Journal…**Journal of Statistics** Article ID…**GSTA419084**

TO: CORRESPONDING AUTHOR

AUTHOR QUERIES - TO BE ANSWERED BY THE AUTHOR

Dear Author
Please address all the numbered queries on this page which are clearly identified on the proof for your convenience.

Thank you for your cooperation

Q1	Please provide department details for affiliations a and b.	
Q2	While converting name and date citations to numbered, we have assumed reference Carroll et al. (2006) to be Carroll et al. (1996) given in the list and numbered it as Ref. [2]. Please confirm if this correct.	
Q3	Please expand p.d. and ML.	

Production Editorial Department, Taylor & Francis
4 Park Square, Milton Park, Abingdon OX14 4RN

Telephone: +44 (0) 2070 176000
Facsimile: +44 (0) 2070 176336

Statistics, Vol. 00, No. 0, Month 2009, 1–16

Taylor & Francis
Taylor & Francis Group

Optimality of quasi-score in the multivariate mean–variance model with an application to the zero-inflated Poisson model with measurement errors

Alexander Kukush[a]*, Andrii Malenko[a], Hans Schneeweiss[b] and Shalabh[b,c]

[a]*National Taras Shevchenko University of Kyiv, Volodymyrska Street 64, 01033, Kyiv, Ukraine;* [b]*University of Munich, Akademiestrasse 1, 80799, Munchen, Germany;* [c] *Indian Institute of Technology Kanpur, 208016, Kanpur, India* Q1

(*Received 24 September 2008; final version received 15 May 2009*)

In a multivariate mean–variance model, the class of linear score (LS) estimators based on an unbiased linear estimating function is introduced. A special member of this class is the (extended) quasi-score (QS) estimator. It is 'extended' in the sense that it comprises the parameters describing the distribution of the regressor variables. It is shown that QS is (asymptotically) most efficient within the class of LS estimators. An application is the multivariate measurement error model, where the parameters describing the regressor distribution are nuisance parameters. A special case is the zero-inflated Poisson model with measurement errors, which can be treated within this framework.

Keywords: multivariate mean–variance model; measurement errors; zero-inflated Poisson model; quasi-score; corrected score

MSC 2000: 62J05; 62H12; 62F12

1. Introduction

We consider a q-dimensional random vector y which is related to a p-dimensional random vector x via a conditional vector-valued mean function $m(x,\theta) := \mathbf{E}\,(y|x)$. The mean function depends on an unknown d-dimensional parameter vector θ to be estimated with the help of an iid sample $(x_i, y_i),\ i = \overline{1,n}$. (All vectors are taken to be column vectors.) The mean function is supplemented by a matrix-valued conditional variance function $v(x,\theta) := \mathbf{V}(y|x)$ depending on the same parameter vector θ as the mean function. This parameter vector also determines the distribution of the regressor variable x, which is supposed to be given by a density function $\rho(x,\theta)$. Such a model may arise in the context of measurement error models.

We can estimate θ by constructing a quasi-score (QS) function. However, the usual QS function

$$\frac{\partial m^{\mathrm{T}}(x,\theta)}{\partial\theta}\, v(x,\theta)^{-1}\{y - m(x,\theta)\},$$

*Corresponding author. Email: alexander_kukush@univ.kiev.ua

ISSN 0233-1888 print/ISSN 1029-4910 online

DOI: 10.1080/02331880903189117
http://www.informaworld.com

Q2 [cf. e.g. 1–4], is not optimal and sometimes not even feasible. We extend the QS function by adding the term $\partial \log \rho(x,\theta)/\partial\theta$. This extended QS function is again called QS function of the model.

We show that the QS estimator of θ based on this QS function is optimal within the class of the so-called linear score (LS) estimators, which are based on linear-in-y unbiased estimating (or score) functions. Optimality is defined in terms of the asymptotic covariance matrices (ACMs) of the QS and LS estimators. We also derive a formula for the rank of the difference of the two ACMs.

This paper is a generalization of some of the results of Kukush *et al.* [5] to the case of a multivariate response variable y, whereas in the previous paper only the univariate case was considered. The proofs, however, carry over with only minor changes and will therefore be omitted.

An application of the multivariate model is the zero-inflated log-linear Poisson measurement error model, which is characterized by the property that the distribution of a count variable y is given by a Poisson law for $y > 0$, while the value $y = 0$ occurs with a separate probability unrelated to the Poisson distribution, [cf. 6–8]. Although this model is univariate, it can be studied under the guise of a two-dimensional multivariate model, where the indicator variable for the event $y = 0$ serves as the second variable.

In the following, we often suppress the arguments in the various functions. For example, we write m instead of $m(x,\theta)$. Derivatives with respect to θ (or other variables) are denoted by a subscript, e.g. $(\log\rho)_\theta := \partial \log \rho(x,\theta)/\partial\theta$, which is a vector of the same dimension as θ. For a vector, like m, the derivative m_θ is a matrix (i.e. $m_\theta := \partial m/\partial\theta^{\mathrm{T}}$), and for a matrix, it is a tensor. For example, if g is a $(d \times q)$ matrix with elements g_{ij}, $i = \overline{1,d}$, $j = \overline{1,q}$, then g_θ is a tensor with elements $g_{ik}^{j} := \partial g_{ij}/\partial\theta_k$, $k = \overline{1,d}$, such that $g_\theta y$ is a matrix with elements $(g_\theta y)_{ik} = \sum_{j=1}^{d} g_{ik}^{j} y_j$, so that $(gy)_\theta = g_\theta y$.

Section 2 introduces the LS and QS estimators in a general mean–variance model and states the main results on the optimality of QS. Section 3 applies the general theory to the zero-inflated Poison measurement error model. Section 4 has some simulation results, and Section 5 concludes.

2. The general mean–variance model

2.1. *LS and QS estimators of a mean–variance model*

Let x and y be random vectors distributed in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Conditional mean and conditional variance of y given x are supposed to be known except for an unknown parameter vector θ with dimension d:

$$m(x,\theta) = \mathbf{E}\,(y|x) \in \mathbb{R}^q, \quad v(x,\theta) = \mathbf{V}(y|x) \in \mathbb{R}^{q\times q}.$$

We assume that $v(x,\theta)$ is a positive-definite matrix for all x and θ. Let x have marginal density $\rho(x,\theta)$.

The class $\mathcal{L}$ of all unbiased linear-in-y scores consists of functions

$$S_{\mathrm{L}}(x,y;\theta) = g(x,\theta)y - h(x,\theta), \tag{1}$$

where g is a matrix of size $d \times q$ and h is a vector of dimension d. Unbiasedness means that, for all θ, $\mathbf{E}S_{\mathrm{L}}(x,y;\theta) = 0$. Note that the expectation of a random function of θ is always taken under the same value of θ as the θ in the argument of the function.

Suppose an iid sample (x_i, y_i), $i = \overline{1,n}$, is given. The LS estimator $\hat\theta_{\mathrm{L}}$ based on S_{L} is given by the solution to the equation

$$\sum_{i=1}^{n} S_{\mathrm{L}}(x_i, y_i; \hat\theta_{\mathrm{L}}) = 0.$$

Under regularity conditions, as detailed for a similar model in [9] (see also [10]), the solution $\hat{\theta}_{\mathrm{L}}$ is, with probability tending to 1, unique for sufficiently large n and $\hat{\theta}_{\mathrm{L}}$ is consistent and asymptotically normal with an ACM given by

$$\Sigma_{\mathrm{L}} = (\mathbf{E}\, S_{\mathrm{L}\theta})^{-1}\, \mathbf{E}\, (S_{\mathrm{L}} S_{\mathrm{L}}^{\mathrm{T}})(\mathbf{E}\, S_{\mathrm{L}\theta})^{-\mathrm{T}}.$$

The most important regularity condition is the condition that $\mathbf{E}\, S_{\mathrm{L}\theta}$ should be nonsingular. We call this the identifiability condition.

