X,b)-\lambda(X,b).$$

We are looking for a "corrected" function $q(W,Y,b)$, such that

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

see [1], Chapter 6. Such function is given by

$$q = Y \ln \lambda(W,b) - \exp\left(-\frac{1}{2} b_1^2 \sigma_u^2\right) \lambda(W,b).$$

The corresponding corrected criterion function is

$$Q_{\mathrm{cor}}(b) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i \ln \lambda(W_i, b) - \mathrm{e}^{-\frac{1}{2} b_1^2 \sigma_u^2} \lambda(W_i, b) \right),$$

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

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

(Q_{cor} does not attain a maximum on the entire plane $\mathbb{R}^2$ because it is unbounded from above.) The 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

$$S(b) = \frac{\partial}{\partial b} q(W, Y, b),$$

and let $S = S(\beta)$ and

$$A = -\mathbb{E} \frac{\partial S}{\partial b^\top}, \qquad B = \operatorname{cov} S, \tag{6}$$

both are taken at the point $b = \beta$. Then, see [1],

$$\Sigma_{\mathrm{cor}} = A^{-1} B A^{-\top}. \tag{7}$$

Hereafter $M^{-\top}$ denotes $(M^{-1})^\top$, where M is a square matrix. We are going to compute this matrix.

3.2 Computation of A

We have

$$S = \begin{pmatrix} Y - \mathrm{e}^{-\frac{1}{2}\beta_1^2 \sigma_u^2} \lambda(W,\beta) \\ YW - (W - \beta_1 \sigma_u^2)\, \mathrm{e}^{-\frac{1}{2}\beta_1^2 \sigma_u^2} \lambda(W,\beta) \end{pmatrix},$$

$$-\frac{\partial S}{\partial \beta^\top} = \mathrm{e}^{-\frac{1}{2}\beta_1^2\sigma_u^2}\,\lambda(W,\beta)\begin{pmatrix} 1 & W-\beta_1\sigma_u^2 \\ W-\beta_1\sigma_u^2 & \left(W-\beta_1\sigma_u^2\right)^2-\sigma_u^2 \end{pmatrix}.$$

By (4), remembering $\sigma_w^2 - \sigma_u^2 = \sigma_x^2$, we have

$$\begin{aligned} A &= \mathrm{e}^{\beta_0+\beta_1\mu_x+\frac{1}{2}\beta_1^2\sigma_x^2}\,\mathbb{E}\begin{pmatrix} 1 & \omega-\beta_1\sigma_u^2 \\ \omega-\beta_1\sigma_u^2 & \left(\omega-\beta_1\sigma_u^2\right)^2-\sigma_u^2 \end{pmatrix} \\ &= \mathrm{e}^{\beta_0+\beta_1\mu_x+\frac{1}{2}\beta_1^2\sigma_x^2}\begin{pmatrix} 1 & \mu_x+\beta_1\sigma_x^2 \\ \mu_x+\beta_1\sigma_x^2 & \left(\mu_x+\beta_1\sigma_x^2\right)^2+\sigma_x^2 \end{pmatrix}. \end{aligned} \qquad (8)$$

3.3 Computation of A^{-1}

The determinant of the latter matrix equals σ_x^2. Therefore

$$A^{-1} = \frac{1}{\sigma_x^2}\,\mathrm{e}^{-(\beta_0+\beta_1\mu_x+\frac{1}{2}\beta_1^2\sigma_x^2)}\begin{pmatrix} \left(\beta_1\sigma_x^2+\mu_x\right)^2+\sigma_x^2 & -(\beta_1\sigma_x^2+\mu_x) \\ -(\beta_1\sigma_x^2+\mu_x) & 1 \end{pmatrix}.$$

3.4 Computation of B

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

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

$$SS^\top = \begin{pmatrix} Y^2 - 2\,\mathrm{e}^{-\frac{1}{2}\beta_1^2\sigma_u^2}\,Y\lambda + \mathrm{e}^{-\beta_1^2\sigma_u^2}\,\lambda^2 & Y^2W - Y(2W-\beta_1\sigma_u^2)\,\mathrm{e}^{-\frac{1}{2}\beta_1^2\sigma_u^2}\,\lambda \\ & +(W-\beta_1\sigma_u^2)\,\mathrm{e}^{-\beta_1^2\sigma_u^2}\,\lambda^2 \\ \dots & Y^2W^2 - 2W(W-\beta_1\sigma_u^2)\,\mathrm{e}^{-\frac{1}{2}\beta_1^2\sigma_u^2}\,\lambda Y \\ & +(W-\beta_1\sigma_u^2)^2\,\mathrm{e}^{-\beta_1^2\sigma_u^2}\,\lambda^2 \end{pmatrix}. \qquad (9)$$

We observe that $\mathbb{E}S = 0$ and so B can be written as

$$B = \mathbb{E}SS^\top. \qquad (10)$$

In order to evaluate this expectation we consider the following three expected values for an arbitrary function, assuming their existence. First,

$$\mathbb{E}\left[f(W)|X\right] = \int_{-\infty}^{\infty} f(X+u)\,\mathrm{e}^{-\frac{u^2}{2\sigma_u^2}}\,\frac{1}{\sqrt{2\pi}\sigma_u}du \qquad (11)$$

because $W|X \sim N(X, \sigma_u^2)$; next

$$\mathbb{E}\left[f(W)\lambda(W,\beta)\, \mathrm{e}^{-\frac{1}{2}\beta_1^2\sigma_u^2} \,\middle|\, X \right] =$$

$$= \int_{-\infty}^{\infty} f(X+u)\, \mathrm{e}^{\beta_0+\beta_1 X+\beta_1 u-\frac{1}{2}\beta_1^2\sigma_u^2}\, \mathrm{e}^{-\frac{u^2}{2\sigma_u^2}} \frac{1}{\sqrt{2\pi}\sigma_u}\, \mathrm{d}u =$$

