

**AN OPTIMAL JOINT ESTIMATOR FOR  
REGRESSION PARAMETERS AND THE DISPERSION PARAMETER  
IN ERRORS-IN-VARIABLES NONLINEAR MODELS**

UDC 519.21

A. L. MALENKO AND O. G. KUKUSH

*Dedicated to our teachers, Anatoliĭ Yakovych Dorogovtsev  
and Mikhaĭlo Iosypovych Yadrenko*

ABSTRACT. We consider an errors-in-variables nonlinear structural model where the density of the response belongs to the exponential family. We estimate regression parameters and the dispersion parameter as well as parameters of the hidden variable. Following the modified quasi-likelihood method we construct a joint estimator that has the minimal asymptotic covariance matrix in a wide class of estimators. The polynomial and gamma models are studied in more detail.

1. INTRODUCTION

Let a relation between a response  $y$  and a random variable  $x$  be described by the conditional mean and conditional variance:

$$(1) \quad \mathbf{E}(y|x) = m(x, \beta, \alpha), \quad \mathbf{V}(y|x) = v(x, \beta, \alpha, \varphi).$$

Here  $\beta$  is a column vector of parameters of the regression,  $\alpha$  is the column vector of nuisance parameters that determine the distribution of  $x$ , and where  $\varphi$  is the dispersion parameter. The above models are called mean-variance models (see [1]). We want to estimate  $\beta$ ,  $\alpha$ , and  $\varphi$  from observations (that is, from a sample of independent identically distributed vectors  $(x_i, y_i)$ ,  $i = 1, \dots, n$ ). In what follows the partial derivatives of functions of several arguments are written with subscripts indicating the corresponding arguments. The derivatives with respect to a vector argument are column vectors of partial derivatives.

If  $\alpha$  and  $\varphi$  are known, the quasi-likelihood estimator  $\hat{\beta}_Q$  of the parameter  $\beta$  is constructed as follows:

a) introduce the score function

$$(2) \quad S_Q(y, x, \beta) = \frac{y - m}{v} m_\beta,$$

where  $m$  and  $v$  are defined by (1),

b) the estimator  $\hat{\beta}_Q$  is defined as a measurable solution of the equation

$$\sum_{i=1}^n S_Q(y_i, x_i, \beta) = 0.$$

---

2000 *Mathematics Subject Classification.* Primary 62J02; Secondary 62J10, 62J12, 62H12, 62F12.

*Key words and phrases.* Exponential family of densities, errors-in-variables models, polynomial model, gamma model, quasi-likelihood method, asymptotic effectiveness of estimators.

The estimator  $\hat{\beta}_Q$  is consistent and asymptotically normal under natural regularity conditions [2]. It is shown in [2, 3] that its asymptotic covariance matrix is minimal in the class of all estimators generated by score functions that are unbiased and linear in  $y$ .

The quasi-likelihood method is extended in [4] to the case where the only known parameter is  $\varphi$ . The corresponding score function is given by

$$S_{Q1}(y, x; \beta, \alpha, \varphi) = \begin{pmatrix} (y - m)v^{-1}m_\beta \\ (y - m)v^{-1}m_\alpha + l_\alpha \end{pmatrix},$$

where  $l$  is the logarithm of the density of  $x$ . The function  $S_{Q1}$  generates the estimators  $(\hat{\beta}_{Q1}^t, \hat{\alpha}_{Q1}^t)^t$  that also have the asymptotic covariance matrix in the class of estimators generated by unbiased and linear-in- $y$  score functions. Here and in what follows the symbol  $^t$  means the transposition.

The case where all three parameters  $\beta$ ,  $\alpha$ , and  $\varphi$  are unknown is considered in the paper [4]. Since the conditional mean does not depend on  $\varphi$ , while the conditional variance does, one needs to estimate  $\varphi$ , too. For this reason, an additional component

$$(3) \quad S_{Q1}^{(\varphi)} = (y - m)^2 - v \text{ or } y^2 - m^2 - v$$

should be added to the score function  $S_{Q1}$ . This leads to the estimator  $(\tilde{\beta}_{Q1}^t, \tilde{\alpha}_{Q1}^t, \tilde{\varphi}_{Q1}^t)^t$ . Its asymptotic covariance matrix is given by

$$\Sigma_{Q1}^{(\beta, \alpha, \varphi)} = \begin{bmatrix} \Sigma^{(\beta, \alpha)} & A \\ A^t & \sigma_{(\varphi)}^2 \end{bmatrix}.$$

Here  $\Sigma^{(\beta, \alpha)}$  is the asymptotic covariance matrix of the estimator  $(\tilde{\beta}_{Q1}^t, \tilde{\alpha}_{Q1}^t)^t$  and  $\sigma_{(\varphi)}^2$  is the asymptotic variance of the estimator  $\tilde{\varphi}_{Q1}$ . It turns out that  $\Sigma^{(\beta, \alpha)}$  coincides with the asymptotic covariance matrix of the estimator  $(\hat{\beta}_{Q1}^t, \hat{\alpha}_{Q1}^t)^t$ . This means that the asymptotic effectiveness of the other estimators does not depend on whether or not  $\varphi$  is known.

