Теорія Ймовір. та Матем. Статист.
Вип. 92, 2015

Theor. Probability and Math. Statist.
No. 92, 2016, Pages 81–91
http://dx.doi.org/10.1090/tpms/984
Article electronically published on August 10, 2016

CONVERGENCE OF ESTIMATORS IN THE POLYNOMIAL MEASUREMENT ERROR MODEL

UDC 519.21

A. G. KUKUSH AND YA. V. TSAREGORODTSEV

Abstract. A polynomial measurement error model is considered. The variance of errors in the regressor variable and the covariance between errors in the regressor variable and errors of the response variable are assumed to be known. The adjusted least squares estimator of regression parameters adopts the ordinary least squares estimator to the errors presented in the regressor. Conditions for the strong consistency of the estimator are found. These conditions are weaker as compared to those by Cheng and Schneeweiss (1998) [Journal of the Royal Statistical Society B, no. 1, 189–199]. Sufficient conditions for the asymptotic normality of the estimator are also found.

1. Introduction

The polynomial relation between the variables is given by

$$y_i = \beta_0 + \beta_1\xi_i + \beta_2\xi_i^2 + \cdots + \beta_k\xi_i^k + \varepsilon_i,$$
$$x_i = \xi_i + \delta_i, \qquad i = 1, \dots, n,$$

where the errors $(\delta_i, \varepsilon_i)$, $i \geq 1$, are independent identically distributed pairs of random variables with zero means and the covariance matrix

$$\Omega := \begin{pmatrix} \sigma_\delta^2 & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix}. \tag{1}$$

Here ξ_i, $i \geq 1$, are unobservable nonrandom variables called *hidden variables.* This setting corresponds to the so-called *functional* measurement error model. In the case of the *structural* model, $\{\xi_i\}$ is a sequence of independent identically distributed random variables.

The parameters σ_δ^2 and $\sigma_{\delta\varepsilon}$ are assumed to be known. Given the observations (x_i, y_i), $i = 1, \dots, n$, the problem is to estimate the regression parameters $\beta_0, \dots, \beta_k$ and, possibly, the variance σ_ε^2.

Cheng and Schneeweiss [4] construct an adjusted least squares estimator of the regression parameters. A survey of the literature on the polynomial regression model can also be found in [4]. Cheng, Schneeweiss, and Thamerus [5] observe that the adjusted least squares estimator is unstable if the size of a sample is small or moderate. A modified adjusted least squares estimator whose behavior is nicer for small and moderate samples and that is asymptotically equivalent to the adjusted least squares estimator as $n \to \infty$ is constructed in [5]. Therefore the modified adjusted least squares estimator better fits practical needs, especially if the size of a sample is not large.

2010 *Mathematics Subject Classification.* Primary 62F12, 62J02.

Key words and phrases. Asymptotic normality, adjusted least squares estimator, consistency of estimators, measurement error model, modification of estimators for small samples, polynomial regression.

Cheng and Kukush [6] used the adjusted least squares estimator to construct a goodness of fit criterion for the polynomial functional measurement error model. Since the adjusted least squares estimator is useful for the structural model, the criterion mentioned above can be used in the structural model as well. Hall and Ma [7] constructed a more powerful goodness of fit criterion for the structural polynomial model.

The aim of this paper is to weaken the conditions used in [4] to prove the consistency of the adjusted least squares estimator of the regression parameter and to find sufficient conditions for the asymptotic normality of this estimator (no condition for the asymptotic normality is mentioned in [4] at all). We do not require the convergence of certain sampling moments of the hidden variable ξ to prove the consistency. Instead, we require that the sampling moment of order $2k$ is bounded and that the determinant of a certain matrix is separated from zero. This matrix is an analogue of the information matrix for a finite sample. We require the convergence of sampling moments of all orders up to $4k-2$ to prove the asymptotic normality. Naturally, the modified adjusted least squares estimator inherits the asymptotic properties of the adjusted least squares estimator. Similar results for the adjusted least squares estimator are obtained in [11, 12] for the linear vector model.

The paper is organized as follows. We describe the model for observations and introduce the adjusted least squares estimators $\hat\beta_A$ and $\hat\sigma^2_{\varepsilon,A}$ in Section 2. Section 3 contains results on the strong consistency of adjusted least squares estimators. The asymptotic normality of $\hat\beta_A$ is proved in Section 4. Also, a consistent estimator of the asymptotic covariance matrix of $\hat\beta_A$ is constructed in Section 4. Section 5 describes some possible directions for further researches.

The following notation is used throughout the paper. The symbol E denotes the mathematical expectation, while cov stands for the covariance matrix of a random vector. The bar means an averaging; for example,

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \qquad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n}x_i y_i,$$

etc. The superscript T denotes the transposition; all vectors in the paper are columns. The convergence with probability one and in distribution are denoted by $\xrightarrow{\mathsf{P1}}$ and $\xrightarrow{\mathrm{d}}$, respectively. Every sequence of random variables or random matrices that converges to zero with probability one is denoted by $o(1)$. The expression $\varepsilon \stackrel{\mathrm{d}}{=} \varepsilon_1$ means that the random variables ε and ε_1 have the same probability distribution. Positive constants that do not depend on n, the size of a sample, are denoted by const; in particular, equalities such as $2\cdot\mathrm{const} = \mathrm{const}$ may happen in the text below.

