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 naïve (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 $\beta = (\beta_0, \beta_1)^T$ is the parameter of interest. Let $W = X + U$ be a surrogate covariate, where $U$ is independent of $X$ and $Y$, and $U \sim N(0, \sigma_u^2)$. $\sigma_u^2$ is positive and known.

We observe independent realizations $(Y_i, W_i)$, $i = 1, \ldots, n$. We suppose that the parameter set $\Theta_\beta$ is a compact set in $\mathbb{R}^2$, and the true value
\( \beta \) is an interior point of \( \Theta_\beta \); only for the naïve 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 naïve estimator (which is inconsistent) and of the corrected and quasi-likelihood estimator (these two are consistent). The comparison was made for small \( \sigma^2_\varepsilon \) and unknown \( \mu_\varepsilon \) and \( \sigma^2_\varepsilon \). In the present paper we compare the two consistent estimators for arbitrary \( \sigma^2_\varepsilon \), but we suppose that \( \mu_\varepsilon \) and \( \sigma^2_\varepsilon \) are known.

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

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

In Section 2 the naïve 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 naïve estimator

For the naïve 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

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

with \( b = (b_0, b_1)^T \in \Theta_\beta \) 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 naïve 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 \( \beta_{\text{naive}} = \infty \). The resulting estimator \( \beta_{\text{naive}} \) coincides with the ML estimator if \( W \) is measured without errors, i.e., if \( W = X \), and in that case \( \beta_{\text{naive}} \) 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

\[ S_{\text{naive}} = \frac{\partial Q_{\text{naive}}}{\partial b} = \sum_{i=1}^{n} (Y_i - \lambda(b, W_i)) \left( \frac{1}{W} \right) \]

is an equivalent way to define \( \beta_{\text{naive}} \).

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 \( n_2 \leq n \) occurs, then almost surely \( Q_{\text{naive}} \) has a unique maximum point. It is the unique zero of \( S_{\text{naive}} \).

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

**Definition 1.** Let \( Z \) be a random variable on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), and \( Z_1, Z_2, \ldots \) be independent observations on \( Z \). A sequence of statements \( A_1(Z_1), A_2(Z_1, Z_2), \ldots \) holds eventually iff
\[
\exists \Omega_0, \ P(\Omega_0) = 1 \forall \omega \in \Omega_0 \ \exists N(\omega) \ \forall n \geq N(\omega) : A_n(Z_1, \ldots, Z_n)
\]
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 \( n \to \infty \),
\[
\frac{1}{n} Q_{\text{naive}}(b) \to Q_{\text{naive}}^\infty(b) \tag{2}
\]
uniformly on any compact set, where
\[
Q_{\text{naive}}^\infty(b) = \mathbb{E}[Y(b_0 + b_1 W) - \lambda(b, W)] = \mathbb{E}[\mathbb{E}[Y_i(b_0 + b_1 W) | X]] - \mathbb{E}\lambda(b, W) = \mathbb{E}[\lambda(\beta, X)(b_0 + b_1 X)] - \mathbb{E}\lambda(b, W). \tag{3}
\]

Now, remember that \( W \sim N(\mu_x, \sigma_w^2) \) and consider, for arbitrary function \( f \), the following expectation, assuming its existence:
\[
\mathbb{E}[f(W) \lambda(W, b)] = \int_{-\infty}^{\infty} f(\mu_x + \tau) e^{b_0 + b_1 \mu_x + b_1 \tau} e^{-\frac{\tau^2}{2\sigma_w^2}} \frac{1}{\sqrt{2\pi\sigma_w}} d\tau
\]
\[
= \int_{-\infty}^{\infty} f(\mu_x + \tau) e^{b_0 + b_1 \mu_x + \frac{b_1^2}{2} \sigma_w^2} e^{-\frac{(\tau - \overline{\tau})^2}{2\sigma_w^2}} \frac{1}{\sqrt{2\pi\sigma_w}} d\tau
\]
\[
eq e^{b_0 + b_1 \mu_x + \frac{b_1^2}{2} \sigma_w^2} \mathbb{E}f(W + b_1 \sigma_w^2). \tag{4}
\]

Analogously, remembering \( X \sim N(\mu_x, \sigma_x^2) \), we can prove that
\[
\mathbb{E}[f(X) \lambda(X, \beta)] = e^{b_0 + \beta_1 \mu_x + \frac{\beta_1^2}{2} \sigma_x^2} \mathbb{E}f(X + \beta_1 \sigma_x^2). \tag{5}
\]

Applying (4) and (5) to (3) we get
\[
Q_{\text{naive}}^\infty(b) = e^{b_0 + \beta_1 \mu_x + \frac{\beta_1^2}{2} \sigma_x^2} (b_0 + b_1 \mu_x + b_1 \beta_1 \sigma_x^2) - e^{b_0 + b_1 \mu_x + \frac{b_1^2}{2} \sigma_w^2},
\]

4
\( Q_{\text{naive}}(b) \) is a concave function. So it attains its maximum at the point where its derivative is equal to zero. We have

\[
\frac{\partial Q_{\text{naive}}(b)}{\partial b} = e^{b_0 + \beta_1 \mu_x + \frac{1}{2} \beta_2 \sigma_x^2} \left( \frac{1}{\mu_x + \beta_1 \sigma_x^2} \right) - e^{b_0 + b_1 \mu_x + \frac{1}{2} \beta_2 \sigma_x^2} \left( \frac{1}{\mu_x + b_1 \sigma_x^2} \right),
\]

and \( \frac{\partial Q_{\text{naive}}(b)}{\partial b} = 0 \) holds if and only if

\[
\begin{align*}
\beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_2 \sigma_x^2 &= b_0 + b_1 \mu_x + \frac{1}{2} \beta_2 \sigma_x^2, \\
\beta_1 \mu_x + \beta_1 \sigma_x^2 &= b_1 \mu_x + b_1 \sigma_x^2.
\end{align*}
\]