$$= \int_{-\infty}^{\infty} f(X+u)\, \mathrm{e}^{\beta_0+\beta_1 X}\, \mathrm{e}^{-\frac{(u-\beta_1\sigma_u^2)^2}{2\sigma_u^2}} \frac{1}{\sqrt{2\pi}\sigma_u} \mathrm{d}u = \mathbb{E}\left[f(W_1)|X\right] \lambda(X,\beta), \tag{12}$$

where $W_1|X \sim N(X+\beta_1\sigma_u^2,\ \sigma_u^2)$.
At last, applying (12) with β replaced by 2β, we have

$$\begin{aligned} & \mathbb{E}\left[f(W)\lambda^2(W,\beta)\, \mathrm{e}^{-\beta_1^2\sigma_u^2} \,\middle|\, X \right] \\ = \quad & \mathrm{e}^{\beta_1^2\sigma_u^2}\, \mathbb{E}\left[f(W)\lambda(W,2\beta)\, \mathrm{e}^{-\frac{1}{2}(2\beta_1)^2\sigma_u^2} \,\middle|\, X \right] \\ = \quad & \mathbb{E}\left[f(W_2)|X\right] \lambda(X,2\beta)\, \mathrm{e}^{\beta_1^2\sigma_u^2} \\ = \quad & \mathbb{E}\left[f(W_2)|X\right] \lambda^2(X,\beta)\, \mathrm{e}^{\beta_1\sigma_u^2}, \end{aligned} \tag{13}$$

where $W_2|X \sim N(X+2\beta_1\sigma_u^2,\ \sigma_u^2)$. Therefore with $\lambda = \lambda(X,\beta)$ we have from (9):

$$\mathbb{E}\left(SS^\top|X,Y\right) =$$

$$\begin{pmatrix} Y^2 - 2\lambda Y + \mathrm{e}^{\beta_1^2\sigma_u^2}\,\lambda^2 & Y^2X - Y\lambda(2X+\beta_1\sigma_u^2) + \mathrm{e}^{\beta_1^2\sigma_u^2}\,\lambda^2(X+\beta_1\sigma_u^2) \\ \dots & Y^2(X^2+\sigma_u^2) - 2(X+\beta_1\sigma_u^2)X\lambda Y - 2\sigma_u^2\lambda Y + \\ & + \mathrm{e}^{\beta_1^2\sigma_u^2}\,\lambda^2\left((X+\beta_1\sigma_u^2)^2+\sigma_u^2\right) \end{pmatrix}.$$

Remember that by the properties of the Poisson distribution

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

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

Therefore, again with $\lambda = \lambda(X, \beta)$,

$$\mathbb{E}\left[SS^\top|X\right] = \lambda \begin{pmatrix} 1 & X \\ X & X^2 + \sigma_u^2 \end{pmatrix}$$
$$+\lambda^2 \begin{pmatrix} -1 + e^{\beta_1^2\sigma_u^2} & -X - \beta_1\sigma_u^2 + e^{\beta_1\sigma_u^2}(X + \beta_1\sigma_u^2) \\ \dots & (X^2 + \sigma_u^2) - 2(X + \beta_1\sigma_u^2)X - 2\sigma_u^2 + \\ & + e^{\beta_1^2\sigma_u^2}\left((X + \beta_1\sigma_u^2)^2 + \sigma_u^2\right) \end{pmatrix}$$
$$= \lambda \begin{pmatrix} 1 & X \\ X & X^2 + \sigma_u^2 \end{pmatrix} + (e^{\beta_1^2\sigma_u^2} - 1)\lambda^2 \begin{pmatrix} 1 & X + \beta_1\sigma_u^2 \\ \dots & (X + \beta_1\sigma_u^2)^2 + \sigma_u^2 \end{pmatrix}$$
$$+\lambda^2 \begin{pmatrix} 0 & 0 \\ 0 & \beta_1^2\sigma_u^4 \end{pmatrix}.$$

Finally we have, see Subsection 1.2,

$$\mathbb{E}[f(X)\lambda(X, \beta)] = \mathbb{E}[f(W_3)]\exp(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2),$$

where $W_3 \sim N(\mu_x + \beta_1\sigma_x^2, \sigma_x^2)$;

$$\mathbb{E}\left[f(X + \beta_1\sigma_u^2)\lambda^2(X, \beta)\right] = \mathbb{E}[f(X + \beta_1\sigma_u^2)\lambda(X, 2\beta)] =$$
$$= \mathbb{E}[f(W_4)]\exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2),$$

where $W_4 \sim N(\mu_x + 2\beta_1\sigma_x^2 + \beta_1\sigma_u^2, \sigma_x^2) = N(\mu_x + \beta_1\sigma_x^2 + \beta_1\sigma_w^2,\ \sigma_x^2)$.

Therefore finally,

$$B = \mathbb{E}[\mathbb{E}(SS^\top|X)] =$$
$$= \exp(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2) \begin{pmatrix} 1 & \mu_x + \beta_1\sigma_x^2 \\ \dots & (\mu_x + \beta_1\sigma_x^2)^2 + \sigma_w^2 \end{pmatrix}$$
$$+ \exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2)(e^{\beta_1^2\sigma_u^2} - 1) \begin{pmatrix} 1 & \mu_x + \beta_1(\sigma_w^2 + \sigma_x^2) \\ \dots & \left[\mu_x + \beta_1(\sigma_w^2 + \sigma_x^2)\right]^2 + \sigma_w^2 \end{pmatrix}$$
$$+ \exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2) \begin{pmatrix} 0 & 0 \\ 0 & \beta_1^2\sigma_u^4 \end{pmatrix}.$$