A question arises on how to extend the quasi-likelihood method to the case where the parameters  $\beta$ ,  $\alpha$ , and  $\varphi$  are unknown and to obtain an optimal joint estimator in the class of unbiased score functions that are quadratic with respect to  $y$ ? This paper is devoted to a solution of this problem.

The idea is to consider the mean-variance model for a vector response

$$Y = (y, (y - m)^2)^t.$$

This model requires not only the first two conditional moments of  $y$  given  $x$  but also the third and fourth conditional moments. The moment functions (of  $x$ ,  $\beta$ ,  $\alpha$ , and  $\varphi$ ) can be constructed explicitly in many cases. The mean-variance model (1) appears in a natural way when considering the structural errors-in-variables regression models where the response belongs to the exponential family of densities [2, 3, 4]. The moment functions for such (nonlinear) models are expressed in terms of  $\varphi$  and derivatives of the structural function  $C(\eta)$  (see Section 2).

The paper is organized as follows. The general errors-in-variables nonlinear regression model is introduced in Section 2. The quasi-likelihood method is extended in two ways to the case where  $\beta$ ,  $\alpha$ , and  $\varphi$  are unknown. We consider the class of quadratic-in- $y$  unbiased score functions and compare the estimators generated by these functions with the quasi-likelihood estimators. The partial cases are discussed in Sections 4 and 5 for the polynomial and gamma models. The appendices contain the proof of the lemma and two theorems.

## 2. GENERAL NONLINEAR REGRESSION MODEL

Throughout the paper, the symbol  $\mathbf{E}$  stands for the expectation of a random variable, vector, or matrix. Let  $\mathbf{V}$  denote the covariance function of a random vector. The operator  $\mathbf{E}$  is applied to the whole expression to the right of the symbol  $\mathbf{E}$ . We assume that  $\mathbf{E}F(z, \theta) = \mathbf{E}_\theta F(z, \theta)$  where  $z$  is a vector of observations and  $\theta$  is the true value of the vector parameter determining the distribution of  $z$ . The convergence of random variables in distribution is denoted by  $\xrightarrow{d}$ .

Consider an errors-in-variables nonlinear regression model. Let  $\nu$  be a  $\sigma$ -finite measure on the Borel  $\sigma$ -algebra in  $\mathbb{R}$ . A random variable  $y$  is observed. This random variable is assumed to have the distribution density  $f(y|\eta)$  with respect to the measure  $\nu$ , and the density belongs to the exponential family:

$$(4) \quad f(y|\eta) = \exp \left\{ \frac{y\eta - C(\eta)}{\varphi} + c(y, \varphi) \right\}.$$

The function  $C(\cdot)$  is smooth enough,  $C'' > 0$ , and  $c(y, \varphi)$  is measurable and does not depend on  $\eta$ . The parameter  $\varphi > 0$  is called the dispersion parameter.

Assume that  $\eta = \eta(\xi, \beta)$ , where  $\xi$  is a random scalar variable,  $\beta$  is a nonrandom column vector of regression parameters,  $\beta \in \mathbb{R}^{k+1}$ . As usual

$$\beta = (\beta_0, \beta_1, \dots, \beta_k)^t,$$

where  $\beta_0$  is a free term (intercept). The random variable  $x = \xi + \delta$  is observed where  $\xi$  and  $\delta$  are independent, and  $\delta$  is a measurement error.

Let the random vectors  $(y_i, \xi_i, \delta_i)$  be independent and identically distributed for  $i = 1, \dots, n$ ;  $\xi_i \sim \mathcal{N}(\mu, \sigma_\xi^2)$  and  $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$ , where  $\sigma_\xi^2$  and  $\sigma_\delta^2$  are positive numbers. The variance  $\sigma_\xi^2$  is assumed to be known. These vectors are a realization of the model (4). We have  $x_i = \xi_i + \delta_i \sim \mathcal{N}(\mu, \sigma^2)$ , where  $\sigma^2 = \sigma_\xi^2 + \sigma_\delta^2$ .

We consider the following two main cases: a) the parameters  $\mu$  and  $\sigma$  are known, and b) the parameters  $\mu$  and  $\sigma$  are unknown. The combined vector of all unknown parameters of the model is denoted by  $\theta$ . In the first case,  $\theta = (\beta^t, \varphi)^t$ , while

$$\theta = (\beta^t, \mu, \sigma, \varphi)^t$$

in the second case. For both cases,  $d := \dim \theta$ . Let  $\theta'$  be the vector constituted of the components of  $\theta$  where the parameter  $\varphi$  is missing. The logarithm of the density of  $x$  is written up to a constant term as follows:

$$l(x, \mu, \sigma) = -\frac{(x - \mu)^2}{2\sigma^2} - \ln \sigma.$$

## 3. ESTIMATORS

We extend the quasi-likelihood method to the case of an unknown parameter  $\varphi$  and construct two estimators of unknown parameters. Consider the conditional expectation and variance of the response given  $x$ :

$$m(x, \theta) = \mathbf{E}[y|x], \quad v(x, \theta) = \mathbf{E}[(y - m(x, \theta))^2|x].$$

The components of the first score functions are

$$S_{Q1}^{(\theta')} = \frac{y - m}{v} m_{\theta'} + l_{\theta'}, \quad S_{Q1}^{(\varphi)} = (y - m)^2 - v$$

and the first score function can be written as follows:

$$(5) \quad S_{Q1} = \left( S_{Q1}^{(\theta')}, S_{Q1}^{(\varphi)} \right)^t.$$

The estimator  $\hat{\theta}_{Q1} = ((\hat{\theta}'_{Q1})^t, \hat{\varphi}_{Q1})^t$  is defined as a measurable solution of the following system of equations:

$$(6) \quad \sum_{i=1}^n S_{Q1}(y_i, x_i, \theta) = 0, \quad \theta \in \Theta,$$

where  $\Theta$  is an open set in  $\mathbb{R}^d$  that contains the true value of the parameter  $\theta$ . When constructing the set  $\Theta$  one should take into account that the true parameters are such that  $\varphi > 0$  and  $\sigma > 0$ .