This system has the only solution \( \beta_{\text{naive}}^* = \left( \beta_0 + \beta_1 \frac{\sigma_x^2}{\sigma_u^2} + \beta_1 \frac{\sigma_u^2}{\sigma_x^2}, \beta_1 \frac{\sigma_x^2}{\sigma_u^2} \right) \). So \( Q_{\text{naive}}^* \) has unique maximum point \( \beta_{\text{naive}}^* \).

Let \( \epsilon > 0 \). Denote by \( K \) the circumference of radius \( \epsilon \) and centre at \( \beta_{\text{naive}}^* \). We have \( \sup_{K} Q_{\text{naive}} < Q_{\text{naive}}(\beta_{\text{naive}}^*) \). Because of (2) \( \sup_{K} Q_{\text{naive}} < Q_{\text{naive}}(\beta_{\text{naive}}^*) \) eventually. Because of concavity of \( Q_{\text{naive}} \), the maximum point \( \beta_{\text{naive}}^* \) eventually lies inside \( K \). Therefore \( \beta_{\text{naive}}^* \) tends to \( \beta_{\text{naive}}^* \) almost surely.

**Statement 2.** The naive estimator is convergent, as \( n \to \infty \),

\[
\beta_{\text{naive}} \xrightarrow{(a.s.)} \left( \frac{\beta_0 + \beta_1 \frac{\sigma_x^2}{\sigma_u^2} + \beta_1 \frac{\sigma_u^2}{\sigma_x^2}}{\beta_1 \frac{\sigma_x^2}{\sigma_u^2}} \right)
\]

**3 The corrected estimator**

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

**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

\[
E \left( q(W, Y, b) \mid Y, X \right) = q_{\text{naive}}(X, Y, b),
\]
see [1]. Chapter 6. Such function is given by

\[ q = Y \ln \lambda(W, b) - \exp \left(-\frac{1}{2} \beta_1^2 \sigma_u^2 \right) \lambda(W, b). \]

The corresponding corrected criterion function is

\[ Q_{\text{cor}}(b) = \frac{1}{n} \sum_{i=1}^{n} \left( Y_i \ln \lambda(W_i, b) - e^{-\frac{1}{2} \beta_2^2 \sigma_u^2} \lambda(W_i, b) \right), \]

and the estimator \( \beta_{\text{cor}} \) is a measurable solution of

\[ \tilde{\beta}_{\text{cor}} \in \arg \max_{b \in \mathbb{R}^2} Q_{\text{cor}}(b). \]

\( Q_{\text{cor}} \) does not attain a maximum on the entire plane \( \mathbb{R}^2 \) because it is unbounded from above.) The estimator is strongly consistent, and \( \sqrt{n}(\beta_{\text{cor}} - \beta) \) converges in distribution to \( N(0, \Sigma_{\text{cor}}) \), where \( \Sigma_{\text{cor}} \) 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

\[ A = -E \frac{\partial S}{\partial b^T}, \quad B = \text{cov} S, \]

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

\[ \Sigma_{\text{cor}} = A^{-1} BA^{-T}. \]

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

### 3.2 Computation of \( A \)

We have

\[ S = \begin{pmatrix}
  Y - e^{-\frac{1}{2} \beta_1^2 \sigma_u^2} \lambda(W, \beta) \\
  YW - (W - \beta_1 \sigma_u^2) e^{-\frac{1}{2} \beta_2^2 \sigma_u^2} \lambda(W, \beta)
\end{pmatrix}, \]
\[-\frac{\partial S}{\partial \beta} = e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda(W, \beta) \left( \frac{1}{W - \beta_1 \sigma_u^2} \left( W - \beta_1 \sigma_u^2 \right) \right). \]

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

\[A = e^{\alpha + \beta_1 \mu + \frac{1}{2} \beta^2 \sigma_y^2} \mathbb{E} \left( \frac{1}{\omega - \beta_1 \sigma_u^2} \left( \omega - \beta_1 \sigma_u^2 \right) \right) \]

\[= e^{\alpha + \beta_1 \mu + \frac{1}{2} \beta^2 \sigma_y^2} \left( \frac{1}{\mu_x + \beta_1 \sigma_x^2} \left( \mu_x + \beta_1 \sigma_x^2 \right) \right). \] (8)

3.3 Computation of \(A^{-1}\)

The determinant of the latter matrix equals \(\sigma_x^2\). Therefore

\[A^{-1} = \frac{1}{\sigma_x^2} e^{-\left(\alpha + \beta_1 \mu + \frac{1}{2} \beta^2 \sigma_y^2\right)} \left( \begin{array}{cc}
(\beta_1 \sigma_x^2 + \mu_x)^2 + \sigma_x^2 & - (\beta_1 \sigma_x^2 + \mu_x) \\
- (\beta_1 \sigma_x^2 + \mu_x) & 1
\end{array} \right). \]