3.5 Factorization

In order to simplify the expression for A and B we introduce

$$g = \mu_x + \beta_1\sigma_x^2$$

and

$$R = \begin{pmatrix} 1 & g \\ 0 & 1 \end{pmatrix}.$$

Then

$$A = R^\top A_1 R$$

with

$$A_1 = \exp(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2) \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix},$$

and

$$B = R^\top B_1 R$$

with

$$\begin{aligned} B_1 &= \exp(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2) \begin{pmatrix} 1 & 0 \\ 0 & \sigma_w^2 \end{pmatrix} \\ &+ \exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2)(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1) \begin{pmatrix} 1 & \beta_1\sigma_w^2 \\ \beta_1\sigma_w^2 & \beta_1^2\sigma_w^4 + \sigma_w^2 \end{pmatrix} \\ &+ \exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2) \begin{pmatrix} 0 & 0 \\ 0 & \beta_1^2\sigma_u^4 \end{pmatrix}. \end{aligned}$$

Here we used the identity

$$\begin{pmatrix} 1 & 0 \\ g & 1 \end{pmatrix} \begin{pmatrix} 1 & x \\ x & x^2 + y \end{pmatrix} \begin{pmatrix} 1 & g \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & g + x \\ g + x & (g + x)^2 + y \end{pmatrix}.$$

3.6 Final expression for Σ_{cor}

From (7) we have

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

and hence

$$\begin{aligned} R\Sigma_{\mathrm{cor}}R^\top &= \mathrm{e}^{-(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2)} \cdot \begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} + \\ &+ \left(\mathrm{e}^{\beta_1^2\sigma_w^2} - \mathrm{e}^{\beta_1^2\sigma_x^2}\right) \begin{pmatrix} 1 & \frac{\sigma_w^2}{\sigma_x^2}\beta_1 \\ \cdots & \frac{\sigma_w^4}{\sigma_x^4}\beta_1^2 + \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} + \mathrm{e}^{\beta_1^2\sigma_x^2} \begin{pmatrix} 0 & 0 \\ 0 & \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2 \end{pmatrix}. \end{aligned} \quad (14)$$

4 An alternative estimator

For the alternative estimator we can assume the parameter set to be the entire plane $\mathbb{R}^2$.

4.1 Definition

Denote the conditional expectation of Y given W by

$$m(W, \beta) = \mathbb{E}(Y|W).$$

Then, see [1], [5],

$$m(W, \beta) = \exp\left(\beta_0 + \beta_1 \mu_{X|W} + \frac{1}{2}\beta_1^2 \sigma_{X|W}^2\right), \tag{15}$$

with

$$\mu_{X|W} = \frac{\mu_x \sigma_u^2 + W \sigma_x^2}{\sigma_w^2}, \qquad \sigma_{X|W}^2 = \frac{\sigma_u^2 \sigma_x^2}{\sigma_w^2}. \tag{16}$$

Another way to improve the naive estimator of β is to substitute in (1) the value $m(W_i, b)$ instead of $\lambda(X_i, b)$. We get the criterion function

$$Q_{\mathrm{al}}(b) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i \ln m(W_i, b) - m(W_i, b)\right), \quad b \in \Theta_\beta. \tag{17}$$

The alternative estimator $\hat{\beta}_{\mathrm{al}}$ is a measurable solution of

$$\hat{\beta}_{\mathrm{al}} \in \arg\min_{b \in \mathbb{R}^2} Q_{\mathrm{al}}(b).$$

4.2 Connection with naive estimator

Consider the transformation

$$\phi\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_0 + b_1 \frac{\sigma_u^2}{\sigma_w^2} + \frac{1}{2} b_1^2 \sigma_{X|W}^2 \\ b_1 \frac{\sigma_x^2}{\sigma_w^2} \end{pmatrix}.$$

It is a homeomorphism of $\mathbb{R}^2$ onto itself. The inverse transformation equals

$$\phi^{-1}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_0 - b_1 \frac{\sigma_u^2}{\sigma_x^2} - b_1^2 \frac{\sigma_u^2 \sigma_w^2}{2\sigma_x^2} \\ b_1 \frac{\sigma_w^2}{\sigma_x^2} \end{pmatrix}.$$

One can observe that $Q_{\rm al}(b) = Q_{\rm naive}(\phi(b))$. Then eventually $Q_{\rm al}(b)$ has a unique maximum point according to Statement 1, and it can be computed as

$$\hat\beta_{\rm al} = \phi^{-1}(\hat\beta_{\rm naive}). \tag{18}$$

The homeomorphism ϕ is differentiable and $\frac{\partial\phi(b)}{\partial b^\top}$ is nonsingular. Then

$$\frac{\partial Q_{\rm al}(b)}{\partial b^\top} = S_{\rm naive}(\phi(b))\frac{\partial\phi(b)}{\partial b^\top}.$$

Hence $\frac{\partial Q_{\rm al}(b)}{\partial b^\top}$ eventually has a unique zero $\hat\beta_{\rm al}$.

4.3 Consistency

Because of (18), continuity of ϕ^{-1}, Statement 2, and relation $\beta^*_{\rm naive} = \phi(\beta)$, the alternative estimator is strongly consistent.

4.4 Asymptotic normality

As $\frac{\partial Q_{\rm al}(\hat\beta_{\rm al})}{\partial\beta} = 0$ and due to the consistency property, we have, see [1], that $\sqrt{n}(\hat\beta_{\rm al} - \beta)$ converges in distribution to $N(0, \Sigma_{\rm al})$, where $\Sigma_{\rm al}$ can be found by a sandwich formula which is similar to (7).

Let

$$q_{\rm al}(b) = Y \ln m(W,b) - m(W,b), \tag{19}$$

$$S_{\rm al}(b) = \frac{\partial q_{\rm al}(b)}{\partial b} = \frac{Y - m(W,b)}{m(W,b)}\frac{\partial m(W,b)}{\partial b}, \quad S_{\rm al} = S_{\rm al}(\beta). \tag{20}$$

$$A_{\rm al} = -\mathbb{E}\frac{\partial S_{\rm al}}{\partial b^\top}, \qquad B_{\rm al} = \operatorname{cov} S_{\rm al}\ , \tag{21}$$

both taken at the point $b = \beta$. Then

$$\Sigma_{\rm al} = A_{\rm al}^{-1} B_{\rm al} A_{\rm al}^{-1} \tag{22}$$

(we will see below that $A_{\rm al}$ is a nonsingular matrix).