QS is a particular element of the class $\mathcal{L}$. It is given by the QS function

$$S_{\mathrm{Q}}(x, y; \theta) = m_\theta^{\mathrm{T}} v^{-1}(y - m) + (\log \rho)_\theta. \tag{2}$$

Under regularity conditions [9], $\hat{\theta}_{\mathrm{Q}}$ is consistent and asymptotically normal with the ACM $\Sigma_{\mathrm{Q}} = (\mathbf{E}\, S_{\mathrm{Q}} S_{\mathrm{Q}}^{\mathrm{T}})^{-1}$. The identifiability condition here boils down to the condition that $\mathbf{E}\, S_{\mathrm{Q}} S_{\mathrm{Q}}^{\mathrm{T}} = \mathbf{E}\, m_\theta^{\mathrm{T}} v^{-1} m_\theta + \mathbf{E}\, (\log \rho)_\theta (\log \rho)_\theta^{\mathrm{T}}$ should be positive-definite. This is equivalent to the condition that the system of $(q + 1)$-dimensional random vectors

$$\{m_{1\theta_k}, m_{2\theta_k}, \ldots, m_{q\theta_k}, (\log \rho)_{\theta_k},\ k = \overline{1, d}\} \quad \text{is linearly independent,} \tag{3}$$

where $m = (m_1, m_2, \ldots, m_q)^{\mathrm{T}}$.

2.2. *Optimality of QS*

The following identity is useful in proving the optimality of QS within the class $\mathcal{L}$:

$$\mathbf{E}\, S_{\mathrm{L}\theta} + \mathbf{E}\, S_{\mathrm{L}} S_{\mathrm{Q}}^{\mathrm{T}} = 0. \tag{4}$$

To prove (4), first note that

$$\begin{aligned} \mathbf{E}\, S_{\mathrm{L}} S_{\mathrm{Q}}^{\mathrm{T}} &= \mathbf{E}\, g(y - m)(y - m)^{\mathrm{T}} v^{-1} m_\theta + \mathbf{E}\, (gm - h)(\log \rho)_\theta^{\mathrm{T}} \\ &= \mathbf{E}\, g m_\theta + \mathbf{E}\, (gm - h)(\log \rho)_\theta^{\mathrm{T}}. \end{aligned} \tag{5}$$

In addition, by differentiating $\mathbf{E}\, S_{\mathrm{L}} = \mathbf{E}\, (gm - h)$, which is identically equal to zero, with respect to θ, we obtain the identity

$$\mathbf{E}\, (gm - h)_\theta + \mathbf{E}\, (gm - h)(\log \rho)_\theta^{\mathrm{T}} = 0. \tag{6}$$

Now,

$$\mathbf{E}\, S_{\mathrm{L}\theta} = \mathbf{E}(g_\theta m - h_\theta),$$

where g_θ is a tensor (see Section 1), and Equation (4) holds as a consequence of Equations (5) and (6).

In a similar way as in [5], we can prove the following theorem by applying Equation (4).

THEOREM 2.1 *In a mean–variance model,*

$$\Sigma_{\mathrm{L}} \geq \Sigma_{\mathrm{Q}},$$

in the sense of the Löwner order.

More details are provided by the following theorem.

THEOREM 2.2 *In a mean–variance model,*

$$\operatorname{rank}(\Sigma_L - \Sigma_Q) = \operatorname{rank}\left[\begin{pmatrix} (gv)_{i1} \\ \vdots \\ (gv)_{iq} \\ (gm-h)_i \end{pmatrix}, \begin{pmatrix} (m_1)_{\theta_i} \\ \vdots \\ (m_q)_{\theta_i} \\ (\log\rho)_{\theta_i} \end{pmatrix}, \ i=\overline{1,d}\right] - d. \tag{7}$$

Proof Kukush *et al.* [5, proof of Theorem 4.2], have shown that

$$\operatorname{rank}(\Sigma_L - \Sigma_Q) = \operatorname{rank}\{(S_L)_i,\ (S_Q)_i,\ i=\overline{1,d}\} - d. \tag{8}$$

The rank on the r.h.s. of Equation (8) can be expressed in terms of the constituents of S_L and S_Q. For this purpose, we evaluate the defect of the system of random variables $\{(S_L)_i,\ (S_Q)_i,\ i=\overline{1,d}\}$, which is the maximum number of linearly independent constant vectors $(c_1^T, c_2^T)^T$ which satisfy the equation

$$c_1^T S_L + c_2^T S_Q = 0, \quad \text{a.s},$$

or, equivalently,

$$c_1^T(gy-h) + c_2^T[m_\theta^T v^{-1}(y-m) + (\log\rho)_\theta] = 0, \quad \text{a.s.}$$

By similar arguments as in [5], using the condition that v is positive-definite for all x, this equation can be rewritten as a system of two equations, one concerning the terms pertaining to y, the other one concerning the remaining terms:

$$c_1^T gv + c_2^T m_\theta^T = 0,$$
$$c_1^T(gm-h) + c_2^T(\log\rho)_\theta = 0,$$

a.s. Thus,

$$\operatorname{def}\{(S_L)_i, (S_Q)_i, i=\overline{1,d}\} = \operatorname{def}\left[\begin{pmatrix} (gv)_{i1} \\ \vdots \\ (gv)_{iq} \\ (gm-h)_i \end{pmatrix}, \begin{pmatrix} (m_1)_{\theta_i} \\ \vdots \\ (m_q)_{\theta_i} \\ (\log\rho)_{\theta_i} \end{pmatrix}, \ i=\overline{1,d}\right]. \tag{9}$$

As both systems in Equation (9) have the same number, $2d$, of random elements (on the left-hand side random variables, on the right-hand side random vectors), the equality of defects implies the equality of ranks. The statement of the theorem now follows from Equations (8) and (9). ■

2.3. *Marginal QS*

Starting from a multivariate mean–variance model, we can always consider a subvector of y and set up the corresponding marginal mean–variance model for this subvector. In particular, the subvector may consist of a single component of y. We can construct a marginal QS function with this marginal model. As long as the identifiability condition (3) for this marginal QS function is satisfied, we can use it to estimate θ.