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 \(\lambda = \lambda(W, \beta)\):

\[SS^\top = \left( \begin{array}{cc}
Y^2 - 2 e^{-\frac{1}{2} \beta^2 \sigma_y^2} Y \lambda + e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda^2 & Y^2 W - W (2 W - \beta_1 \sigma_u^2) e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda \\
Y^2 W - W (2 W - \beta_1 \sigma_u^2) e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda + (W - \beta_1 \sigma_u^2) e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda^2 & Y^2 W^2 - 2 W (W - \beta_1 \sigma_u^2) e^{-\frac{1}{2} \beta^2 \sigma_y^2} \lambda^3
\end{array} \right). \] (9)

We observe that \(ES = 0\) and so \(B\) can be written as

\[B = ES^\top. \] (10)

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

\[E[f(W)|X] = \int_{-\infty}^{\infty} f(X + u) e^{-\frac{u^2}{2\sigma_y^2}} \frac{1}{\sqrt{2\pi\sigma_u}} du \] (11)
because $W|X \sim N(X, \sigma^2_u)$; next

$$
\begin{align*}
\mathbb{E} \left[ f(W) \lambda(W, \beta) e^{-\frac{1}{2} \beta^2 \sigma^2_u} \right] X &= \\
&= \int_{-\infty}^{\infty} f(X + u) e^{\beta_0 + \beta_1 X + \beta_2 u - \frac{1}{2} \beta^2 \sigma^2_u} e^{-\frac{u^2}{2 \sigma_u^2}} \frac{1}{\sqrt{2\pi \sigma_u}} \, du = \\
&= \int_{-\infty}^{\infty} f(X + u) e^{\beta_0 + \beta_1 X} e^{-\frac{u^2 + \beta^2 \sigma^2_u}{2 \sigma_u^2}} \frac{1}{\sqrt{2\pi \sigma_u}} \, du = \mathbb{E} [f(W_1)|X] \lambda(X, \beta),
\end{align*}
$$

(12)

where $W_1|X \sim N(X + \beta_1 \sigma^2_u, \sigma^2_u)$.

At last, applying (12) with $\beta$ replaced by $2\beta$, we have

$$
\begin{align*}
\mathbb{E} \left[ f(W) \lambda^2(W, \beta) e^{-\frac{1}{2} \beta^2 \sigma^2_u} \right] X &= \\
&= e^{\beta_0^2 \sigma_u^2} \mathbb{E} \left[ f(W) \lambda(W, 2\beta) e^{-\frac{1}{2}(2\beta)^2 \sigma^2_u} \right] X = \\
&= \mathbb{E} [f(W_2)|X] \lambda(X, 2\beta) e^{\beta_1^2 \sigma^2_u} \\
&= \mathbb{E} [f(W_2)|X] \lambda^2(X, \beta) e^{\beta_1 \sigma^2_u},
\end{align*}
$$

(13)

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

$$
\mathbb{E} (SS^T|X, Y) = \\
\begin{pmatrix}
Y^2 - 2\lambda Y + e^{\beta_0^2 \sigma_u^2} \lambda^2 & Y^2 X - Y \lambda(2X + \beta_1 \sigma^2_u) + e^{\beta_0^2 \sigma_u^2} \lambda^2 (X + \beta_1 \sigma^2_u) \\
\vdots & \vdots \\
Y^2 (X^2 + \sigma^2_u) - 2(X + \beta_1 \sigma^2_u) \lambda Y - 2\sigma^2_u \lambda Y + e^{\beta_0^2 \sigma_u^2} \lambda^2 (X + \beta_1 \sigma^2_u)^2 + \sigma^2_u)
\end{pmatrix}.
$$

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 $\lambda = \lambda(X, \beta)$,

\[
\begin{align*}
\mathbb{E}[SS^\top|X] &= \lambda \left( \begin{array}{c} X \\ X^2 + \sigma_u^2 \\
\end{array} \right) \\
&\quad + \lambda^2 \left( \begin{array}{c} -1 + e^{\beta_2 \sigma_u^2} \\
- X - \beta_1 \sigma_u^2 + e^{\beta_1 \sigma_u^2}(X + \beta_1 \sigma_u^2) \\
\vdots \\
(X^2 + \sigma_u^2) - 2(X + \beta_1 \sigma_u^2)X - 2\sigma_u^2 + e^{\beta_1 \sigma_u^2}((X + \beta_1 \sigma_u^2)^2 + \sigma_u^2) \\
\end{array} \right) \\
&= \lambda \left( \begin{array}{c} X \\
X^2 + \sigma_u^2 \\
\end{array} \right) + (e^{\beta_2 \sigma_u^2} - 1) \lambda^2 \left( \begin{array}{c} X + \beta_1 \sigma_u^2 \\
(X + \beta_1 \sigma_u^2)^2 + \sigma_u^2 \\
\end{array} \right) \\
&\quad + \lambda^2 \left( \begin{array}{cc} 0 & 0 \\
0 & \beta_1^4 \sigma_u^4 \\
\end{array} \right). 
\end{align*}
\]

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

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

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

Therefore finally,

\[
B = \mathbb{E}[\mathbb{E}(SS^\top|X)] = 
= \exp(\beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_1^2 \sigma_x^2) \left( \begin{array}{c} 1 \\
\mu_x + \beta_1 \sigma_x^2 \\
\mu_x + \beta_1 \sigma_x^2 + \sigma_u^2 \\
\end{array} \right) \\
+ \exp(2\beta_0 + 2\beta_1 \mu_x + 2\beta_1^2 \sigma_u^2)(e^{\beta_2 \sigma_u^2} - 1) \left( \begin{array}{c} 1 \\
\mu_x + \beta_1 \sigma_u^2 + \sigma_u^2 \\
\mu_x + \beta_1 \sigma_u^2 + \sigma_u^2 \\
\end{array} \right) \\
+ \exp(2\beta_0 + 2\beta_1 \mu_x + 2\beta_1^2 \sigma_u^2) \left( \begin{array}{cc} 0 & 0 \\
b_{\beta_1}^4 \sigma_u^4 & 0 \\
\end{array} \right). 
\]