It is proved in [4] for the case where  $\varphi$  is known that the estimator  $\hat{\theta}'_{Q1}$  has the minimal asymptotic covariance matrix among all estimators constructed from unbiased and linear-in- $y$  score functions.

The conditional expectation  $m$  does not depend on  $\varphi$ , while the conditional variance  $v$  does depend on  $\varphi$ . Thus the class of unbiased linear-in- $y$  score functions is not adequate to estimate the unknown parameter  $\varphi$ . It is natural to consider unbiased score functions that are polynomials of the second order in  $y$  (cf. (3)), namely

$$S_L = f(x, \theta)y^2 + g(x, \theta)y + h(x, \theta), \quad \mathbf{E} S_L = 0,$$

where  $\dim f, g, h = d$ , and to assume some regularity conditions sufficient for the consistency and asymptotical normality of the estimator constructed with the help of the function  $S_L$ . The regularity conditions of this kind related to the theory of score equations [5] are discussed in [2].

Put

$$Y = \begin{pmatrix} y \\ (y - m)^2 \end{pmatrix},$$

$$u(x, \theta) = u := \mathbf{E} [(y - m)^3 | x], \quad w(x, \theta) = w := \mathbf{E} [(y - m)^4 | x].$$

An unbiased score function  $S_L$  that is a second order polynomial in  $y$  can be represented as follows:

$$S_L(y, x, \theta) = G(x, \theta) \cdot Y - H(x, \theta), \quad \mathbf{E} S_L = 0,$$

where  $G(x, \theta)$  is a  $d \times 2$  matrix and  $\dim H(x, \theta) = d$ .

In addition to the assumption that the score function is unbiased, one should require that the limit equation have a unique solution. To assure the uniqueness we consider the limit equation

$$(7) \quad \mathbf{E}_{\theta_0} S_L(y, x, \theta) = 0, \quad \theta \in \Theta.$$

Here  $\theta_0$  is a true value of the parameter  $\theta$ . We assume that the family of parameters  $\Theta$  is a convex compact set. If the matrix  $\Phi(\theta, \theta_0) := \mathbf{E}_{\theta_0} S_{L\theta}(y, x, \theta)$  is nondegenerate for all  $\theta \in \Theta$ , then equation (7) has a unique solution  $\theta = \theta_0$ . Usually we need the property that  $\theta_0$  is an interior point of  $\Theta$ . Since  $\theta_0$  is unknown, it is natural to assume that

$$(8) \quad \text{the matrix } \Phi(\theta, \theta_1) \text{ is nondegenerate for all } \theta \in \Theta, \theta_1 \in \Theta^0,$$

to achieve the uniqueness of the solution of equation (7), where  $\Theta^0$  is the set of all interior points of  $\Theta$ . Note that the matrix  $\Phi(\theta, \theta)$  is nondegenerate for all  $\theta \in \Theta$  (in particular, this matrix is negative definite for quasi-likelihood estimators). Since the matrix-valued function  $\Phi$  is continuous in all its arguments, condition (8) holds if  $\theta$  belongs to a sufficiently small neighborhood of a fixed point  $\theta_1 \in \Theta^0$ .

For particular models, one needs to keep condition (8) and control the size of the set  $\Theta$ . In what follows we assume that condition (8) is satisfied for all score functions  $S_L(y, x, \theta)$  considered in the paper.

To define a new estimator consider the conditional expectation of the random vector  $Y$  and its conditional covariance matrix given  $x$ :

$$M(x, \theta) = M := \mathbf{E}[Y|x] = \begin{pmatrix} m \\ v \end{pmatrix}, \quad V(x, \theta) = V := \mathbf{V}[Y|x] = \begin{pmatrix} v & u \\ u & w - v^2 \end{pmatrix}.$$

We assume that the matrix  $V(x, \theta)$  is positive definite for all  $x$  and  $\theta$ . Define a new score function by

$$(9) \quad S_{Q2} = M_\theta^t \cdot V^{-1} \cdot (Y - M) + l_\theta.$$

The estimator  $\hat{\theta}_{Q2} = ((\hat{\theta}'_{Q2})^t, \hat{\varphi}_{Q2})^t$  is defined as a measurable solution of the system of equations

$$(10) \quad \sum_{i=1}^n S_{Q2}(y_i, x_i, \theta) = 0, \quad \theta \in \Theta.$$

Under some regularity conditions, similar to those in Theorem 4.1 of [2], the systems (6) and (10) possess solutions eventually (that is, with probability one as  $n \geq n_0(\omega)$ ); if one of the systems has no solution, then we take  $\hat{\theta} = \theta_0$  as the estimator of  $\theta$ , where  $\theta_0$  is a fixed point of  $\Theta$ .