4.5 Computation of A_{al}

We have by (20)

$$S_{\mathrm{al}} = (Y - m(W,\beta)) \begin{pmatrix} 1 \\ \mu_{X|W} + \beta_1 \sigma^2_{X|W} \end{pmatrix} \qquad (23)$$

and by (15)

$$-\frac{\partial S_{\mathrm{al}}}{\partial \beta^\top} = m(W,\beta) \begin{pmatrix} 1 \\ \mu_{X|W} + \beta_1 \sigma^2_{X|W} \end{pmatrix} \left(1,\ \mu^2_{X|W} + \beta_1 \sigma^2_{X|W}\right) + E,$$

where

$$E = [-Y + m(W,\beta)] \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2_{X|W} \end{pmatrix}.$$

Hence

$$-\frac{\partial S_{\mathrm{al}}}{\partial \beta^\top} = m(W,\beta) \begin{pmatrix} 1 & \mu_{X|W} + \beta_1 \sigma^2_{X|W} \\ \cdots & (\mu_{X|W} + \beta_1 \sigma^2_{X|W}) + \sigma^2_{X|W} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & Y\sigma^2_{X|W} \end{pmatrix}.$$

We want to compute the expected value of $\frac{\partial S_{\mathrm{al}}}{\partial \beta^\top}$. We first find

$$\mathbb{E}E = \mathbb{E}[\mathbb{E}(E|W)] = 0.$$

In order to compute the expectation of the other terms we derive a general expression for all these terms.

As $\mu_{X|W} \sim N\left(\mu_x, \frac{\sigma_x^4}{\sigma_w^2}\right)$, see (16), we have, using (15), for any function f for which the following expectation exists

$$\mathbb{E}\left[f(\mu_{X|W}) m(W,\beta)\right] = \mathbb{E}\left[f(\mu_{X|W}) \exp\left(\beta_0 + \beta_1 \mu_{X|W} + \frac{1}{2}\beta_1^2 \sigma^2_{X|W}\right)\right] =$$

$$= \int_{-\infty}^{\infty} f(\mu_x + \tau) \exp\left(\beta_0 + \beta_1 \mu_x + \beta_1 \tau + \frac{1}{2}\beta_1^2 \sigma^2_{X|W}\right) \frac{\exp\left(-\frac{\tau^2 \sigma_w^2}{2\sigma_x^4}\right)}{\sqrt{2\pi}\frac{\sigma_x^2}{\sigma_w}} \mathrm{d}\tau =$$

$$= \int_{-\infty}^{\infty} f(\mu_x + \tau) \exp\left(\beta_0 + \beta_1 \mu_x + \frac{1}{2}\beta_1^2 \sigma_x^2\right) \frac{\exp\left(-\frac{1}{2}\frac{\sigma_w^2}{\sigma_x^4}\left(\tau - \beta_1 \frac{\sigma_x^4}{\sigma_w^2}\right)^2\right)}{\sqrt{2\pi}\sigma_x^2/\sigma_w} \mathrm{d}\tau =$$

$$= \mathbb{E}[f(\nu_1)] \exp\left(\beta_0 + \beta_1 \mu_x + \frac{1}{2}\beta_1^2 \sigma_x^2\right) \qquad (24)$$

with $\nu_1 \sim N\left(\mu_x + \beta_1 \frac{\sigma_x^4}{\sigma_w^2},\ \frac{\sigma_x^4}{\sigma_w^2}\right)$. Here we used (16) with $\sigma_u^2 = \sigma_w^2 - \sigma_x^2$. Summarizing these results, we finally get

$$A_{\text{al}} = -\mathbb{E}\frac{\partial S_{\text{al}}}{\partial \beta^\top}$$
$$= \exp\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\begin{pmatrix} 1 & \mu_x + \beta_1\sigma_x^2 \\ \mu_x + \beta_1\sigma_x^2 & (\mu_x + \beta_1\sigma_x^2)^2 + \frac{\sigma_x^4}{\sigma_w^2}\end{pmatrix} \quad (25)$$

4.6 Computation of B_{al}

As in the case of the corrected estimator, B_{al} can be computed as

$$B_{\text{al}} = \mathbb{E}(S_{\text{al}}S_{\text{al}}^\top).$$

We have from (23)

$$S_{\text{al}}S_{\text{al}}^\top = [Y - m(W,\beta)]^2 \begin{pmatrix} 1 & \mu_{X|W} + \beta_1\sigma_{X|W}^2 \\ \cdots & (\mu_{X|W} + \beta_1\sigma_{X|W}^2)^2 \end{pmatrix}.$$

Like in the Subsection 4.5 we have

$$\mathbb{E}\left[(Y - m(W,\beta))^2 \,\middle|\, W\right] = v(W,\beta) = m(W,\beta) + m^2(W,\beta)\left(\mathrm{e}^{\beta_1^2\sigma_{X|W}^2} - 1\right),$$

see [5], and therefore

$$\mathbb{E}\left(S_{\text{al}}S_{\text{al}}^\top \,\middle|\, W\right) = \left[m(W,\beta) + m^2(W,\beta)\left(\mathrm{e}^{\beta_1^2\sigma_{X|W}^2} - 1\right)\right] \times$$
$$\times \begin{pmatrix} 1 & \mu_{X|W} + \beta_1\sigma_{X|W}^2 \\ \cdots & (\mu_{X|W} + \beta_1\sigma_{X|W}^2)^2 \end{pmatrix}. \quad (26)$$