### 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
\[ A_1 = \exp(\beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_2^2 \sigma_x^2) \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}, \]
and
\[ B = R^\top B_1 R \]
with
\[
B_1 = \exp(\beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_2^2 \sigma_x^2) \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix} + \exp(2\beta_0 + 2\beta_1 \mu_x + 2\beta_2^2 \sigma_x^2)(e^{\beta_2^2 \sigma_x^2} - 1) \begin{pmatrix} 1 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \beta_1^2 \sigma_x^4 + \sigma_x^2 \end{pmatrix} + \exp(2\beta_0 + 2\beta_1 \mu_x + 2\beta_2^2 \sigma_x^2) \begin{pmatrix} 0 & 0 \\ 0 & \beta_1^2 \sigma_x^4 \end{pmatrix}.\]

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 $\Sigma_{cor}$

From (7) we have
\[ R\Sigma_{cor} R^\top = A_1^{-1} B_1 A_1^{-1}, \]
and hence
\[
R\Sigma_{cor} R^\top = e^{-\frac{1}{2}(\beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_2^2 \sigma_x^2)} \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix} + \\
\bigg( e^{\beta_2^2 \sigma_x^2} - e^{\beta_2^2 \sigma_x^2} \bigg) \begin{pmatrix} \frac{1}{2} \sigma_x^2 \beta_1 \\ \frac{1}{2} \sigma_x^2 \beta_1 \end{pmatrix} + e^{\beta_2^2 \sigma_x^2} \begin{pmatrix} 0 & 0 \\ 0 & \beta_1^2 \sigma_x^4 \end{pmatrix}.
\]

(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],

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

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

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

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

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

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

4.2 Connection with naive estimator

Consider the transformation

$$\phi \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_0 + b_1 \frac{\sigma_x^2}{\sigma_u^2} + \frac{1}{2} b_1 \sigma_{X|W}^2 \\ b_1 \frac{\sigma_x^2}{\sigma_u^2} \end{pmatrix}. \tag{18}$$

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

$$\phi^{-1} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_0 - b_1 \frac{\sigma_x^2}{2 \sigma_u^2} - b_1 \frac{\sigma_{X|W}^2}{\sigma_u^2} \\ b_1 \frac{\sigma_x^2}{2 \sigma_u^2} \end{pmatrix}. \tag{19}$$
One can observe that \( Q_{\alpha}(b) = Q_{\text{naive}}(\phi(b)) \). Then eventually \( Q_{\alpha}(b) \) has a unique maximum point according to Statement 1, and it can be computed as
\[
\beta_{\alpha} = \phi^{-1}(\beta_{\text{naive}}),
\]
(18)
The homeomorphism \( \phi \) is differentiable and \( \frac{\partial \phi(b)}{\partial b} \) is nonsingular. Then
\[
\frac{\partial Q_{\alpha}(b)}{\partial b} = S_{\text{naive}}(\phi(b)) \frac{\partial \phi(b)}{\partial b}.
\]
Hence \( \frac{\partial Q_{\alpha}(b)}{\partial b} \) eventually has a unique zero \( \tilde{\beta}_{\alpha} \).

4.3 Consistency

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

4.4 Asymptotic normality

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

Let
\[
q_{\alpha}(b) = Y \ln m(W, b) - m(W, b),
\]
(19)
\[
S_{\alpha}(b) = \frac{\partial q_{\alpha}(b)}{\partial b} = \frac{Y - m(W, b)}{m(W, b)} \frac{\partial m(W, b)}{\partial b}, \quad S_{\alpha} = S_{\alpha}(\beta).
\]
(20)
\[
A_{\alpha} = -E \frac{\partial S_{\alpha}}{\partial b}, \quad B_{\alpha} = \text{cov} S_{\alpha}, \quad (21)
\]
both taken at the point \( b = \beta \). Then
\[
\Sigma_{\alpha} = A_{\alpha}^{-1} B_{\alpha} A_{\alpha}^{-1}
\]
(22)
(we will see below that \( A_{\alpha} \) is a nonsingular matrix).
4.5 Computation of $A_{\alpha}$

We have by (20)

$$S_{\alpha} = (Y - m(W, \beta)) \left( \begin{matrix} 1 \\ \mu_{X|W} + \beta_1 \sigma_{X|W}^2 \end{matrix} \right)$$

(23)

and by (15)

$$-\frac{\partial S_{\alpha}}{\partial \beta} = m(W, \beta) \left( \begin{matrix} 1 \\ \mu_{X|W} + \beta_1 \sigma_{X|W}^2 \end{matrix} \right) \left( \begin{matrix} 1 & 0 \\ \frac{\mu_{X|W}^2 + \beta_1 \sigma_{X|W}^2}{\sigma_{X|W}^2} & 1 \end{matrix} \right) + E,$$

where

$$E = [-Y + m(W, \beta)] \left( \begin{matrix} 0 & 0 \\ 0 & \sigma_{X|W}^2 \end{matrix} \right).$$