We study the relation between the full and the marginal QS estimator of θ. For simplicity, let $q=2$. We consider the marginal QS estimator which uses only y_1 and is based on the marginal

QS function

$$S_{Q^*} = \frac{m_{1\theta}(y_1 - m_1)}{v_{11}} + (\log \rho)_\theta. \tag{10}$$

This estimator is most efficient in the class of estimators based on a linear-in-y_1 estimating function. Above we considered estimators linear in $(y_1, y_2)^T$, and it is obvious from Theorem 2.1 that

$$\Sigma_{Q^*} \geq \Sigma_Q.$$

We can compute the rank of $\Sigma_{Q^*} - \Sigma_Q$. Consider the functions

$$g^* = \left(\frac{m_{1\theta}}{v_{11}}, 0\right),$$

$$h^* = \frac{m_{1\theta}\, m_1}{v_{11}} - (\log \rho)_\theta.$$

Then $S_{Q^*} = g^* y - h^*$. Furthermore,

$$g^* v = \left(m_{1\theta},\ \frac{v_{12}}{v_{11}} m_{1\theta}\right),$$

$$g^* m - h^* = (\log \rho)_\theta.$$

By Equation (7), rank $(\Sigma_{Q^*} - \Sigma_Q) + d =$

$$\text{rank}\begin{bmatrix} m_{1\theta}^T & m_{1\theta}^T \\ m_{1\theta}^T v_{12} v_{11}^{-1} & m_{2\theta}^T \\ (\log \rho)_\theta^T & (\log \rho)_\theta^T \end{bmatrix} = \text{rank}\begin{bmatrix} 0 & m_{1\theta}^T \\ m_{1\theta}^T v_{12} v_{11}^{-1} - m_{2\theta}^T & m_{2\theta}^T \\ 0 & (\log \rho)_\theta^T \end{bmatrix}.$$

As by assumption $v(x, \theta)$ is positive-definite for all x; therefore, v_{11} is positive for all x and hence

$$\text{rank}\,(\Sigma_{Q^*} - \Sigma_Q) = \text{rank}\begin{bmatrix} 0 & m_{1\theta}^T \\ m_{1\theta}^T v_{12} - m_{2\theta}^T v_{11} & m_{2\theta}^T v_{11} \\ 0 & (\log \rho)_\theta^T \end{bmatrix} - d, \tag{11}$$

where rank[$\cdot$] is the column rank of the system of random variables [$\cdot$].

3. Zero-inflated Poisson model

3.1. *The model and its QS estimator*

Consider a scalar response variable y_1 and a scalar regressor variable ξ such that the conditional distribution of y_1 given ξ is a mixture of a Poisson distribution $\text{Po}(\eta)$ with parameter η and a one-point distribution δ_0 at point zero with mixing parameter $\alpha \in (0, 1)$:

$$y_1|\xi \sim \alpha\delta_0 + (1 - \alpha)\text{Po}(\eta).$$

Let $\eta = \exp(\beta_0 + \beta_1 \xi)$. In addition to y_1, we introduce the indicator variable

$$y_2 = I(y_1 = 0),$$

so that $y = (y_1, y_2)^{\mathrm{T}}$ is a bivariate response variable. This zero-inflated log-linear Poisson model is a special case of our general model with $p = 1$ and $q = 2$. The distribution of $y|\xi$ is given by

$$p(y|\xi) = (1-\alpha)\,\frac{\mathrm{e}^{-\eta}\eta^{y_1}}{y_1!}\,\left(1 + \frac{\alpha}{1-\alpha}\,\mathrm{e}^{\eta}\right)^{y_2}.$$

The variable ξ is not directly observable. Instead, we observe

$$x = \xi + \delta$$

with a measurement error $\delta \sim N(0, \sigma_\delta^2),\ \sigma_\delta^2 > 0$, which is independent of ξ and y. The error variance σ_δ^2 is assumed to be known. In addition, we assume $\xi \sim N(\mu, \sigma_\xi^2),\ \sigma_\xi^2 > 0$, so that

$$\log \rho(x, \theta) = -\frac{(x-\mu)^2}{2\sigma^2} - \log\sigma + \text{const}, \quad \sigma^2 = \sigma_\xi^2 + \sigma_\delta^2.$$

The unknown parameter vector of this model is $\theta = (\alpha, \beta_0, \beta_1, \mu, \sigma)^{\mathrm{T}}$, and thus $d = 5$. To derive the mean–variance model for $y|x$, we need to compute $\mu_1(x) := \mathbf{E}\,(\xi|x)$ and $\tau^2 := \mathbf{V}(\xi|x)$. We have

$$\mu_1(x) = Kx + (1-K)\mu \ \text{ with } \ K = K(\sigma) = 1 - \sigma_\delta^2\sigma^{-2} \quad \text{and} \quad \tau^2 = \sigma_\delta^2 K,$$

where K is the reliability ratio [cf. 5, Section 6.2]. The mean function $m(x, \theta) = (m_1(x, \theta), m_2(x, \theta))^{\mathrm{T}}$ is given by

$$\begin{aligned} m_1(x, \theta) &= (1-\alpha)\exp\{\beta_0 + \mu_1(x)\beta_1 + \beta_1^2\tau^2/2\}, \\ m_2(x, \theta) &= \alpha + (1-\alpha)\,\mathbf{E}\,(f|x), \end{aligned} \tag{12}$$

where $f = f(t) = \exp\{-e^t\}$ with $t = \beta_0 + \beta_1\mu_1(x) + \beta_1\tau\gamma$ and $\gamma \sim N(0, 1)$, independent of x. The matrix $v = v(x, \theta)$ is expressed in terms of m_1 and m_2 as follows:

$$v = \begin{pmatrix} m_1(1-m_1) + \dfrac{1}{1-\alpha}\,\mathrm{e}^{\beta_1^2\tau^2}m_1^2 & -m_1m_2 \\ -m_1m_2 & m_2(1-m_2) \end{pmatrix}. \tag{13}$$

With these mean and variance functions, we can set up the QS estimator as in Section 2.1.

Q3 It can be proved that $v(x, \theta)$ is p.d. for all x and θ, a.s. Indeed, we have

$$v(x, \theta) = \mathbf{E}\,(\mathbf{V}(y|\xi)|x) + \mathbf{V}(\mathbf{E}\,(y|\xi)|x) \geq \mathbf{E}\,(\mathbf{V}(y|\xi)|x),$$

and it is enough to show that $\mathbf{V}(y|\xi)$ is p.d. for all ξ and θ a.s. Let z be an indicator variable independent of ξ, with $P(z = 0) = \alpha$ and $P(z = 1) = 1 - \alpha$, such that $y_1|(\xi,\ z = 1) \sim \mathrm{Po}(\eta)$ and $y_1|(\xi,\ z = 0) \sim \delta_0$. Then

$$\mathbf{V}(y|\xi) \geq \mathbf{E}\,\mathbf{V}(y|\xi, z) = (1-\alpha)\mathbf{V}(y|\xi, z = 1) = (1-\alpha)\begin{pmatrix} \eta & -\eta e^{-\eta} \\ -\eta e^{-\eta} & e^{-\eta}(1-e^{-\eta}) \end{pmatrix},$$

which is positive-definite.

We can prove that the QS estimator of μ is just the empirical mean. Indeed, consider m_θ:

$$\begin{aligned}
m_{1\theta} &= \left(-\frac{1}{1-\alpha}, 1, \mu_1(x)+\tau^2\beta_1, (1-K)\beta_1, \beta_1(x-\mu)\frac{\partial K}{\partial\sigma}+\frac{1}{2}\beta_1^2\frac{\partial\tau^2}{\partial\sigma}\right)^{\mathrm{T}} m_1,\\
m_{2\alpha} &= 1-\mathbf{E}\,(f|x),\\
m_{2\beta_0} &= (1-\alpha)\,\mathbf{E}\,(f'|x),\\
m_{2\beta_1} &= (1-\alpha)(\mu_1(x)\,\mathbf{E}\,(f'|x)+\beta_1\tau^2\,\mathbf{E}\,(f''|x)),\\
m_{2\mu} &= (1-\alpha)(1-K)\beta_1\,\mathbf{E}\,(f'|x),\\
m_{2\sigma} &= (1-\alpha)\beta_1\left((x-\mu)\frac{\partial K}{\partial\sigma}\,\mathbf{E}\,(f'|x)+\frac{1}{2}\beta_1\frac{\partial\tau^2}{\partial\sigma}\,\mathbf{E}\,(f''|x)\right).
\end{aligned} \tag{14}$$