The following generalization of the main theorem of the paper [4] to the case of a vector response  $Y$  can be proved similarly to [4].

**Theorem 3.1.** *Let  $\Sigma_L$  and  $\Sigma_{Q2}$  denote the asymptotic covariance matrices of the estimators generated by the score functions  $S_L$  and  $S_{Q2}$ , respectively. Then:*

- a)  $\Sigma_L \geq \Sigma_{Q2}$ ; moreover, if  $\Sigma_L = \Sigma_{Q2}$  for all  $\theta$ , then  $\hat{\theta}_L = \hat{\theta}_{Q2}$  almost surely.
- b)  $\text{rank}(\Sigma_L - \Sigma_{Q2}) + d$  is equal to the rank of the system of random vectors of the Hilbert space  $L_2(\Omega, \mathcal{F}, \mathbf{P}; \mathbb{R}^3)$ :

$$\text{rank} \left[ \begin{pmatrix} (GV)_{i1} \\ (GV)_{i2} \\ (GM - H)_i \end{pmatrix}, \begin{pmatrix} (M_1)\theta_i \\ (M_2)\theta_i \\ l_{\theta_i} \end{pmatrix}, i = 1, \dots, d \right].$$

The asymptotic covariance matrices of the estimators generated by the score functions  $S_{Q1}$  and  $S_{Q2}$  can be compared with the help of Theorem 3.1. Indeed, the score function  $S_{Q1}$  belongs to the class of functions that are linear in  $Y$ :

$$S_{Q1} = G_1 \cdot Y - H_1, \quad G_1 = \begin{pmatrix} v^{-1}m_{\theta'} & 0 \\ 0 & 1 \end{pmatrix}, \quad H_1 = \begin{pmatrix} v^{-1}mm_{\theta'} - l_{\theta'} \\ v \end{pmatrix}.$$

**Corollary 3.1.** *Let the parameters  $\mu$  and  $\sigma$  be known and  $\theta = (\beta^t, \varphi)^t$ . Then*

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2})$$

*is equal to*

$$\text{rank} \left[ \begin{pmatrix} 0 \\ v^{-1}um_{\beta_i} - v_{\beta_i} \end{pmatrix}, \begin{pmatrix} m_{\beta_i} \\ v_{\beta_i} \end{pmatrix}, i = 0, \dots, k, \begin{pmatrix} u \\ w - v^2 \end{pmatrix}, \begin{pmatrix} 0 \\ v_\varphi \end{pmatrix} \right] - d.$$

**Corollary 3.2.** *Let the parameters  $\mu$  and  $\sigma$  be unknown and  $\theta = (\beta^t, \mu, \sigma, \varphi)^t$ . Then*

$$\begin{aligned} & \text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) \\ &= \text{rank} \left[ \begin{pmatrix} 0 \\ v^{-1}um_{\beta_i} - v_{\beta_i} \end{pmatrix}, \begin{pmatrix} m_{\beta_i} \\ v_{\beta_i} \end{pmatrix}, i = 0, \dots, k, \right. \\ & \quad \left. \begin{pmatrix} 0 \\ v^{-1}m_\mu u - v_\mu \end{pmatrix}, \begin{pmatrix} 0 \\ v^{-1}m_\sigma u - v_\sigma \end{pmatrix}, \begin{pmatrix} 0 \\ v_\varphi \end{pmatrix}, \begin{pmatrix} u \\ w - v^2 \end{pmatrix} \right] - (d - 2). \end{aligned}$$

In the rest of the paper we consider particular cases of model (4). These models are such that the regularity conditions, similar to [2], hold, and this implies that the estimators  $\hat{\theta}_{Q_1}$  and  $\hat{\theta}_{Q_2}$  are asymptotically normal.

#### 4. POLYNOMIAL MODEL

One observes the random variables

$$(11) \quad \begin{cases} y = \beta_0 + \beta_1 \xi + \cdots + \beta_k \xi^k + \varepsilon, \\ x = \xi + \delta, \end{cases}$$

such that  $\xi \sim N(\mu, \sigma_\xi^2)$ ,  $\delta \sim N(0, \sigma_\delta^2)$ , and  $\varepsilon \sim N(0, \varphi)$ . We assume that the random variables  $\xi$ ,  $\delta$ ,  $\varepsilon$  are jointly independent and that all their variances are positive. Put

$$\sigma^2 = \sigma_\xi^2 + \sigma_\delta^2, \quad K = 1 - \sigma_\delta^2 \sigma^{-2},$$

and  $\tau^2 = \sigma_\delta^2 K$ . Model (11) is a particular case of (4) for  $C(\eta) = \eta^2/2$ ,

$$\eta(\xi, \beta) = \beta_0 + \beta_1 \xi + \cdots + \beta_k \xi^k.$$

The polynomial model is such that

$$m(x, \beta) = \beta^t r, \quad v(x, \beta) = \beta^t (R - r r^t) \beta + \varphi,$$

where the components of the vector  $r = (r_0, r_1, \dots, r_k)^t$  are such that  $r_i(x) = \mathbf{E}[\xi^i | x]$  and  $R = (r_{i+j}, i, j = 0, \dots, k)$  (see [7]).