Hence

$$-\frac{\partial S_{\alpha}}{\partial \beta} = m(W, \beta) \left( \begin{matrix} 1 \\ \mu_{X|W} + \beta_1 \sigma_{X|W}^2 \end{matrix} \right) \left( \begin{matrix} 1 & 0 \\ \mu_{X|W}^2 + \beta_1 \sigma_{X|W}^2 & \sigma_{X|W}^2 \end{matrix} \right) - \left( \begin{matrix} 0 \\ 0 \\ Y \sigma_{X|W}^2 \end{matrix} \right).$$

We want to compute the expected value of $\frac{\partial S_{\alpha}}{\partial \beta}$. We first find

$$EE = E[E(E|W)] = 0.$$

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

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

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

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

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

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

(24)
with \( y_1 \sim N(\mu_x + \beta_1 \sigma_x^2, \frac{\sigma^2}{\sigma_x^2}) \). Here we used (16) with \( \sigma_u^2 = \sigma_w^2 - \sigma_x^2 \).

Summarizing these results, we finally get

\[
A_{al} = -E \frac{\partial S_{al}}{\partial \beta_1} = \exp \left( \beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_1^2 \sigma_x^2 \right) \left( \frac{1}{\mu_x + \beta_1 \sigma_x^2} \mu_x + \beta_1 \sigma_x^2 \right) \left( \frac{1}{\mu_x + \beta_1 \sigma_x^2 + \sigma_u^2} \right) \] (25)

### 4.6 Computation of \( B_{al} \)

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

\[
B_{al} = E(S_{al} S_{al}^T).
\]

We have from (23)

\[
S_{al} S_{al}^T = \left[ Y - m(W, \beta) \right]^2 \left( \begin{array}{c}
\mu_{X|W} + \beta_1 \sigma_{X|W}^2 \\
\vdots \\
\mu_{X|W} + \beta_1 \sigma_{X|W}^2 \\
\end{array} \right) \left( \mu_{X|W} + \beta_1 \sigma_{X|W}^2 \right)^2 \]

Like in the Subsection 4.5 we have

\[
E\left[(Y - m(W, \beta))^2 \mid W \right] = \nu(W, \beta) = m(W, \beta) + m^2(W, \beta) \left( e^{\beta_1^2 \sigma_{X|W}^2} - 1 \right),
\]

see [5], and therefore

\[
E\left(S_{al} S_{al}^T \mid W \right) = \left[ m(W, \beta) + m^2(W, \beta) \left( e^{\beta_1^2 \sigma_{X|W}^2} - 1 \right) \right] \times \\
\times \left( \begin{array}{c}
\mu_{X|W} + \beta_1 \sigma_{X|W}^2 \\
\vdots \\
\mu_{X|W} + \beta_1 \sigma_{X|W}^2 \\
\end{array} \right) \left( \mu_{X|W} + \beta_1 \sigma_{X|W}^2 \right)^2. \] (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):

\[
E[f(\mu_{X|W}) m^2(W, \beta)] = E\left[f(\mu_{X|W}) m(W, 2\beta) e^{-\beta_1^2 \sigma_{X|W}^2} \right] = \\
= E[f(n_2)] \exp \left( 2\beta_0 + 2\beta_1 \mu_x + \frac{(2\beta_1)^2 \sigma_x^2}{2} - \beta_1^2 \sigma_{X|W}^2 \right) = \\
= E[f(n_2)] \exp \left( 2(\beta_0 + \beta_1 \mu_x) + \beta_1^2 \left( \sigma_x^2 + \frac{\sigma_u^2}{\sigma_w^2} \sigma_{X|W}^2 \right) \right) \] (27)
with $\nu_2 \sim N \left( \mu_x + 2\beta_1 \frac{\sigma^2}{\sigma^2_x}, \frac{\sigma^4}{\sigma^2_x} \right)$. Here we used the identity $\sigma^2_x - \sigma^2_{X|Y} = \frac{\sigma^4}{\sigma^2_x}$.

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

$$B_{al} = E(S_a S^\top_a) = \exp \left( \beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_1^2 \sigma_x^2 \right) \left( \begin{array}{cccc} 1 & \mu_x + \beta_1 \sigma_x^2 & \cdots & (\mu_x + \beta_1 \sigma_x^2)^2 \\
 & + \frac{\beta_1^4 \sigma_x^4}{\sigma_x^2} \\
 & \vdots & \mu_x + \beta_1 \left( \frac{\sigma_x^2}{\sigma_x^2} + \frac{\sigma^4}{\sigma^2_x} \right) \\
 & \cdots & \left[ \mu_x + \beta_1 \left( \frac{\sigma_x^2}{\sigma_x^2} + \frac{\sigma^4}{\sigma^2_x} \right) \right]^2 \\
\end{array} \right).$$

(28)

$4.7$ Factorization

As in Subsections 3.5 and 1.6 we use

$$g = \mu_x + \beta_1 \sigma_x^2, \quad R = \left( \begin{array}{cc} 1 & g \\
0 & 1 \end{array} \right).$$

From (25) we have

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

with

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

and from (28)

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

with

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

$$+ \exp \left( 2\beta_0 + 2\beta_1 \mu_x + 2\beta_1^2 \sigma_x^2 \right) \left( \begin{array}{cccc} 1 & \frac{\beta_1 \sigma_x^4}{\sigma_x^2} & \cdots & \left( \frac{\beta_1 \sigma_x^4}{\sigma_x^2} \right)^2 \\
& \vdots & \frac{\beta_1 \sigma_x^4}{\sigma_x^2} \\
& \cdots & \frac{\beta_1 \sigma_x^4}{\sigma_x^2} + \frac{\sigma^4}{\sigma^2_x} \end{array} \right).$$
4.8 Final expression for $\Sigma_{al}$

We have $R\Sigma_{al} R^\top = A_2^{-1} B_2 A_2^{-1}$.