Here, we used the identity

$$\mathbf{E}\,[\gamma f'(a(x)+c\gamma)|x] = c\,\mathbf{E}\,[f''(a(x)+c\gamma)|x]$$

with $c=\beta_1\tau$. We see that

$$(m_{1\mu}, m_{2\mu}) = (1-K)\beta_1(m_{1\beta_0}, m_{2\beta_0}). \tag{15}$$

This implies that from the second and fourth equations for $\hat{\theta}_{\mathrm{Q}}$, i.e. from

$$\sum_{i=1}^{n}\{m_{1\beta_0}(x_i,\theta), m_{2\beta_0}(x_i,\theta)\}v^{-1}\{y_i-m(x_i,\theta)\}=0,$$

$$\sum_{i=1}^{n}\{m_{1\mu}(x_i,\theta), m_{2\mu}(x_i,\theta)\}v^{-1}\{y_i-m(x_i,\theta)\}+\sum_{i=1}^{n}\{\log\rho(x_i,\theta)\}_\mu=0,$$

we obtain $\hat{\mu}_Q=\overline{x}$.

On the other hand, the QS estimator of σ^2 is *not* the empirical variance of $x_i,\ i=\overline{1,n}$. We will give an indirect proof of this fact in the next section.

Note It is interesting to note that a marginal QS method, which uses only the conditional mean and variance of y_1, does not work. Indeed, such a method would be based on the QS function

$$S_{\mathrm{Q}^*} = m_{1\theta}v_{11}^{-1}(y_1-m_1)+(\log\rho)_\theta,$$

alone (Equation (10)). But since the first two components of $m_{1\theta}$ are linearly dependent (Equation (14)) the estimating equations based on S_{Q^*} are not sufficient to produce a unique solution $\hat{\theta}_{\mathrm{Q}^*}$. Looked at it from another angle, it is seen that the identifiability condition (3) is violated.

3.2. *Modified corrected score*

In this section, we construct a score function to estimate θ, which does not use any information about the distribution of x.

Q3 Consider the ML score for (α, β) in the error-free model:

$$S_{\mathrm{ML}} = \frac{y_1}{\eta}\begin{pmatrix} 0 \\ \eta_\beta \end{pmatrix} + \frac{y_2}{\alpha + (1-\alpha)\mathrm{e}^{-\eta}}\begin{pmatrix} \frac{1}{(1-\alpha)} \\ \alpha\eta_\beta \end{pmatrix} - \begin{pmatrix} \frac{1}{(1-\alpha)} \\ \eta_\beta \end{pmatrix}, \quad \eta_\beta = \eta\begin{pmatrix} 1 \\ \xi \end{pmatrix}. \tag{16}$$

It is not possible to construct the so-called corrected score (CS) function $S_{\mathrm{C}}^{(\alpha,\beta)}$ as the solution to the deconvolution problem

$$\mathbf{E}\,(S_{\mathrm{C}}^{(\alpha,\beta)}(x, y; \theta)|\xi, y) = S_{\mathrm{ML}}(\xi, y; \theta),$$

[cf. 11], because there are complex zeros in the common denominator of S_{ML}, [cf. 12]. Therefore, we modify S_{ML} by multiplying the first component of S_{ML} by $(1-\alpha)(\alpha + (1-\alpha)\mathrm{e}^{-\eta})$ and the other two components by $\alpha + (1-\alpha)\mathrm{e}^{-\eta}$. It should be noted that this modified S_{ML} is no more optimal in the context of the error-free model. Nevertheless, we use it to construct a modified score $S_{\mathrm{C}}^{(\alpha,\beta)}$, which is the solution to the modified deconvolution problem

$$\mathbf{E}\,(S_{\mathrm{C}}^{(\alpha,\beta)}(x, y; \theta)|\xi, y) = (\alpha + (1-\alpha)\mathrm{e}^{-\eta})\mathrm{diag}\{1-\alpha, 1, 1\}S_{\mathrm{ML}}(\xi, y; \theta).$$

The parameters μ and σ^2 are estimated as empirical mean and variance, respectively.

The modified CS (MCS) S_{C} is a linear unbiased score function, $S_{\mathrm{C}} = gy - h$, where

$$\mathbf{E}\,(g|\xi) = \begin{pmatrix} 0 & 1 \\ \alpha + (1-\alpha)\mathrm{e}^{-\eta} & \alpha\eta \\ (\alpha + (1-\alpha)\mathrm{e}^{-\eta})\xi & \alpha\eta\xi \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{E}\,(h|\xi) = \begin{pmatrix} \alpha + (1-\alpha)\mathrm{e}^{-\eta} \\ (\alpha + (1-\alpha)\mathrm{e}^{-\eta})\eta \\ (\alpha + (1-\alpha)\mathrm{e}^{-\eta})\eta\xi \\ \xi - \mu \\ (\xi - \mu)^2 - \sigma_\xi^2 \end{pmatrix}. \tag{17}$$

The last two components of h are

$$h_4 = x - \mu,$$
$$h_5 = (x - \mu)^2 - \sigma^2.$$

The other components of h and the elements of g are given below (see Section 4).

The estimator based on this score function is the MCS estimator.

The following theorem states the efficiency of the QS estimator vis-à-vis the MCS estimator as measured by the difference of the ACMs.

THEOREM 3.1 *Under* $\alpha \in (0, 1)$, $\beta_1 \neq 0$,

$$\Sigma_{\mathrm{Q}}^{(\alpha,\beta,\sigma)} < \Sigma_{\mathrm{C}}^{(\alpha,\beta,\sigma)}.$$

These matrices are the ACMs of the QS and MCS estimators of $(\alpha, \beta_0, \beta_1, \sigma)^{\mathrm{T}}$, *respectively. Under* $\beta_1 = 0$ *and* $\alpha \in (0, 1)$, *we have rank* $(\Sigma_{\mathrm{C}}^{(\alpha,\beta,\sigma)} - \Sigma_{\mathrm{Q}}^{(\alpha,\beta,\sigma)}) = 1$.

It follows that, under $\beta_1 \neq 0$, we have $\Sigma_{\mathrm{Q}}^{(\sigma)} < \Sigma_{\mathrm{C}}^{(\sigma)}$, therefore, $\hat{\sigma}_{\mathrm{Q}}^2$ is not the empirical variance.

If α is known, then, under $\beta_1 \neq 0$, the QS estimator of $(\beta_0, \beta_1, \sigma)$ is strictly more efficient than the MCS estimator.

3.3. *Proof of Theorem 3.1*

To prove the statement of the theorem, we compute the rank of the system in Equation (7). First note that $\hat{\mu}_C = \hat{\mu}_Q = \bar{x}$. Therefore, from the inequality $\Sigma_C \geq \Sigma_Q$, we have

$$\Sigma_C - \Sigma_Q = \begin{pmatrix} \Sigma_C^{(\alpha,\beta,\sigma)} - \Sigma_Q^{(\alpha,\beta,\sigma)} & 0 \\ 0 & 0 \end{pmatrix}.$$

The right-hand side of Equation (7) can be written as the rank of the following system of three-dimensional random vectors minus 5:

$$\begin{bmatrix} (gv)_{11} & (gv)_{21} & (gv)_{31} & m_{1\alpha} & m_{1\beta_0} & m_{1\beta_1} & m_{1\sigma} & 0 & 0 \\ (gv)_{12} & (gv)_{22} & (gv)_{32} & m_{2\alpha} & m_{2\beta_0} & m_{2\beta_1} & m_{2\sigma} & 0 & 0 \\ (gm-h)_1 & (gm-h)_2 & (gm-h)_3 & 0 & 0 & 0 & 0 & x-\mu & (x-\mu)^2-\sigma^2 \end{bmatrix}, \tag{18}$$

where the column $(m_{1\mu}, m_{2\mu}, 0)^T$ was dropped because of Equation (15).