The following result helps to make the system of score equations simpler in the case of unknown nuisance parameters  $\mu$  and  $\sigma$ .

**Lemma 4.1.** *There is a linear relation between the derivatives of  $m$  and  $v$ , namely*

$$(12) \quad m_\mu = (1 - K)\beta^t D m_\beta, \quad v_\mu = (1 - K)\beta^t D v_\beta.$$

*All the elements  $d_{ij}$ ,  $i, j = 0, \dots, k$ , of the matrix  $D$  are zero except for  $d_{i, i-1} = i$ ,  $i = 1, \dots, k$ .*

Consider a *linear* model, that is, model (11) for  $k = 1$ . Then

$$m = \beta_0 + \beta_1(Kx + (1 - K)\mu), \quad v = \varphi + \beta_1^2 \tau^2, \quad u = 0, \quad w = 3v^2.$$

It is easy to see that  $S_{Q_2} = H(\theta)S_{Q_1}$ , where the matrix  $H$  is nonrandom, nondegenerate, and depends on  $\theta$  only. Thus the estimators  $\hat{\theta}_{Q_1}$  and  $\hat{\theta}_{Q_2}$  coincide almost surely. This implies that the estimator obtained with the help of the score function  $S_{Q_1}$  has a maximal asymptotic effectiveness in the class of unbiased quadratic polynomials of  $y$ .

**Theorem 4.1.** *Let the true value in the model (11) be such that  $\beta_k \neq 0$  and let  $k \geq 2$ . Then*

- a)  $\Sigma_{Q_1} > \Sigma_{Q_2}$  if the parameters  $\mu$  and  $\sigma$  are known and  $\theta = (\beta^t, \varphi)$ ;
  - b)  $\hat{\mu}_{Q_1} = \hat{\mu}_{Q_2} = \bar{x}$  if the parameters  $\mu$  and  $\sigma$  are unknown and  $\theta = (\beta^t, \mu, \sigma, \varphi)^t$ ;
- moreover

$$d - 2 \leq \text{rank} \left( \Sigma_{Q_1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q_2}^{(\beta, \sigma, \varphi)} \right) \leq d - 1.$$

*In particular, if  $k = 2$  and  $d = 6$ , then*

$$\text{rank} \left( \Sigma_{Q_1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q_2}^{(\beta, \sigma, \varphi)} \right) = 4.$$

*Here  $\Sigma^{(\beta, \sigma, \varphi)}$  denotes the joint asymptotic covariance matrix of the estimators  $\hat{\beta}$ ,  $\hat{\sigma}$ , and  $\hat{\varphi}$ .*

## 5. GAMMA MODEL

Let

$$(13) \quad f(y|\eta) = \frac{1}{\Gamma(b)} \left(\frac{b}{a}\right)^b y^{b-1} e^{-yb/a}, \quad a = \exp\{\beta_0 + \beta_1 \xi\}, \quad b = \varphi^{-1},$$

where  $\beta = (\beta_0, \beta_1)^t$ . We observe  $y$  and let  $x = \xi + \delta$ . The assumptions are  $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$  and  $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$ . We also assume that both variances are positive and the random variables  $\xi$  and  $\delta$  are independent.

The gamma model is a particular case of (4) for  $\eta(\xi, \beta) = -a^{-1}$  and  $C(\eta) = -\ln(-\eta)$ . Note that (13) is a generalized linear model.

Put  $c = \exp\{\beta_1^2 \tau^2\}$ . Then

$$\begin{aligned} m(x, \theta) &= \exp\{\beta_0 + \beta_1 \mu_1(x) + \beta_1^2 \tau^2 / 2\}, & v(x, \theta) &= (c(\varphi + 1) - 1)m^2, \\ u(x, \theta) &= (c^3(1 + 2\varphi)(1 + \varphi) - 3c(\varphi + 1) + 2)m^3, \\ w(x, \theta) &= (c^6(1 + 3\varphi)(1 + 2\varphi)(1 + \varphi) - 4c^3(1 + 2\varphi)(1 + \varphi) + 6c(\varphi + 1) - 3)m^4. \end{aligned}$$

Since  $c \geq 1$ ,

$$c^3(1 + 2\varphi)(1 + \varphi) - 3c(\varphi + 1) + 2 \geq 2\varphi^2 > 0,$$

whence  $u(x, \theta) > 0$  for all  $x$  and  $\theta$ . Similarly,  $w - v^2 > 0$  for all  $x$  and  $\theta$ .

Now we evaluate the derivatives of the functions  $m$  and  $v$ :

$$\begin{aligned} m_{\beta_0} &= m, & m_{\beta_1} &= (\mu_1(x) + \beta_1 \tau^2) m, \\ v_{\beta_0} &= 2v, & v_{\beta_1} &= 2(\mu_1(x) + \beta_1 \tau^2) v + 2c\beta_1 \tau^2 (\varphi + 1) m^2, & v_\varphi &= cm^2. \end{aligned}$$

**Theorem 5.1.** *Let the true value of the parameter in the model (13) be such that  $\beta_1 \neq 0$ . Then*

- $\text{rank}(\Sigma_{Q_1} - \Sigma_{Q_2}) = 2$  if the parameters  $\mu$  and  $\sigma$  are known and  $\theta = (\beta_0, \beta_1, \varphi)$ ;
- $\hat{\mu}_{Q_1} = \hat{\mu}_{Q_2} = \bar{x}$  if the parameters  $\mu$  and  $\sigma$  are unknown and  $\theta = (\beta_0, \beta_1, \mu, \sigma, \varphi)^t$ ;  
moreover  $\text{rank}(\Sigma_{Q_1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q_2}^{(\beta, \sigma, \varphi)}) = 2$ .