Now,

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

and hence

$$R\Sigma_{al} R^\top = \exp \left( - \left( \beta_0 + \beta_1 \mu_x + \frac{1}{2} \beta_1^2 \sigma_x^2 \right) \right) \left( \begin{array}{cc} 1 & 0 \\ 0 & \frac{\beta_1}{\sigma_x^2} \end{array} \right) +$$

$$+ e^{\beta_1^2 \sigma_x^2} \left( 1 - e^{-\beta_1^2 \sigma_x^2} \right) \left( \begin{array}{cc} 1 & \beta_1 \frac{1}{\sigma_x^2} + \frac{\sigma_1^2}{\sigma_x^2} \\ \beta_1 & \beta_1^2 + \frac{\sigma_1^2}{\sigma_x^2} \end{array} \right).$$

(29)

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

From (14) and (29) we get

$$R(\Sigma_{cor} - \Sigma_{al}) R^\top = e^{\beta_1^2 \sigma_x^2} \left( e^{\beta_1^2 \sigma_x^2} - 1 \right) \left( \begin{array}{cc} 1 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 \\ \frac{\sigma_1^2}{\sigma_x^2} \beta_1 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 + \frac{\sigma_1^2}{\sigma_x^2} \end{array} \right) +$$

$$e^{\beta_1^2 \sigma_x^2} \left( 1 - e^{-\beta_1^2 \sigma_x^2} \right) \left( \begin{array}{cc} 1 & \beta_1 \frac{1}{\sigma_x^2} + \frac{\sigma_1^2}{\sigma_x^2} \\ \beta_1 & \beta_1^2 + \frac{\sigma_1^2}{\sigma_x^2} \end{array} \right) + e^{\beta_1^2 \sigma_x^2} \left( \begin{array}{cc} 0 & 0 \\ 0 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 \end{array} \right).$$

We factorize once more. Let

$$D = \left( \begin{array}{cc} 1 & \beta_1 \\ 0 & 1 \end{array} \right).$$

Then

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

where

$$F = \left( e^{\beta_1^2 \sigma_x^2} - 1 \right) \left( \begin{array}{cc} 1 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 \\ \frac{\sigma_1^2}{\sigma_x^2} \beta_1 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 + \frac{\sigma_1^2}{\sigma_x^2} \end{array} \right) +$$

$$\left( 1 - e^{-\beta_1^2 \sigma_x^2} \right) \left( \begin{array}{cc} 1 & 0 \\ 0 & \frac{\sigma_1^2}{\sigma_x^2} \beta_1 \end{array} \right) + \left( \begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \right).$$
which can also be written as

\[
F = \left( \frac{e^{2\beta u \sigma_u^2} - 2 + e^{-\frac{\beta^2 u^2 \sigma_u^2}{\sigma_w^2}}}{\sigma_u^2 \beta_1 \left( e^{\beta u \sigma_u^2} - 1 \right)} \right) + \left( 0 \left( e^{2\beta u \sigma_u^2} - 2 + e^{-\frac{\beta^2 u^2 \sigma_u^2}{\sigma_w^2}} \right) \frac{\sigma_u^2}{\sigma_w^2} \right) = F_1 + F_2. \tag{30}
\]

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

\[
e^{2\beta u \sigma_u^2} - 2 + e^{-\frac{\beta^2 u^2 \sigma_u^2}{\sigma_w^2}} > \beta_1 \sigma_u^2 - \beta_1^2 \sigma_u^2 \frac{\sigma_u^2}{\sigma_w^2} = \beta_1^2 \frac{\sigma_u^2}{\sigma_w^2} > 0,
\]

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

\[
det F_1 = \frac{\sigma_u^4}{\sigma_w^2 \beta_1} \left[ e^{2\beta u \sigma_u^2} - 2 e^{\beta u \sigma_u^2} + e^{-\frac{\beta^2 u^2 \sigma_u^2}{\sigma_w^2}} - \left( e^{\beta u \sigma_u^2} - 1 \right)^2 \right]
\]

\[
= \frac{\sigma_u^4}{\sigma_w^2 \beta_1} \left( e^{\beta u \sigma_u^2} - 1 \right) > 0.
\]

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 \( \mu_x \) and \( \sigma_x^2 \) 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 \( \Sigma_{\text{cor}} \) and \( \Sigma_{\text{al}} \)

We show that for \( \beta_1 \neq 0 \)

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

where \( S_4 \) is positive definite.

17
We analyze the matrix $F$ of (30) when $\sigma_u^2 \to 0$. First,

\[
\begin{align*}
\exp(\beta_1 \sigma_u^2) - 2 + \exp(-\beta_1 \sigma_u^2) = & \frac{\beta_1^2 \sigma_u^2}{2} + \frac{1}{2} \beta_1^4 \sigma_u^4 + \frac{1}{2} \beta_1^2 \sigma_u^2 + \frac{1}{2} \beta_1^4 \frac{\sigma_u^2}{\sigma_w^2} + O(\sigma_u^6) = \\
= & \beta_1^2 \frac{\sigma_u^2}{\sigma_w^2} + \beta_1^4 \sigma_u^4 + O(\sigma_u^6) = \beta_1^2 \frac{\sigma_u^2}{\sigma_w^2} + \beta_1^4 \sigma_u^4 + O(\sigma_u^6).
\end{align*}
\]