We divide the proof into two parts. First we show that

$$\{(gm-h)_1, (gm-h)_2, (gm-h)_3, x-\mu, (x-\mu)^2-\sigma^2\} \tag{19}$$

are linearly independent functions of x. Then we show that the functions of x

$$\{m_{2\alpha}, m_{2\beta_0}, m_{2\beta_1}, m_{2\sigma}\} \tag{20}$$

are linearly independent.

With these two sets of linearly independent functions, we immediately obtain that the column rank of the system in Equation (18) is 9; therefore, the rank of $\Sigma_C^{(\alpha,\beta,\sigma)} - \Sigma_Q^{(\alpha,\beta,\sigma)}$ is 4, and this matrix is positive-definite.

3.3.1. *Part 1 of the proof*

We want to show that the functions (19) are linearly independent under $\beta_1 \neq 0$. We consider only the case $\beta_1 > 0$. The case $\beta_1 < 0$ can be treated similarly. We divide the proof into three steps:

(1) We prove that $(gm-h)_i \to 0$ as $x \to -\infty$, $i = 1, 2, 3$, while x and x^2 converge to infinity. This means that we can exclude $(x-\mu)$ and $(x-\mu)^2 - \sigma^2$ and consider only the linear independence of $(gm-h)_i$.
(2) We show that $(gm-h)_3 \sim x e^{\beta_1 K x}$, while $(gm-h)_{1,2} \sim e^{\beta_1 K x}$ as $x \to -\infty$. This allows us to consider only $(gm-h)_1$ and $(gm-h)_2$.
(3) We show that any linear combination of $(gm-h)_1$ and $(gm-h)_2$ can be split into two parts with different order of convergence to zero. This will yield linear independence of $(gm-h)_1$ and $(gm-h)_2$.

Taking all three arguments together, we obtain the linear independence of the total system (19).

However, before we start with these steps, we need to introduce some preliminary considerations. We define functions $u_i(x)$, $i = \overline{1,4}$, which are the solutions to the following deconvolution problems:

$$\mathbf{E}\,(u_1|\xi) = \mathrm{e}^{-\eta}, \quad \mathbf{E}\,(u_2|\xi) = \mathrm{e}^{-\eta}\xi, \quad \mathbf{E}\,(u_3|\xi) = \mathrm{e}^{-\eta}\eta, \quad \mathbf{E}\,(u_4|\xi) = \mathrm{e}^{-\eta}\eta\xi. \tag{21}$$

The explicit forms of $u_k(x)$ are:

$$u_1(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \exp\left\{k\beta_0 + k\beta_1 x - \frac{k^2\beta_1^2\sigma_\delta^2}{2}\right\}, \tag{22}$$

$$u_2(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (x - k\beta_1\sigma_\delta^2) \exp\left\{k\beta_0 + k\beta_1 x - \frac{k^2\beta_1^2\sigma_\delta^2}{2}\right\}, \tag{23}$$

$$u_3(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \exp\left\{(k+1)\beta_0 + (k+1)\beta_1 x - (k+1)^2\frac{\beta_1^2\sigma_\delta^2}{2}\right\}, \tag{24}$$

$$u_4(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \{x - (k+1)\beta_1\sigma_\delta^2\} \exp\left\{(k+1)\beta_0 + (k+1)\beta_1 x - (k+1)^2\frac{\beta_1^2\sigma_\delta^2}{2}\right\}. \tag{25}$$

Due to Fubinis theorem, we can exchange the order of summation and of computing $\mathbf{E}\,(u_k|\xi)$ and can thus check that the functions $u_k(x)$ given in Equations (22)–(25) are indeed the solutions to Equation (21). The series in Equations (22)–(25) converge uniformly on $(-\infty, x_0)$ for arbitrary $x_0 \in \mathbb{R}$. This yields the following asymptotic expansions for u_k as $x \to -\infty$:

$$\begin{aligned}
u_1(x) &= 1 - \exp\left\{\beta_0 + \beta_1 x - \frac{\beta_1^2\sigma_\delta^2}{2}\right\} + \frac{1}{2}\exp\{2\beta_0 + 2\beta_1 x - 2\beta_1^2\sigma_\delta^2\} + o(\mathrm{e}^{2\beta_1 x}),\\
u_2(x) &= x + o(x),\\
u_3(x) &= \exp\left\{\beta_0 + \beta_1 x - \frac{\beta_1^2\sigma_\delta^2}{2}\right\} - \exp\{2\beta_0 + 2\beta_1 x - 2\beta_1^2\sigma_\delta^2\} + o(\mathrm{e}^{2\beta_1 x}),\\
u_4(x) &= x\exp\left\{\beta_0 + \beta_1 x - \frac{\beta_1^2\sigma_\delta^2}{2}\right\} + o(x\mathrm{e}^{\beta_1 x}).
\end{aligned}$$

With the help of the functions $u_k(x)$, $k = \overline{1,4}$, we can write expressions for the first three rows of the matrix g and the vector h:

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \alpha + (1-\alpha)u_1 & \alpha\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\} \\ \alpha x + (1-\alpha)u_2 & \alpha(x - \beta_1\sigma_\delta^2)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\} \end{pmatrix}, \tag{26}$$

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = \begin{pmatrix} \alpha + (1-\alpha)u_1 \\ \alpha\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\} + (1-\alpha)u_3 \\ \alpha(x - \beta_1\sigma_\delta^2)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\} + (1-\alpha)u_4 \end{pmatrix}. \tag{27}$$

Therefore, the first three components of the vector $(gm - h)$ are:

$$\begin{pmatrix} m_2 - \alpha - (1-\alpha)u_1 \\ \alpha(m_1 - (1-m_2)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\}) + (1-\alpha)(u_1 m_1 - u_3) \\ \alpha(x m_1 - (1-m_2)(x - \beta_1\sigma_\delta^2)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\}) + (1-\alpha)(u_2 m_1 - u_4) \end{pmatrix}.$$

Remember that $m_1 = \mathrm{const}\cdot \mathrm{e}^{\beta_1 K x}$.

We establish the asymptotics of m_2 as $x \to -\infty$. Denote $f_0 = \mathbf{E}(f|x)$, where

$$f = f(t) = \exp\{-e^t\}, \quad t = \beta_0 + \beta_1\mu_1(x) + \beta_1\tau\gamma, \ \ \gamma \sim N(0,1).$$

Then $m_2 = \alpha + (1-\alpha)f_0$. Obviously, $t \sim \beta_1 Kx$ and $f(t) = 1 - e^t + 1/2e^{2t} + o(e^{2t})$ as $x \to -\infty$. By the dominated convergence theorem,

$$f_0 = 1 - \exp\left\{\beta_0 + \beta_1\mu_1(x) + \frac{\beta_1^2\tau^2}{2}\right\} + \frac{1}{2}\exp\{2\beta_0 + 2\beta_1\mu_1(x) + 2\beta_1^2\tau^2\} + o(e^{2\beta_1 Kx}).$$

It is now easy to see that $m_2 \to 1$ as $x \to -\infty$.