To evaluate the expectation of (26) we need again a general formula for certain expectations with arbitrary functions f. Assuming that the following expectation is finite, we have by (24) using the definition of m in (15):

$$\mathbb{E}[f(\mu_{X|W})m^2(W,\beta)] = \mathbb{E}\left[f(\mu_{X|W})m(W,2\beta)\,\mathrm{e}^{-\beta_1^2\sigma_{X|W}^2}\right] =$$
$$= \mathbb{E}[f(\nu_2)]\exp\left(2\beta_0 + 2\beta_1\mu_x + \frac{(2\beta_1)^2}{2}\sigma_x^2 - \beta_1^2\sigma_{X|W}^2\right) =$$
$$= \mathbb{E}[f(\nu_2)]\exp\left(2(\beta_0 + \beta_1\mu_x) + \beta_1^2\left(\sigma_x^2 + \frac{\sigma_x^4}{\sigma_w^2}\right)\right) \quad (27)$$

with $\nu_2 \sim N\left(\mu_x + 2\beta_1 \frac{\sigma_x^4}{\sigma_w^2},\ \frac{\sigma_x^4}{\sigma_w^2}\right)$. Here we used the identity $\sigma_x^2 - \sigma_{X|W}^2 = \frac{\sigma_x^4}{\sigma_w^2}$.

From (26), (24), and (27) we obtain

$$B_{\text{al}} = \mathbb{E}(S_{\text{al}} S_{\text{al}}^\top) = \exp\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\begin{pmatrix} 1 & \mu_x + \beta_1\sigma_x^2 \\ \cdots & (\mu_x + \beta_1\sigma_x^2)^2 \\ & +\frac{\sigma_x^4}{\sigma_w^2} \end{pmatrix} +$$

$$\exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2)\left(1 - \mathrm{e}^{-\beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\begin{pmatrix} 1 & \mu_x + \beta_1\left(\sigma_x^2 + \frac{\sigma_x^4}{\sigma_w^2}\right) \\ \cdots & \left[\mu_x + \beta_1\left(\sigma_x^2 + \frac{\sigma_x^4}{\sigma_w^2}\right)\right]^2 \\ & +\frac{\sigma_x^4}{\sigma_w^2} \end{pmatrix}. \tag{28}$$

4.7 Factorization

As in Subsections 3.5 and 1.6 we use

$$g = \mu_x + \beta_1\sigma_x^2, \qquad R = \begin{pmatrix} 1 & g \\ 0 & 1 \end{pmatrix}.$$

From (25) we have

$$A_{\text{al}} = R^\top A_2 R$$

with

$$A_2 = \exp\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_x^4}{\sigma_w^2} \end{pmatrix}$$

and from (28)

$$B_{\text{al}} = R^\top B_2 R$$

with

$$B_2 = \exp\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_x^4}{\sigma_w^2} \end{pmatrix} +$$

$$+\exp(2\beta_0 + 2\beta_1\mu_x + 2\beta_1^2\sigma_x^2)\left(1 - \mathrm{e}^{-\beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\begin{pmatrix} 1 & \beta_1\frac{\sigma_x^4}{\sigma_w^2} \\ \cdots & \left(\beta_1\frac{\sigma_x^4}{\sigma_w^2}\right)^2 + \frac{\sigma_x^4}{\sigma_w^2} \end{pmatrix}.$$

4.8 Final expression for $\Sigma_{\rm al}$

We have $R\Sigma_{\rm al}R^\top = A_2^{-1}B_2A_2^{-1}$.

Now,

$$A_2^{-1} = \exp\left[-\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\right]\begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix}.$$

and hence

$$R\Sigma_{\rm al}R^\top = \exp\left(-\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\beta_1^2\sigma_x^2\right)\right)\begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} + \\ + \mathrm{e}^{\beta_1^2\sigma_x^2}\left(1 - \mathrm{e}^{-\beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\begin{pmatrix} 1 & \beta_1 \\ \beta_1 & \beta_1^2 + \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix}. \quad (29)$$

5 Comparison of $\Sigma_{\rm cor}$ and $\Sigma_{\rm al}$

From (14) and (29) we get

$$R(\Sigma_{\rm cor} - \Sigma_{\rm al})R^\top = \mathrm{e}^{\beta_1^2\sigma_x^2}\left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1\right)\begin{pmatrix} 1 & \frac{\sigma_w^2}{\sigma_x^2}\beta_1 \\ \frac{\sigma_w^2}{\sigma_x^2}\beta_1 & \frac{\sigma_w^4}{\sigma_x^4}\beta_1^2 + \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} \\ - \mathrm{e}^{\beta_1^2\sigma_x^2}\left(1 - \mathrm{e}^{-\beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\begin{pmatrix} 1 & \beta_1 \\ \beta_1 & \beta_1^2 + \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} + \mathrm{e}^{\beta_1^2\sigma_x^2}\begin{pmatrix} 0 & 0 \\ 0 & \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2 \end{pmatrix}.$$

We factorize once more. Let

$$D = \begin{pmatrix} 1 & \beta_1 \\ 0 & 1 \end{pmatrix}.$$

Then

$$R(\Sigma_{\rm cor} - \Sigma_{\rm al})R^\top = \mathrm{e}^{\beta_1^2\sigma_x^2}\, D^\top F D,$$

where

$$F = \left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1\right)\begin{pmatrix} 1 & \frac{\sigma_u^2}{\sigma_x^2}\beta_1 \\ \frac{\sigma_u^2}{\sigma_x^2}\beta_1 & \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2 + \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} \\ - \left(1 - \mathrm{e}^{-\beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\begin{pmatrix} 1 & 0 \\ 0 & \frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2 \end{pmatrix}$$

which can also be written as

$$
\begin{aligned}
F &= \begin{pmatrix} \mathrm{e}^{\beta_1^2\sigma_u^2} - 2 + \mathrm{e}^{-\beta_1^2 \frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}} & \frac{\sigma_u^2}{\sigma_x^2}\beta_1\left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1\right) \\ \frac{\sigma_u^2}{\sigma_x^2}\beta_1\left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1\right) & \mathrm{e}^{\beta_1^2\sigma_u^2}\,\beta_1^2\frac{\sigma_u^4}{\sigma_x^4} \end{pmatrix} \\
&\quad + \begin{pmatrix} 0 & 0 \\ 0 & \left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 2 + \mathrm{e}^{-\beta_1^2 \frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}}\right)\frac{\sigma_w^2}{\sigma_x^4} \end{pmatrix} \\
&= F_1 + F_2. \qquad (30)
\end{aligned}
$$