Similarly,

\[
\exp(\beta_1 \sigma_u^2) - 1 = \beta_1^2 \sigma_u^2 + O(\sigma_u^4), \quad \frac{\sigma_u^2}{\sigma_w^2} = \frac{1}{\sigma_u^2} + O(\sigma_u^4).
\]

Therefore

\[
F = \sigma_u^4 \left( \begin{array}{cc} \frac{\beta_1^2}{\sigma_w^2} + \frac{\beta_1^4}{\sigma_w^2} & \frac{\beta_1^3}{\sigma_w^2} \\ \frac{\beta_1^3}{\sigma_w^2} & \frac{\beta_1^4}{\sigma_w^2} + \frac{\beta_1^2}{\sigma_w^2} \end{array} \right) + O(\sigma_u^6).
\]

The second matrix in brackets is nonnegative semidefinite, the first matrix has positive diagonal elements and its determinant $\frac{\beta_1^2}{\sigma_w^2}$ is positive. Hence

\[
F = \sigma_u^4 F_1 + O(\sigma_u^6),
\]

where $F_1$ is a positive definite matrix. Let $S_4 = \exp(\beta \sigma_u^2) \cdot R^{-1} D^T S D R^{-T}$, then $S_4$ is a positive definite and (31) holds true.

We proved the following theorem.

**Theorem 2.** Let $\mu_x, \sigma_x^2$ be known, and $\beta_1 \neq 0$. Then, as $\sigma_u^2 \to 0$,

\[
\Sigma_{\text{or}} - \Sigma_{\text{al}} = \sigma_u^4 \cdot S_4 + O(\sigma_u^6)
\]

with a positive definite matrix $S_4$.

7 Quasi-likelihood estimator

7.1 The estimator

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

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - m(W_i, b)}{v(W_i, b)} \frac{\partial m(W_i, b)}{\partial b} = 0.
\]
Here
\[ v(W, \beta) = \text{var}(Y|W). \]

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),
\[ \Sigma_{\text{QL}} = A_{\text{QL}}^{-1} B_{\text{QL}} A_{\text{QL}}^{-T}. \]

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):
\[ \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - m(W_i, b)}{m(W_i, b)} \frac{\partial m(W_i, b)}{\partial b} = 0. \quad (33) \]

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 \( \sigma^2 = 0 \), i.e., the measurement error vanishes, the equations (32) and (33) coincide, because \( Y|X \) has a Poisson distribution, and
\[ E(Y|X) = \text{var}(Y|X) = \lambda(X, \beta). \]

7.2 Asymptotic optimality of quasi-likelihood estimator

We use the theorem 2.1 from [2] to prove that \( \Sigma_{\text{QL}} < \Sigma_{\text{ML}} \).

Consider the family of unbiased estimation functions, which linearly depend on \( Y \):
\[ \mathcal{H} = \{ (Y - m(W, b)) \chi(W, b) \} , \]

where \( \chi \) 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 \( \chi(W, b) = \frac{1}{m(W, b)} \frac{\partial m(W, b)}{\partial b} \) and \( \chi(W, b) = \frac{1}{m(W, b)} \frac{\partial m(W, b)}{\partial b} \), respectively). For all \( G \in \mathcal{H} \)
\[ \frac{\partial G(b)}{\partial b} = -\chi(W, b) \frac{\partial m(W, b)}{\partial b} + (Y - m(W, b)) \frac{\partial \lambda(W, b)}{\partial b}. \]
As \( m(W, \beta) = \mathbb{E}[Y|W] \) and \( \chi(W, \beta) \frac{\partial m(W, \beta)}{\partial \beta} \) depend only on \( W \), for \( b = \beta \),
the expectation of the second item is equal to 0. Then
\[
\mathbb{E} \frac{\partial G}{\partial \beta^T} = - \mathbb{E} \left( \chi(W, \beta) \frac{\partial m(W, \beta)}{\partial \beta^T} \right).
\]

Let us compute \( \mathbb{E} G \mathbf{S}^\top_{\text{QL}} \).

\[
G(\beta) \mathbf{S}^\top_{\text{QL}}(\beta) = \chi(W, \beta) \frac{(Y - m(W, \beta))^2}{\epsilon(W, \beta)} \chi(W, \beta) \frac{\partial m(W, \beta)}{\partial \beta^T}.
\]

As \( \mathbb{E} [(Y - m(W, \beta))^2 | W] = \epsilon(W, \beta) \),
\[
\mathbb{E} [G(\beta) \mathbf{S}^\top_{\text{QL}}(\beta) | W] = \chi(W, \beta) \frac{\partial m(W, \beta)}{\partial \beta^T},
\]
\[
\mathbb{E} (G(\beta) \mathbf{S}^\top_{\text{QL}}(\beta)) = \mathbb{E} \left( \chi(W, \beta) \frac{\partial m(W, \beta)}{\partial \beta^T} \right).
\]

So
\[
\left( \mathbb{E} \frac{\partial G}{\partial \beta^T} \right)^{-1} \mathbb{E} G(\beta) \mathbf{S}^\top_{\text{QL}}(\beta) = - \mathbf{I}
\]
is a constant matrix.