(1) Now we are ready for the first step. Consider a linear combination of the functions (19), which is zero for all x:

$$c_1(gm-h)_1 + c_2(gm-h)_2 + c_3(gm-h)_3 + c_4(x-\mu) + c_5((x-\mu)^2 - \sigma^2) \equiv 0. \quad (28)$$

From the asymptotic expressions for the functions $m_1, m_2, u_1, \dots, u_4$ it is easily seen that the functions $(gm-h)_i$ vanish as $x \to -\infty$, $i = \overline{1,3}$. Therefore, the coefficients c_4 and c_5 in Equation (28) must be equal to zero.

(2) Now we establish the asymptotic behavior of $(gm-h)_i$. Consider $(gm-h)_1$:

$$(gm-h)_1 = (1-\alpha)(f_0 - 1) + (1-\alpha)(1-u_1).$$

As $f_0 - 1 \sim \mathrm{const}\cdot e^{\beta_1 Kx}$, $1 - u_1 \sim \mathrm{const}\cdot e^{\beta_1 x}$ and $e^{\beta_1 x} = o(e^{\beta_1 Kx})$, we have

$$(gm-h)_1 \sim \ \mathrm{const}\cdot e^{\beta_1 Kx}, \ \ x \to -\infty.$$

Consider $(gm-h)_2$. We have $(1-m_2)e^{\beta_1 x} = o(e^{\beta_1 Kx})$, $u_3 = o(e^{\beta_1 Kx})$ and $u_1 m_1 \sim m_1$. Therefore,

$$(gm-h)_2 \sim m_1 \sim \mathrm{const}\cdot e^{\beta_1 Kx}, \ \ x \to -\infty.$$

Consider $(gm-h)_3$. We have $(1-m_2)(x - \beta_1\sigma_\delta^2)e^{\beta_1 x} = o(xe^{\beta_1 Kx})$, $u_4 = o(xe^{\beta_1 Kx})$, and $u_2 m_1 \sim xm_1$. Therefore,

$$(gm-h)_3 \sim xm_1 \sim \mathrm{const}\cdot xe^{\beta_1 Kx}, \quad x \to -\infty.$$

We see that $(gm-h)_1 = o((gm-h)_3)$ and $(gm-h)_2 = o((gm-h)_3)$ as $x \to -\infty$. Therefore, the coefficient c_3 in Equation (28) must be equal to zero.

(3) Now we can rewrite Equation (28) in the equivalent form:

$$\begin{aligned} &c_1(1-\alpha)(f_0 - u_1) + c_2\alpha(m_1 - (1-\alpha)(1-f_0)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\}) \\ &\quad + c_2(1-\alpha)(u_1 m_1 - u_3) \equiv 0. \end{aligned}$$

We rewrite it once again:

$$\begin{aligned} c_1(1-\alpha)(f_0 - 1) + c_2 m_1(\alpha + (1-\alpha)u_1) &\equiv c_1(1-\alpha)(u_1 - 1) \\ &\quad + c_2\alpha(1-\alpha)(1-f_0)\exp\{\beta_0 + \beta_1 x - \beta_1^2\sigma_\delta^2/2\} + c_2(1-\alpha)u_3. \quad (29) \end{aligned}$$

The left-hand side of Equation (29) is approximated by

$$a_1(1-\alpha)(c_2-c_1)\mathrm{e}^{\beta_1 K x} + a_2(1-\alpha)c_1\mathrm{e}^{2\beta_1 K x} - a_3(1-\alpha)^2 c_2 \mathrm{e}^{\beta_1(K+1)x} + o(\mathrm{e}^{2\beta_1 K x})$$
$$= a_1(1-\alpha)(c_2-c_1)\mathrm{e}^{\beta_1 K x} + a_2(1-\alpha)c_1\mathrm{e}^{2\beta_1 K x} + o(\mathrm{e}^{2\beta_1 K x}), \quad x\to-\infty,$$

where a_i are positive constants. The right-hand side of Equation (29) is approximated by

$$a_4(1-\alpha)(c_2-c_1)\mathrm{e}^{\beta_1 x} + \alpha(1-\alpha)c_2 a_5 \mathrm{e}^{\beta_1(K+1)x} + o(\mathrm{e}^{\beta_1(K+1)x}), \quad x\to-\infty,$$

where a_i are also positive constants. We see that Equation (29) is possible only if $c_1=c_2=0$.

We proved that all the coefficients in Equation (28) are zero; therefore, the functions (19) are linearly independent.

3.3.2. *Part 2 of the proof*

We want to prove that the functions (20) are linearly independent. Due to the expressions for $m_{2\theta}$ in Section 3.1, we have to prove the linear independence of the functions

$$\left\{1-f_0,\ f_1,\ \mu_1(x)f_1+\beta_1\tau^2 f_2,\ \beta_1(x-\mu)\frac{\partial K}{\partial\sigma}f_1+\frac{1}{2}\beta_1^2\frac{\partial\tau^2}{\partial\sigma}f_2\right\},$$

where we denoted $f_i := \mathbf{E}\,(f^{(i)}|x)$. This can be transformed into the equivalent set

$$\left\{1-f_0,\ f_1,\ Kxf_1+\beta_1\tau^2 f_2,\ \beta_1\frac{\partial K}{\partial\sigma}xf_1+\frac{1}{2}\beta_1^2\frac{\partial\tau^2}{\partial\sigma}f_2\right\}.$$

The last two functions are a linear transformation of the functions xf_1 and f_2 with transformation matrix

$$T=\begin{pmatrix} K & \beta_1\tau^2 \\ \beta_1\dfrac{\partial K}{\partial\sigma} & \dfrac{1}{2}\beta_1^2\dfrac{\partial\tau^2}{\partial\sigma}\end{pmatrix}.$$

We have

$$\det T = -\beta_1^2 K\frac{\sigma_\delta^4}{\sigma^3},$$

which is not zero under $\beta_1\neq 0$. Therefore, to prove linear independence of the functions (20), we have to prove linear independence of the functions $\{1-f_0, f_1, f_2, xf_1\}$.

Consider a linear combination of these functions, which is zero:

$$c_0(1-f_0)+c_1f_1+c_2f_2+c_3xf_1\equiv 0. \tag{30}$$

We establish the asymptotic behavior of the functions in Equation (30) as $x\to-\infty$. We use the dominated convergence theorem. We have $1-f\sim\mathrm{e}^t$, $t\to-\infty$ and thus for $\gamma\sim N(0,1)$:

$$1-f_0\sim\mathbf{E}\,(\exp\{\beta_0+\beta_1\mu_1(x)+\beta_1\tau\gamma\}|x)=\mathrm{const}\cdot\mathrm{e}^{\beta_1 K x},\quad x\to-\infty.$$

Consider f_1. We have $f'=-\exp\{-\mathrm{e}^t\}\mathrm{e}^t$ and $f'\sim-\mathrm{e}^t$, $t\to-\infty$. Therefore,

$$f_1\sim\mathrm{const}\cdot\mathrm{e}^{\beta_1 K x},\quad x\to-\infty.$$

Consider f_2. We have $f''=\exp\{-\mathrm{e}^t\}\mathrm{e}^{2t}-\exp\{-\mathrm{e}^t\}\mathrm{e}^t$ and $f''\sim-\mathrm{e}^t$, $t\to-\infty$. Therefore,

$$f_2\sim\mathrm{const}\cdot\mathrm{e}^{\beta_1 K x},\quad x\to-\infty.$$

If we divide Equation (30) by xf_1 and take the limit as $x\to-\infty$, we see that $c_3=0$.