If $\beta_1 = 0$, then $F = 0$. Suppose that $\beta_1 \neq 0$. Because

$$
\mathrm{e}^{\beta_1^2\sigma_u^2} - 2 + \mathrm{e}^{-\beta_1^2 \frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}} > \beta_1^2\sigma_u^2 - \beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2} = \beta_1^2\frac{\sigma_u^4}{\sigma_w^2} > 0,
$$

the matrix F_2 is positive semidefinite and the 1–1 entry of F_1 is positive. In addition the determinant of F_1 is positive. Indeed

$$
\begin{aligned}
\det F_1 = \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2 &\left[\mathrm{e}^{2\beta_1^2\sigma_u^2} - 2\,\mathrm{e}^{\beta_1^2\sigma_u^2} + \mathrm{e}^{\beta_1^2\frac{\sigma_u^4}{\sigma_w^2}} - \left(\mathrm{e}^{\beta_1^2\sigma_u^2} - 1\right)^2\right] \\
&= \frac{\sigma_u^4}{\sigma_x^4}\beta_1^2\left(\mathrm{e}^{\beta_1^2\frac{\sigma_u^4}{\sigma_w^2}} - 1\right) > 0.
\end{aligned}
$$

Therefore F is positive definite, and so is $\Sigma_{\text{cor}} - \Sigma_{\text{al}}$.

We proved the following result.

Theorem 1. Let $\sigma_u^2 > 0$ and let μ_x and σ_x^2 be known. If $\beta_1 = 0$, then $\Sigma_{\text{cor}} = \Sigma_{\text{al}}$; otherwise, if $\beta_1 \neq 0$, then $\Sigma_{\text{cor}} > \Sigma_{\text{al}}$.

6 Comparison of asymptotic expansions for Σ_{cor} and Σ_{al}

We show that for $\beta_1 \neq 0$

$$
\Sigma_{\text{cor}} - \Sigma_{\text{al}} = \sigma_u^4 S_4 + O(\sigma_u^6), \quad \sigma_u^2 \to 0, \qquad (31)
$$

where S_4 is positive definite.

We analyze the matrix F of (30) when $\sigma_u^2 \to 0$. First

$$\mathrm{e}^{\beta_1^2\sigma_u^2} - 2 + \mathrm{e}^{-\beta_1^2 \frac{\sigma_x^2\sigma_u^2}{\sigma_w^2}} = \beta_1^2\sigma_u^2 + \frac{1}{2}\beta_1^4\sigma_u^4 - \beta_1^2\frac{\sigma_x^2\sigma_u^2}{\sigma_w^2} + \frac{1}{2}\beta_1^4\frac{\sigma_x^4\sigma_u^4}{\sigma_w^4} + O(\sigma_u^6) =$$

$$= \beta_1^2\frac{\sigma_u^4}{\sigma_w^2} + \beta_1^4\sigma_u^4 + O(\sigma_u^6) = \beta_1^2\frac{\sigma_u^4}{\sigma_x^2} + \beta_1^4\sigma_u^4 + O(\sigma_u^6).$$

Similarly,

$$\mathrm{e}^{\beta_1^2\sigma_u^2} - 1 = \beta_1^2\sigma_u^2 + O(\sigma_u^4), \qquad \frac{\sigma_w^2}{\sigma_x^4} = \frac{1}{\sigma_x^2} + O(\sigma_u^2).$$

Therefore

$$F = \sigma_u^4\left[\begin{pmatrix} \frac{\beta_1^2}{\sigma_x^2} + \beta_1^4 & \frac{\beta_1^3}{\sigma_x^2} \\ \frac{\beta_1^3}{\sigma_x^2} & \frac{\beta_1^2}{\sigma_x^4} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{\beta_1^2}{\sigma_x^4} + \frac{\beta_1^4}{\sigma_x^2} \end{pmatrix}\right] + O(\sigma_u^6).$$

The second matrix in brackets is nonnegative semidefinite, the first matrix has positive diagonal elements and its determinant $\frac{\beta_1^4}{\sigma_x^6}$ is positive. Hence

$$F = \sigma_u^4 F_4 + O(\sigma_u^6),$$

where F_4 is a positive definite matrix. Let $S_4 = \mathrm{e}^{\beta_1^2\sigma_x^2} \cdot R^{-1}D^\top SDR^{-\top}$, then S_4 is a positive definite and (31) holds true.

We proved the following theorem.

Theorem 2. Let μ_x, σ_x^2 be known, and $\beta_1 \neq 0$. Then, as $\sigma_u^2 \to 0$,

$$\Sigma_{\mathrm{cor}} - \Sigma_{\mathrm{al}} = \sigma_u^4 \cdot S_4 + O(\sigma_u^6)$$

with a positive definite matrix S_4.

7 Quasi-likelihood estimator

7.1 The estimator

The structural quasi-likelihood (QL) estimator of β is defined as a measurable solution of the equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Y_i - m(W_i, b)}{v(W_i, b)} \frac{\partial m(W_i, b)}{\partial b} = 0. \qquad (32)$$

Here

$$v(W,\beta) = \mathrm{var}(Y|W).$$

The quasi-likelihood estimator is consistent by arguments, which are similar to the ones for a polynomial model, see [3], see also [4]. It is asymptotically normal, and the asymptotic covariance matrix is given by a corresponding sandwich formula [4], which is similar to (22),

$$\Sigma_{\mathrm{QL}} = A_{\mathrm{QL}}^{-1} B_{\mathrm{QL}} A_{\mathrm{QL}}^{-\top}.$$

Now, the problem is to compare the asymptotic covariance matrices of the three consistent estimators.