By theorem 2.1 [2] the QL-estimation function is \( \mathcal{O}_p \)-optimal, see the
definition of this optimality in [2]. Because \( \Sigma_G = \left( \mathbb{E} G^{(s)} \mathbf{G}^{(s)}^\top \right)^{-1} \), where
\[
G^{(s)}(\beta) = \mathbb{E} \frac{\partial G(\beta)^T}{\partial \beta} \left( \mathbb{E} G(\beta) \mathbf{G}^{(s)} \right)^{-1} G(\beta)
\]
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 \( S_2 \), \( S_1 \leq S_2 \) iff \( S_1 - S_2 \) is positive semidefinite).

### 7.3 Comparison of asymptotic covariance matrices of alternative and QL estimators

Now we show, that \( \Sigma_{\text{al}} - \Sigma_{\text{QL}} \) is strictly positive definite. Suppose that
\( \Sigma_{\text{al}} - \Sigma_{\text{QL}} \) is singular. Then so is \( \Sigma_{\text{al}}^{-1} - \Sigma_{\text{QL}}^{-1} = \text{cov} S^{(s)}_{\text{QL}} - \text{cov} S^{(s)}_{\text{al}} \). But
the first formula of the proof implies
\[
\text{cov} (S^{(s)}_{\text{QL}} - S^{(s)}_{\text{al}}) = \text{cov} S^{(s)}_{\text{QL}} - \text{cov} S^{(s)}_{\text{al}}.
\]
(34)
Indeed, in Subsection 7.2 we obtained that 
\[ \left( E_{\beta} \frac{\partial S_{\alpha}(\beta)}{\partial \beta} \right)^{-1} E S_{\alpha}(\beta) S_{Q\ell}(\beta) = \left( E_{\beta} \frac{\partial S_{\alpha}(\beta)}{\partial \beta} \right)^{-1} E S_{Q\ell}(\beta) S_{\alpha}(\beta). \]
Hence, with \( E S_{\alpha}(\beta) = E S_{Q\ell}(\beta) = 0 \), and using the definition of standardized estimator, we obtain 
\[ E S_{\alpha} S_{Q\ell}^{T} = A_{\alpha}, \quad E S_{\alpha}(\beta) S_{Q\ell}(\beta) = A_{\alpha} S_{\alpha}^{T} A_{\alpha}^{-1} = \text{cov} S_{\alpha}^{(s)} \]. Hence (34) holds.

Then the vector \( S_{Q\ell}^{(s)} - S_{\alpha}^{(s)} \) with probability 1 lies in own non-stochastic subspace, and there exists a linear functional \( f \neq 0 \) such that \( f(S_{Q\ell}^{(s)} - S_{\alpha}^{(s)}) = 0 \) a.s. Denote \( f_{1} = f \cdot \left( E_{\beta} \frac{\partial S_{\alpha}(\beta)}{\partial \beta} \right) (\text{cov} S_{Q\ell})^{-1}. \)
\[ f_{2} = f \cdot \left( E_{\beta} \frac{\partial S_{\alpha}(\beta)}{\partial \beta} \right) (\text{cov} S_{\alpha})^{-1}. \]
Then \( f_{1}(S_{Q\ell}) - f_{2}(S_{\alpha}) = 0 \) a.s., and a.s.
\[ \frac{1}{\nu(W, \beta)} (Y - m(W, \beta)) f_{1} \left( \frac{\partial m}{\partial \beta} \right) = \frac{1}{\nu(W, \beta)} (Y - m(W, \beta)) f_{2} \left( \frac{\partial m}{\partial \beta} \right). \]

Now choose coefficients \( a_{k}, \ c_{k} \) such that \( f_{k} \left( \mu_{X|W} + \beta_{1} \sigma_{X|W}^{2} \right) = a_{k} + c_{k} W, \ k = 1, 2. \) As
\[ \frac{\partial m(W, \beta)}{\partial \beta} = m(W, \beta) \left( \mu_{X|W} + \frac{1}{\beta_{1}} \sigma_{X|W}^{2} \right), \]
\[ m(W, \beta) \frac{Y - m(W, \beta)}{\nu(W, \beta)} (a_{1} + c_{1} W) = (Y - m(W, \beta)) (a_{2} + c_{2} W) \text{ a.s.} \]
So almost surely
\[ a_{1} + c_{1} W = \left( 1 + m(W, \beta) \left( \text{exp}(\beta_{1} \sigma_{X|W}^{2}) - 1 \right) \right) (a_{2} + c_{2} W). \]
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 \( f_{2} = 0, \ f_{1} = 0 \). So we get contradiction.
We have proved that \( \Sigma_{\alpha} = \Sigma_{Q\ell} \) is a positive definite nonsingular matrix.
8 Conclusion

In the case of known $\mu_x$ and $\sigma^2_x$, we compared $\Sigma_{cor}$ and $\Sigma_{al}$. This result holds true for arbitrary $\sigma^2_x$.

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

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

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

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

The next question remains open. To compare $\Sigma_{QL}$, $\Sigma_{al}$, and $\Sigma_{cor}$, when $\mu_x$, $\sigma^2_x$ are unknown and we plug-in $\hat{\mu}_x = \hat{\mu}_w$ and $\hat{\sigma}^2_x = \hat{\sigma}^2_w - \hat{\sigma}^2_w$ instead of $\mu_x$ and $\sigma^2_x$, while we construct $\hat{\beta}_{QL}$ and $\hat{\beta}_{al}$ (here $\hat{\mu}_w$ and $\hat{\sigma}^2_w$ are sample mean and sample variance of $W_1, W_2, \ldots, W_n$). In this situation an additional term in $\Sigma_{QL}$ and $\Sigma_{al}$ appears, see [3]. But $\Sigma_{cor}$ remains unchanged, because $\beta_{cor}$ 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