Consider the asymptotics of the functions f_i, $i = 0, 1, 2$, as $x \to +\infty$. We have for arbitrary $a \in \mathbb{R}$ that

$$\exp\{-\mathrm{e}^t\}\mathrm{e}^{at} \to 0, \quad t \to \infty.$$

Therefore,

$$f \to 0, \quad f' \to 0, \quad f'' \to 0, \quad t \to +\infty,$$

and by the dominated convergence theorem

$$f_i \to 0, \quad i = 0, 1, 2, \quad x \to +\infty.$$

Thus, in Equation (30), we have $c_0 = 0$. Now Equation (30) can be rewritten as

$$(c_1 - c_2)\mathbf{E}\,(\exp\{-\mathrm{e}^t\}\mathrm{e}^t|x) \equiv c_2\,\mathbf{E}\,(\exp\{-\mathrm{e}^t\}\mathrm{e}^{2t}|x).$$

We see that the asymptotics of the left- and the right-hand sides are different as $x \to -\infty$ because

$$\exp\{-\mathrm{e}^t\}\mathrm{e}^{2t} = o(\exp\{-\mathrm{e}^t\}\mathrm{e}^t), \quad t \to -\infty.$$

Therefore, $c_1 = c_2 = 0$, and the functions in Equation (30) are linearly independent.

4. Simulations

A simulation study with the zero-inflated Poisson model of Section 3 was conducted with a threefold objective: first to show that the estimation methods QS and MCS work, at least for large samples; second to corroborate the asymptotic results of the preceding theory and third to study the behaviour of the methods for small samples. A sample size of $n = 100$ was taken to be a small sample, while $n = 1000$ stood for a large sample. For smaller sample size (e.g. $n = 40$) the results became rather unstable and so we did not report on them. The following parameter values were fixed: $\mu_\xi = 0.5$, $\sigma_\xi^2 = 0.1$, $\sigma_\delta^2 = 0.1$, $\alpha = 0.6$, $\beta_0 = 0$, $\beta_1 = 0.5$. As a variant, $\sigma_\delta^2 = 0.05$ was also tried. We simulated $R = 1000$ samples (x_i, y_i), $i = 1, \ldots, n$, of size n and computed the QS and MCS estimates for each sample. Bias and variance were then estimated from the 1000 replications. In addition to QS and MCS, we also computed a naive estimator (NA), which estimates the parameters by ML without taking the measurement errors into account and, as a benchmark, the ML estimator from the error-free data (ξ_i, y_i), $i = 1, \ldots, n$, (Equation (16)).

For all estimation methods, μ is estimated by $\hat{\mu} = \bar{x}$. For MCS, σ^2 is estimated by s_x^2, the empirical variance of the sample x_i, $i = 1, \ldots, n$, while for QS, σ^2 has to be estimated jointly with the other parameters.

For QS, the multivariate conditional mean–variance model given x is set up with Equation (12) as the bivariate mean function and Equation (13) as the covariance matrix. The parameter μ is replaced with its estimate $\bar{x}$. Deviating from the definition in Section 3, we here denote the main parameter vector $(\alpha, \beta_0, \beta_1)^{\mathrm{T}}$ by θ, while σ is treated separately.

The QS, estimators of θ and σ are found by applying the method of iteratively reweighted least squares: Let θ_k and σ_k be the estimated values of θ and σ after the kth iteration. For each sample point (x_i, y_{1i}), $i = 1, \ldots, n$, the vectors and matrices $m(x_i, \theta_k, \sigma_k)$, $v(x_i, \theta_k, \sigma_k)$, $M(x_i, \theta_k, \sigma_k) := \partial/\partial\theta^{\mathrm{T}} m(x_i, \theta_k, \sigma_k)$ and the vector $y_i := (y_{1i}, y_{2i})$ are computed. The bivariate linear regression

$$y_i - m(x_i, \theta_k, \sigma_k) = M(x_i, \theta_k, \sigma_k)d_k + u_i, \quad i = 1, \ldots, n, \tag{31}$$

with $\mathrm{var}(u_i) = v(x_i, \theta_k, \sigma_k)$ is set up and is solved for d_k by weighted least squares. The value of θ in the next iteration is then given by $\theta_{k+1} = \theta_k + d_k$. The value of σ in step $k + 1$ is found by solving the last equation of the of the system (2), i.e.

$$\sum_{i=1}^{n}\{m_{1\sigma}(x_i), m_{2\sigma}(x_i)\}v^{-1}(x_i)\{y_i - m(x_i)\} + \sum_{i=1}^{n}\{\log\rho(x_i)\}_\sigma = 0$$

and is given by

$$\sigma_{k+1}^2 = s_x^2 + \frac{1}{n}\sum_{i=1}^{n} w(x_i,\theta_k,\sigma_k)^{\mathrm{T}} v^{-1}(x_i,\theta_k,\sigma_k)\{y_i - m(x_i,\theta_k,\sigma_k)\},$$

where

$$w(x,\theta,\sigma)^{\mathrm{T}} = 2\sigma_\delta^2\beta_1\left(x - \bar{x} + \frac{1}{2}\beta_1\sigma_\delta^2,\ (\alpha-1)\left\{(x-\bar{x})E[b\mathrm{e}^{-b}|x] + \frac{\sigma_\delta^2}{2\tau}E[b\mathrm{e}^{-b}\gamma|x]\right\}\right).$$

The elements of the matrix M are given in Equation (14). For the sake of convenience, we repeat the expressions for $m_{2\theta}$ but in a somewhat different form. Let $b = b(\gamma) = \exp(\beta_0 + \beta_1\mu_1(x) + \beta_1\tau\gamma)$, then

$$\begin{aligned} m_{2\alpha} &= 1 - E[\mathrm{e}^{-b}|x] \\ m_{2\beta_0} &= (\alpha-1)E[b\mathrm{e}^{-b}|x] \\ m_{2\beta_1} &= (\alpha-1)\{\mu_1(x)E[b\mathrm{e}^{-b}|x] + \tau E[b\mathrm{e}^{-b}\gamma|x]. \end{aligned}$$

In addition,

$$m_{2\sigma} = 2(\alpha-1)\beta_1\frac{\sigma_\delta^2}{\sigma^3}\left\{(x-\bar{x})E[b\mathrm{e}^{-b}|x] + \frac{\sigma_\delta^2}{2\tau}E[b\mathrm{e}^{-b}\gamma|x]\right\}.$$

The last two formulae differ from the corresponding formulae in Equation (14) in that the partial integration has not been carried out. Note that $K_\sigma = 2\sigma_\delta^2/\sigma^3$ and $\tau_\sigma = K_\sigma\sigma_\delta^2/(2\tau)$.

For MCS, we have to compute the elements of the first three rows of g and h, (Equation (17)). These are given by Equations (26) and (27). The infinite series (22)–(25) needed to compute Equations (26) and (27) have been truncated at the order of $k = 20$. With g and h so constructed, we can set up the MCS estimating equations:

$$\sum g(x_i,\theta)y_i - \sum h(x_i,\theta) = 0$$

and solve them for $\theta = (\alpha, \beta_0, \beta_1)^{\mathrm{T}}$.