## 6. CONCLUDING REMARKS

We considered two quasi-likelihood estimators in the errors-in-variables linear structural regression model with the response belonging to the exponential family of densities.

The second (new) estimator has a smaller asymptotic covariance matrix for the polynomial model whose degree is greater than one if the parameters of the distribution of the independent variable are known. This means that the estimator  $\hat{\theta}_{Q_2}$  has a better asymptotic effectiveness than another estimator  $\hat{\theta}_{Q_1}$ .

We showed that the difference between the asymptotic covariance matrices is nonnegative definite and nonzero if the nuisance parameters  $\mu$  and  $\sigma$  are unknown (this means that the estimator  $\hat{\theta}_{Q_2}$  is more effective in some directions than the estimator  $\hat{\theta}_{Q_1}$  is).

The difference of matrices is a nonnegative definite matrix of rank 2 in the gamma model for both cases of known or unknown parameters  $\mu$  and  $\sigma$ . Since  $\dim \theta \geq 3$ , the estimator  $\hat{\theta}_{Q_2}$  is more effective than  $\hat{\theta}_{Q_1}$  in some directions only in the case of the gamma model. Both estimators of the unknown parameter  $\mu$  coincide with the empirical mean value of the variable  $x$ , which is a characterizing feature of generalized linear models.

The question on the rank of the difference of asymptotic covariance matrices is still open for the polynomial model of degree  $k \geq 3$  if the nuisance parameters are unknown. This question will be discussed elsewhere.

## APPENDIX A. PROOF OF LEMMA 4.1

The distribution of the random variable  $\xi|x$  is normal with parameters  $Kx + (1-K)\mu$  and  $\tau^2$ . Then  $r_i(x) = \mathbf{E}[(Kx + (1-K)\mu + \tau\gamma)^i|x]$  for a normal random variable  $\gamma$  that is independent of  $x$ . Thus

$$\frac{\partial r_i(x)}{\partial \mu} = i(1-K) \mathbf{E}[(Kx + (1-K)\mu + \tau\gamma)^{i-1}|x] = i(1-K)r_{i-1}(x).$$

The vector  $r$  and matrix  $R$  are such that

$$r_\mu = (1-K)Dr, \quad R_\mu = (1-K)(DR + RD^t).$$

Now the first statement of the lemma follows:

$$m_\mu = \beta^t r_\mu = \beta^t(1-K)Dr = (1-K)\beta^t Dm_\beta.$$

For the proof of the second equality, consider two quadratic forms

$$\begin{aligned} (1-K)\beta^t Dv_\beta &= (1-K)\beta^t 2D(R - rr^t)\beta, \\ v_\mu &= (1-K)\beta^t(DR + RD^t - Drr^t - rr^t D^t)\beta. \end{aligned}$$

The matrices of both forms coincide and they are equal to the matrix

$$(1-K)(D(R - rr^t) + (R - rr^t)D^t).$$

## APPENDIX B. PROOF OF THEOREM 4.1

Let both parameters  $\mu$  and  $\sigma^2$  be known and  $\theta = (\beta^t, \varphi)^t$ . We assume that the true value of the parameter  $\beta_k$  is nonzero.

Then  $v_\varphi = 1$ . By Corollary 3.1,  $\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) + d$  is equal to

$$(14) \quad \text{rank} \left[ \begin{pmatrix} 0 \\ v^{-1}um_{\beta_i} - v_{\beta_i} \end{pmatrix}, \begin{pmatrix} m_{\beta_i} \\ v_{\beta_i} \end{pmatrix}, i = 0, \dots, k, \begin{pmatrix} u \\ w - v^2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right].$$

Let  $k = 2$ . Then  $\deg u = 2$ ,  $\deg m_\beta = (0, 1, 2)^t$ ,  $u = f^t m_\beta$ ,  $\deg f^t v_\beta = 2$ , and  $\deg(w - v^2 - f^t v_\beta) = 4$ . Thus

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}[um_{\beta_i} - vv_{\beta_i}, i = 0, 1, 2, pv, v] - 1, \quad p = w - v^2 - f^t v_\beta.$$

The degrees of the polynomials are such that  $\deg u = 2$ ,

$$\deg(um_{\beta_1} - vv_{\beta_1}) = 3, \quad \deg(um_{\beta_2} - vv_{\beta_2}) = 4, \quad \deg pv = 6,$$

and  $\deg v = 2$ . It is easy to see that

$$u + 2\tau^2\beta_2(3\varphi + 2\tau^4\beta_2^2) = 6\beta_2\tau^2v.$$

Hence  $u$  and  $v$  are linearly independent. The family of polynomials is of rank 5 and  $\Sigma_{Q2} < \Sigma_{Q1}$ .