Using (20), we see that the alternative estimator eventually satisfies an equation which resembles (32):

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Y_i - m(W_i,b)}{m(W_i,b)} \, \frac{\partial m(W_i,b)}{\partial b} = 0. \tag{33}$$

Both QL and alternative estimators are not ML estimators, therefore a priori it is not clear which estimator is more efficient. We mention that if $\sigma_u^2 = 0$, i.e., the measurement error vanishes, the equations (32) and (33) coincide, because $Y|X$ has a Poisson distribution, and

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

7.2 Asymptotic optimality of quasi-likelihood estimator

We use the theorem 2.1 from [2] to prove that $\Sigma_{\mathrm{QL}} < \Sigma_{\mathrm{al}}$.

Consider the family of unbiased estimation functions, which linearly depend on Y:

$$\mathcal{H} = \left\{ (Y - m(W,b))\chi(W,b) \right\},$$

where χ is a smooth two-dimensional vector such that taking expectation and differentiation with respect to b are interchangeable and Sandwich-formula holds. QL and al-estimation functions belong to this family (for $\chi(W,b) = \frac{1}{v(W,b)}\frac{\partial m(W,b)}{\partial b}$ and $\chi(W,b) = \frac{1}{m(W,b)}\frac{\partial m(W,b)}{\partial b}$, respectively). For all $G \in \mathcal{H}$

$$\frac{\partial G(b)}{\partial b^\top} = -\chi(W,b)\frac{\partial m(W,b)}{\partial b^\top} + (Y - m(W,b))\frac{\partial \chi(W,b)}{\partial b^\top}.$$

As $m(W,\beta) = \mathbb{E}[Y|W]$ and $\chi(W,\beta)\frac{\partial m(W,\beta)}{\partial b}$ depend only on W, for $b = \beta$, the expectation of the second item is equal to 0. Then

$$\mathbb{E}\frac{\partial G}{\partial \beta^\top} = -\mathbb{E}\left(\chi(W,\beta)\frac{\partial m}{\partial \beta^\top}\right).$$

Let us compute $\mathbb{E}GS_{\mathrm{QL}}^\top$.

$$G(\beta)S_{\mathrm{QL}}^\top(\beta) = \chi(W,\beta)\frac{(Y - m(W,\beta))^2}{v(W,\beta)}\chi(W,\beta)\frac{\partial m(W,\beta)}{\partial \beta^\top}.$$

As $\mathbb{E}\left[(Y - m(W,\beta))^2 \middle| W\right] = v(W,\beta)$,

$$\mathbb{E}\left[G(\beta)S_{\mathrm{QL}}^\top(\beta) \middle| W\right] = \chi(W,\beta)\frac{\partial m(W,\beta)}{\partial \beta^\top},$$

$$\mathbb{E}\left(G(\beta)S_{\mathrm{QL}}^\top(\beta)\right) = \mathbb{E}\left(\chi(W,\beta)\frac{\partial m(W,\beta)}{\partial \beta^\top}\right).$$

So

$$\left(\mathbb{E}\frac{\partial G}{\partial \beta^\top}\right)^{-1}\mathbb{E}G(\beta)S_{\mathrm{QL}}^\top(\beta) = -\mathrm{I}$$

is a constant matrix.

By theorem 2.1 [2] the QL-estimation function is $\mathrm{O_F}$-optimal, see the definition of this optimality in [2]. Because $\Sigma_G = \left(\mathbb{E}G^{(\mathrm{s})}G^{(\mathrm{s})\top}\right)^{-1}$, where $G^{(\mathrm{s})}(\beta) = \mathbb{E}\frac{\partial G(\beta)^\top}{\partial \beta}(\mathbb{E}G(\beta)G(\beta)^\top)^{-1}G(\beta)$ is a standardized estimating function, the QL-estimator has the least asymptotic covariance matrix within $\mathcal{H}$, with respect to Loewner order (i.e., for two symmetrical matrices S_1 and S_2, $S_1 \le S_2$ iff $S_1 - S_2$ is positive semidefinite).

7.3 Comparison of asymptotic covariance matrices of alternative and QL estimators

Now we show, that $\Sigma_{\mathrm{al}} - \Sigma_{\mathrm{QL}}$ is strictly positive definite. Suppose that $\Sigma_{\mathrm{al}} - \Sigma_{\mathrm{QL}}$ is singular. Then so is $\Sigma_{\mathrm{QL}}^{-1} - \Sigma_{\mathrm{al}}^{-1} = \operatorname{cov} S_{\mathrm{QL}}^{(\mathrm{s})} - \operatorname{cov} S_{\mathrm{al}}^{(\mathrm{s})}$. But the first formula of the proof implies

$$\operatorname{cov}(S_{\mathrm{QL}}^{(\mathrm{s})} - S_{\mathrm{al}}^{(\mathrm{s})}) = \operatorname{cov} S_{\mathrm{QL}}^{(\mathrm{s})} - \operatorname{cov} S_{\mathrm{al}}^{(\mathrm{s})}. \tag{34}$$

Indeed, in Subsection 7.2 we obtained that $\left(\mathbb{E}\frac{\partial S_{\rm al}(\beta)}{\partial\beta^\top}\right)^{-1}\mathbb{E}S_{\rm al}(\beta)S_{\rm QL}^\top(\beta) = \left(\mathbb{E}\frac{\partial S_{\rm QL}(\beta)}{\partial\beta^\top}\right)^{-1}\mathbb{E}S_{\rm QL}(\beta)S_{\rm QL}^\top(\beta)$. Hence, with $\mathbb{E}S_{\rm al}(\beta) = \mathbb{E}S_{\rm QL}(\beta) = 0$, and using the definition of standardized estimator, we obtain $\mathbb{E}S_{\rm al}S_{\rm QL}^{(\rm s)\top} = A_{\rm al}$, $\mathbb{E}S_{\rm al}^{(\rm s)}S_{\rm QL}^{(\rm s)\top} = A_{\rm al}^\top B_{\rm al}^{-1}A_{\rm al} = \operatorname{cov} S_{\rm al}^{(\rm s)}$. Hence (34) holds.