The results for $\sigma_\delta^2 = 0.1$ are presented in Table 1. They show that the asymptotic theory is fully corroborated in samples of size $n = 1000$. There is only a negligible bias in the three-parameter estimates, except, of course, for the naive estimator. The variance of the QS estimates are all smaller than the corresponding ones of the MCS estimates. The variance of the naive estimates are still smaller and even smaller than those of ML, but then these estimates are inconsistent anyway. When compared with the other two parameters of the model, α is estimated very precisely by all estimation methods.

For small samples ($n = 100$), we have similar results, although they are not so clear. Some of the estimates have a small, but noticeable, bias (e.g. $\hat{\beta}_{0\mathrm{MCS}}$), and $\hat{\alpha}_{\mathrm{MCS}}$ has a slightly smaller variance than $\hat{\alpha}_{\mathrm{QS}}$. The variances for $n = 100$ are a bit more than 10-fold the variance for $n = 1000$. In 1% of the runs, the QS estimate could not be computed because of the occurrence of a nearly singular covariance matrix v.

Table 1. The bias and variance of α, β_0 and β_1 when $\sigma_\delta^2 = 0.1$.

	α		β_0		β_1	
Method	Bias	Var	Bias	Var	Bias	Var
$n = 100$						
QS	−0.0258649	0.0074825	−0.0924049	0.3186765	0.0077364	0.7629999
MCS	−0.0247392	0.0071527	−0.1171952	0.3455757	0.0659780	0.9449513
NA	−0.0129181	0.0063493	0.0799786	0.0953049	−0.2340004	0.1661142
ML	−0.0168983	0.0065117	−0.0736062	0.1465186	0.0387977	0.3267195
$n = 1000$						
QS	−0.0015413	0.0005445	−0.0006907	0.0234746	−0.0136166	0.0575812
MCS	−0.0016397	0.0005462	−0.0075067	0.0261445	−0.0003143	0.0665571
NA	0.0022237	0.0005151	0.1316479	0.0084422	−0.2441716	0.0136816
ML	−0.0011093	0.0005352	−0.0116880	0.0132224	0.0124668	0.0286879

Table 2. The bias and variance of α, β_0 and β_1 when $\sigma_\delta^2 = 0.05$.

	α		β_0		β_1	
Method	Bias	Var	Bias	Var	Bias	Var
$n = 100$						
QS	−0.0196895	0.0067698	−0.0803810	0.2183384	0.0142384	0.5357211
MCS	−0.0273286	0.0069701	−0.1479973	0.2983405	0.1260944	0.7589774
NA	−0.0108817	0.0065266	0.0551384	0.1174278	−0.1771726	0.2180307
ML	−0.0133309	0.0067456	−0.0403318	0.1522869	−0.0083097	0.3301717
$n = 1000$						
QS	−0.0022525	0.0005487	−0.0075886	0.0169651	−0.0019069	0.0404219
MCS	−0.0036343	0.0005601	−0.0106284	0.0192654	0.0072325	0.0444614
NA	0.0000362	0.0005220	0.0869474	0.0095582	−0.1688009	0.0186527
ML	−0.0019372	0.0005317	−0.0035439	0.0124334	−0.0056925	0.0274534

When $\sigma_\delta^2 = 0.05$, we have similar results both for $n = 100$ and $n = 1000$ (Table 2). The variances are somewhat smaller than the corresponding ones for $\sigma_\delta^2 = 0.1$. For $n = 1000$, the difference in the variances of QS and MCS estimates is very small.

As noted above, the variance of x is estimated differently depending on whether QS or MCS is the estimation method. For QS, σ_x^2 is estimated along with α, β_0 and β_1, while for MCS, σ_x^2 is estimated by the empirical variance of the sample values $x_1, \ldots, x_n$. Both estimates, however, differ only by a negligible amount.

5. Conclusion

We proved that in a multivariate mean–variance model, QS is optimal within the class of LS estimators, in the sense that the ACM of the QS estimator is smaller (in the Löwner-order sense) than the ACM of any LS estimator. The QS estimator that we considered is an extended QS estimator, which comprises the estimation of the (nuisance) parameters describing the distribution of the regressor variables.

An important model, where this result can be applied, is the measurement error model given by a mean–variance model in the error-free variables supplemented by a measurement equation, which relates the latent regressor variables to observable variables. In such a model, the parameters describing the distribution of the regressor variables can be considered to be nuisance parameters. In this context, the CS estimator, which is a special LS estimator, has been introduced as an alternative to QS. It is well known that QS is (asymptotically) more efficient than CS, albeit under the assumption that the nuisance parameters are known.

Recently, this result has been generalized to the case of unknown nuisance parameters by extending QS in the way indicated above. But this generalization was restricted to a univariate mean–variance model. With the extension to a multivariate model, we are able to analyse the zero-inflated log-linear Poisson measurement error model with a normally distributed regressor variable. (Before this extension, only the ordinary log-linear Poisson measurement error model was amenable to an analysis.) In this model, QS is strictly more efficient than CS if the slope parameter is not zero. The mean of the regressor is estimated in the usual way as the arithmetic mean of the observations, but the variance of the regressor must be estimated in a more complicated way taking the complete model into account.

A simulation study confirms these results. The results can be extended to a more realistic case where the distribution of regressor variable is a mixture of normals.

Acknowledgements

Support by the German Research Foundation and by the Alexander von Humboldt Foundation is gratefully acknowledged. A. Kukush is supported as well by the Swedish Institute grant SI-01424/2007.

References

[1] B. Armstrong, *Measurement error in the generalized linear model*, Commun. Stat. Simul. Comput. 14 (1985), pp. 529–544.

[2] R.J. Carroll, D. Ruppert, L.A. Stefanski, and C.M. Crainiceanu, *Measurement Error in Nonlinear Models–A Modern Perspective*, 2^{nd} edn., Chapman and Hall, London, 1996.

[3] C.C. Heyde, *Quasi-Likelihood and Its Application*, Springer, New York, 1997.

[4] R.W.M. Wedderburn, *Quasi likelihood functions, generalized linear models, and the Gauss–Newton method*, Biometrika 61 (1974), pp. 439–447.

[5] O.G. Kukush, A.L. Malenko, and H. Schneeweiss, *Optimality of the quasi score estimator in a mean–variance model with applications to measurement error models*, Discussion Paper 494 (2006), SFB 386, Universität München.

[6] A.C. Cameron and P.K. Trivedi, *Regression Analysis of Count Data*, Cambridge University Press, Cambridge, 1998.

[7] C. Czado and A. Min, *Testing for zero-modification in count regression models*, Discussion Paper 494 (2006), SFB 474, Universität München.

[8] D. Lambert, *Zero-inflated poisson regression, with an application to defects in manufacturing*, Technometrics 34 (1992), pp. 1–14.

[9] O.G. Kukush and H. Schneeweiss, *Comparing different estimators in a nonlinear measurement error model. I*, Math. Methods Stat., 14 (2005), pp. 53–79.

[10] M.J. Schervish, *Theory of Statistics*, Springer, New York, 1995.

[11] T. Nakamura, *Corrected score function for errors-in-variables models*, Biometrika 77 (1990), pp. 127–137.

[12] L. Stefanski, *Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models*, Commun. Stat., Part A Theory Methods 18 (1989), pp. 4335–4358.

[13] O.G. Kukush, A.L. Malenko, and H. Schneeweiss, *Comparing the efficiency of estimates in concrete errors-in-variables models under unknown nuisance parameters.* Theory Stoch. Process. 13 (29) (2007), N4, pp. 69–81.