Let  $k = 3$ . Then  $\deg u = 3k - 4 > k$  and  $\{u, m_{\beta_i}, i = 0, \dots, 3\}$  are linearly independent. One can exclude the first row in the system of random vectors (14):

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}[um_{\beta_i} - vv_{\beta_i}, i = 0, \dots, 3, v].$$

The degrees of the polynomials are  $(5, 6, 7, 7, 4)$ . One can show that, for some constants  $c_1, c_2, c_3$ , and  $c_4$ , the polynomial

$$p = um_{\beta_3} - vv_{\beta_3} - (c_0u + c_1(um_{\beta_1} - vv_{\beta_1}) + c_2(um_{\beta_2} - vv_{\beta_2}) + c_4v)$$

is nonzero if  $\beta_3 \neq 0$ . Thus  $\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = 5$  and  $\Sigma_{Q2} < \Sigma_{Q1}$ .

Let  $k \geq 4$ . Then  $\{u, m_{\beta_i}, i = 0, \dots, k\}$  are linearly independent, since

$$\deg u = 3k - 4 > k.$$



Hence  $\deg(um_{\beta_i} - vv_{\beta_i}) = 3k + i - 4$ ,  $i = 0, \dots, k$ ,  $\deg v = 2k - 2$ , and

$$3k + i - 4 \geq 3k - 4 > 2k - 2.$$

Thus  $\{v, um_{\beta_i} - vv_{\beta_i}, i = 0, \dots, k\}$  are linearly independent and  $\Sigma_{Q2} < \Sigma_{Q1}$ .

Statement a) is proved.

Consider the case where the parameters  $\mu$  and  $\sigma^2$  are unknown and  $\theta = (\beta^t, \mu, \sigma, \varphi)^t$ . As above, we assume that the true value is nonzero:  $\beta_k \neq 0$ .

Lemma 4.1 implies that the equation for the estimator of  $\mu$  is of the same form for both score functions:

$$(15) \quad \sum_{i=1}^n (x_i - \mu) = 0, \quad \hat{\mu}_{Q2} = \hat{\mu}_{Q1} = \bar{x}.$$

Thus the estimators of  $\mu$  coincide and the corresponding diagonal entry of the difference of the asymptotic covariance matrices is equal to zero. Since the difference  $\Sigma_{Q1} - \Sigma_{Q2}$  is a symmetric nonnegative definite matrix, the corresponding row and column are zero. Thus we can evaluate

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}\left(\Sigma_{Q1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q2}^{(\beta, \sigma, \varphi)}\right)$$

by using Corollary 3.2.

We derive from equalities (12) that

$$m_{\mu}u - v_{\mu}v = (1 - K)\beta^t D(m_{\beta}u - v_{\beta}v).$$

Then we exclude the vector  $(0, v^{-1}m_{\mu}u - v_{\mu})^t$  from the system of vectors:

$$(16) \quad \left[ \begin{pmatrix} 0 \\ v^{-1}um_{\beta_i} - v_{\beta_i} \end{pmatrix}, \begin{pmatrix} m_{\beta_i} \\ v_{\beta_i} \end{pmatrix}, i = 0, \dots, k, \begin{pmatrix} 0 \\ v^{-1}m_{\sigma}u - v_{\sigma} \end{pmatrix}, \begin{pmatrix} 0 \\ v \end{pmatrix}, \begin{pmatrix} u \\ w - v^2 \end{pmatrix} \right].$$

We consider the *quadratic* model (that is, we consider the case of  $k = 2$ ). Then  $u = f^t m_{\beta}$  and system (16) can be rewritten as follows:

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}[m_{\beta_i}u - v_{\beta_i}v, i = 0, 1, 2, m_{\sigma}u - v_{\sigma}v, v, pv] - 1,$$

where  $p = w - v^2 - f^t v_{\beta}$ . The degrees of the polynomials in the system are  $(2, 3, 4, 3, 2, 6)$ . Thus  $m_{\sigma}u + v_{\sigma}v$  is a linear combination of components of the vectors  $m_{\beta}u - v_{\beta}v$  and  $v = c_1u + c_2$ , where  $c_2$  is nonzero if  $\beta_2 \neq 0$ . After nondegenerate linear transformations, the degrees of the polynomials become equal to  $(2, 3, 4, 0, 6)$ . Therefore

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = 4,$$

and the rank of the difference of the matrices equals 6.

If  $k \geq 3$ , then the polynomials  $\{u, m_{\beta_i}, i = 0, \dots, k\}$  are linearly independent, since  $\deg u = 3k - 4 > k$  and  $\deg m_{\beta_i} = i \leq k$ . Then (16) implies that

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}[m_{\beta_i}u - v_{\beta_i}v, i = 0, \dots, k, m_{\sigma}u - v_{\sigma}v, v].$$

The polynomials

$$\{m_{\beta_i}u - v_{\beta_i}v, i = 0, \dots, k - 1, m_{\sigma}u - v_{\sigma}v, v\}$$

have different degrees,  $\deg(m_{\sigma}u - v_{\sigma}v) = \deg(m_{\beta_k}u - v_{\beta_k}v)$ , whence

$$d - 2 \leq \text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) \leq d - 1.$$

Indeed, the leading terms of the polynomials are

$$m_{\beta_i}u - v_{\beta_i}v = (3k - 3 - 2i)k^3 \beta_k^3 \tau^4 K^{3k+i-4} x^{3k+i-4} + \dots,$$

$$m_{\sigma}u - v_{\sigma}v = (k - 2)k^4 \beta_k^4 \tau^2 \frac{\partial \tau^2}{\partial \sigma} K^{4k-4} x^{4k-4} + \dots,$$

$$v = k^2 \beta_k^2 \tau^2 K^{2k-2} x^{2k-2} + \dots.$$

We see that  $3k - 3 - 2i > k - 1 > 0$  and  $2k - 2 < 3k + i - 4 < 4k - 4$  if  $0 \leq i \leq k - 1$ . Therefore the polynomials are of different degrees, indeed. The statement b) is proved.