Then the vector $S_{\rm QL}^{(\rm s)} - S_{\rm al}^{(\rm s)}$ with probability 1 lies in own non-stochastic subspace, and there exists a linear functional $f \neq 0$ such that $f(S_{\rm QL}^{(\rm s)} - S_{\rm al}^{(\rm s)}) = 0$ a.s. Denote

$$f_1 = f\cdot\left(\mathbb{E}\frac{\partial S_{\rm QL}^\top}{\partial\beta}\right)(\operatorname{cov} S_{\rm QL})^{-1},$$

$$f_2 = f\cdot\left(\mathbb{E}\frac{\partial S_{\rm al}^\top}{\partial\beta}\right)(\operatorname{cov} S_{\rm al})^{-1}.$$

Then $f_1(S_{\rm QL}) - f_2(S_{\rm al}) = 0$ a.s., and a.s.

$$\frac{1}{v(W,\beta)}(Y - m(W,\beta))f_1\left(\frac{\partial m}{\partial\beta}\right) = \frac{1}{m(W,\beta)}(Y - m(W,\beta))f_2\left(\frac{\partial m}{\partial\beta}\right).$$

Now choose coefficients $a_k,\ c_k$ such that $f_k\begin{pmatrix}1\\ \mu_{X|W} + \beta_1\sigma^2_{X|W}\end{pmatrix} = a_k + c_kW$, $k = 1,2$. As

$$\frac{\partial m(W,\beta)}{\partial b} = m(W,\beta)\begin{pmatrix}1\\ \mu_{X|W} + \beta_1\sigma^2_{X|W}\end{pmatrix},$$

$$\frac{m(W,\beta)}{v(W,\beta)}(Y - m(W,\beta))\,(a_1 + c_1W) = (Y - m(W,\beta))\,(a_2 + c_2W)\ \text{a.s.}$$

So almost surely

$$a_1 + c_1W = \left(1 + m(W,\beta)(\exp(\beta_1^2\sigma^2_{X|W}) - 1)\right)(a_2 + c_2W).$$

As the functions in this equality are continuous and the support of W is $\mathbb{R}$, the formula holds for all real non-stochastic W. Considering a behaviour at infinity, he get $a_2 = c_2 = 0$, then $f_2 = 0$, $f = 0$. So we get contradiction. We have proved that $\Sigma_{\rm al} - \Sigma_{\rm QL}$ is a positive definite nonsingular matrix.

8 Conclusion

In the case of known μ_x and σ_x^2, we compared Σ_{cor} and Σ_{al}. This result holds true for arbitrary σ_u^2.

Theorem 2 shows that $\Sigma_{\text{cor}} = \Sigma_{\text{al}} + O(\sigma_u^4)$, as $\sigma_u^2 \to 0$. This implies that $\Sigma_{\text{cor}} \approx \Sigma_{\text{al}} \approx \Sigma_{\text{QL}}$ up to $O(\sigma_u^4)$, because earlier it was shown that $\Sigma_{\text{QL}} = \Sigma_{\text{cor}} + O(\sigma_u^4)$, see [4].

Compare Σ_{al} with the asymptotic covariance matrix Σ_{naive} for ordinary (naive) MLE. It was shown in [4] that

$$\Sigma_{\text{cor}} - \Sigma_{\text{naive}} = \frac{2\sigma_u^2}{\sigma_w^2} \exp\left(-\left(\beta_0 + \beta_1\mu_x + \frac{1}{2}\sigma_x^2\beta_1^2\right)\right) \cdot \begin{pmatrix} g^2 & -g \\ -g & 1 \end{pmatrix} + \\ + O(\sigma_u^4), \quad \sigma_u^2 \to 0,$$

where $g = \mu_x + \sigma_x^2\beta_1$. The same expansion holds true for $\Sigma_{\text{al}} - \Sigma_{\text{naive}}$. This shows that in a sense Σ_{al} is larger than Σ_{naive}, for small σ_u^2. (But we mention that $\hat{\beta}_{\text{naive}}$ is inconsistent, while all the three remaining estimators are consistent).

The next question remains open. To compare Σ_{QL}, Σ_{al}, and Σ_{cor}, when μ_x, σ_x^2 are unknown and we plug–in $\hat{\mu}_x = \hat{\mu}_w$ and $\hat{\sigma}_x^2 = \hat{\sigma}_w^2 - \sigma_u^2$ instead of μ_x and σ_x^2, while we construct $\hat{\beta}_{\text{QL}}$ and $\hat{\beta}_{\text{al}}$ (here $\hat{\mu}_w$ and $\hat{\sigma}_w^2$ are sample mean and sample variance of $W_1,\ W_2, \ldots, W_n$). In this situation an additional term in Σ_{QL} and Σ_{al} appears, see [3]. But Σ_{cor} remains unchanged, because $\hat{\beta}_{\text{cor}}$ does not use the distribution of X.

Acknowledgement

We are grateful to Hans Schneeweiss for fruitful discussions. Support by Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

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

[2] Heyde C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation.* Springer-Verlag, New York, Berlin, Heidelberg.

[3] Kukush, A., and Schneeweiss, H. (2000). *A Comparison of Asymptotic Covariance Matrices of Adjusted Least Squares and Structural Least Squares in Error Ridden Polynomial Regression.* Discussion Paper 218, Sonderforschungsbereich 386, University of Munich.

[4] Kukush, A., Schneeweiss, H. and Wolf, R. (2001) *Three Estimators for the Poisson Regression Model with Measurement Errors.* Discussion Paper 243, Sonderforschungsbereich 386, University of Munich.

[5] Thamerus, M. (1998). *Different Nonlinear Regression Models with Incorrectly Observed Covariates.* In R. Galata and H. Küchenhoff (Eds.): Econometrics in Theory and Practice, Festschrift for Hans Schneeweiss. Physica, Heidelberg, New York.