#### APPENDIX C. PROOF OF THEOREM 5.1

We apply Corollary 3.1. The functions  $\{m_{\beta_0}, m_{\beta_1}, u\}$  are linearly independent and thus the formula for the rank is simpler:

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}\{um_{\beta_0} - vv_{\beta_0}, um_{\beta_1} - vv_{\beta_1}, vv_{\varphi}\}.$$

The functions  $um_{\beta_0} - vv_{\beta_0}$  and  $vv_{\varphi}$  are proportional to  $m^4$ , while the third function is expressed in terms of  $xm^4$  and  $m^4$ :

$$um_{\beta_1} - vv_{\beta_1} = cK(\varphi + 1)(c^2(2\varphi + 1) - 2c(\varphi + 1) + 1)x \cdot m^4 + \text{const} \cdot m^4.$$

If  $\beta_1 \neq 0$ , then  $c^2(2\varphi + 1) - 2c(\varphi + 1) + 1 > 0$ . The rank of this family is 2.

Let both parameters  $\mu$  and  $\sigma$  be unknown and  $\theta = (\beta_0, \beta_1, \mu, \sigma, \varphi)^t$ . Then

$$m_{\mu} = \beta_1(1 - K)m_{\beta_0}, \quad v_{\mu} = \beta_1(1 - K)v_{\beta_0}.$$

As in the proof of Theorem 4.1, we obtain equation (15) for the estimator of  $\mu$  for both score functions, whence  $\hat{\mu}_{Q2} = \hat{\mu}_{Q1} = \bar{x}$ . As above we make sure that

$$\text{rank}(\Sigma_{Q1} - \Sigma_{Q2}) = \text{rank}\left(\Sigma_{Q1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q2}^{(\beta, \sigma, \varphi)}\right).$$

Corollary 3.2 implies that

$$\text{rank}\left(\Sigma_{Q1}^{(\beta, \sigma, \varphi)} - \Sigma_{Q2}^{(\beta, \sigma, \varphi)}\right) = \text{rank}\{m_{\beta_1}u - v_{\beta_1}v, m_{\sigma}u - v_{\sigma}v, v_{\varphi}v\}.$$

If  $\beta_1 \neq 0$ , the rank of the latter family is 2, since

$$m_{\beta_1}u - v_{\beta_1}v = c_1xm^4 + c_2m^4, \quad m_{\sigma}u - v_{\sigma}v = c_3xm^4 + c_4m^4, \quad v_{\varphi}v = c_5m^4.$$

#### BIBLIOGRAPHY

1. R. J. Carroll, D. Ruppert, and L. A. Stefanski, *Measurement Error in Nonlinear Models*, Chapman and Hall, London, 1995. MR1630517 (2000c:62001)
2. A. Kukush and H. Schneeweiss, *Comparing different estimators in a nonlinear measurement error model. I*, Math. Methods Statist. **14** (2005), 53–79. MR2158071 (2006j:62068a)
3. A. Kukush and H. Schneeweiss, *Asymptotic optimality of the quasi-score estimator in a class of linear score estimators*, Discussion Paper, vol. 477, Universität München, SFB 386, 2006.
4. A. Kukush, A. Malenko, and H. Schneeweiss, *Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models*, Discussion Paper, vol. 494, Universität München, SFB 386, 2006.
5. M. J. Schervish, *Theory of Statistics*, Springer, New York, 1995. MR1354146 (96m:62001)
6. S. Shklyar, H. Schneeweiss, and A. Kukush, *Quasi Score is more efficient than Corrected Score in a polynomial measurement error model*, Metrika **65** (2007), no. 3, 275–295. MR2299552 (2008h:62071)
7. H. Schneeweiss, *The polynomial and the Poisson measurement error models: Some further results on quasi score and corrected score estimation*, Discussion Paper, vol. 446, Universität München, SFB 386, 2005.

DEPARTMENT OF PROBABILITY THEORY AND MATHEMATICAL STATISTICS, FACULTY FOR MECHANICS AND MATHEMATICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE 6, KYIV 03127, UKRAINE

*E-mail address:* `expilis@yandex.ru`

DEPARTMENT OF MATHEMATICAL ANALYSIS, FACULTY FOR MECHANICS AND MATHEMATICS, NATIONAL TARAS SHEVCHENKO UNIVERSITY, ACADEMICIAN GLUSHKOV AVENUE 6, KYIV 03127, UKRAINE

*E-mail address:* `alexander_kukush@univ.kiev.ua`

Received 28/DEC/2006

Translated by N. SEMENOV