2. A model for observations and the adjusted least squares estimator

Let $k \ge 1$ be fixed and consider the following model for observations:

$$y_i = \rho_i^{\mathsf{T}}\beta + \varepsilon_i, \qquad x_i = \xi_i + \delta_i, \qquad i = 1,\dots,n, \tag{2}$$

where $\rho_i := (1, \xi_i, \xi_i^2, \dots, \xi_i^k)^{\mathsf{T}}$ and $\beta := (\beta_0, \beta_1, \dots, \beta_k)^{\mathsf{T}}$; ξ_i, $i = 1,\dots,n$, are nonrandom values of the hidden variable ξ. The errors $(\delta_i, \varepsilon_i)$, $i = 1,\dots,n$, form a sequence of independent identically distributed random pairs with zero means and the covariance matrix Ω defined in (1), and with positive variances σ_δ^2 and σ_ε^2.

All moments $\mathsf{E}\,\delta^j$, $j = 2,\dots,2k$, and $\mathsf{E}\,\delta^j\varepsilon$, $j = 1,\dots,k$, are assumed to be known. Following [4] we construct an adjusted least squares estimator of the parameter β.

If (ξ_i, y_i), $i = 1, \dots, n$, are observed, then the usual least squares estimator is constructed with the help of the estimating function

$$S_{LS}^{(\beta)}(\xi, y; \beta) := \rho y - \rho\rho^{\mathsf{T}}\beta, \qquad \rho = \rho(\xi) := \left(1, \xi, \dots, \xi^k\right)^{\mathsf{T}}.$$

This is an unbiased estimating function, that is,

$$\mathsf{E}_\beta\, S_{LS}^{(\beta)}(\xi, y; \beta) = 0 \tag{3}$$

for all $\beta \in \mathbb{R}^{k+1}$. Here and throughout the paper the symbol E_β denotes the mathematical expectation if β is the true value of the regression parameter, that is $y = \rho^{\mathsf{T}}(\xi)\beta + \varepsilon$, $\varepsilon \overset{\mathrm{d}}{=} \varepsilon_1$; ξ is a nonrandom variable.

We observe $x = \xi + \delta$, $(\delta, \varepsilon) \overset{\mathrm{d}}{=} (\delta_1, \varepsilon_1)$, instead of ξ . Then we construct an adjusted estimating function $S_C^{(\beta)}(x, y; \beta)$ such that

$$\mathsf{E}_\beta\, S_C^{(\beta)}(x, y; \beta) = \mathsf{E}_\beta\, S_{LS}^{(\beta)}(\xi, y; \beta), \qquad \beta \in \mathbb{R}^{k+1}.$$

The new estimating function is unbiased in view of equality (3). This property is a necessary condition for the consistency of the adjusted least squares estimator that is determined by the estimating function $S_C^{(\beta)}$.

To construct $S_C^{(\beta)}$, one needs to construct a matrix $H(x)$ and vector $h(x, y)$ such that

$$\mathsf{E}\, H(x) = \rho\rho^{\mathsf{T}}, \tag{4}$$

$$\mathsf{E}_\beta\, h(x, y) = \mathsf{E}_\beta\, \rho y; \qquad \xi \in \mathbb{R},\ \beta \in \mathbb{R}^{k+1}. \tag{5}$$

Then the adjusted estimating function is equal to

$$S_C^{(\beta)} = h(x, y) - H(x)\beta. \tag{6}$$

To construct $H(x)$, consider the polynomials

$$t_r = \sum_{j=0}^{r} a_{rj} x^j \tag{7}$$

for every $0 \le r \le 2k$, where a_{rj} are functions of $\mathsf{E}\,\delta^p$, $p = 0, \dots, r$. The polynomials defined by (7) are such that

$$\mathsf{E}\, t_r(x) = \xi^r, \qquad \xi \in \mathbb{R}.$$

For example, $t_0 = 1$, $t_1 = x$, $t_2 = x_2 - \sigma_\delta^2$, and $t_3 = x^3 - 3x\sigma_\delta^2 - \mathsf{E}\,\delta^3$. The details of the construction of the polynomials t_r can be found in [4]. Then the solution of equation (4) is determined by the matrix H with entries

$$H_{ij}(x) = t_{i+j}(x), \qquad i, j = 0, \dots, k. \tag{8}$$

Recall that the moments $\mathsf{E}\,\delta^p$, $p = 0, \dots, 2k$, are known.

Further, a solution of equation (5) is written as follows:

$$h = (h_r)_{r=0}^k, \qquad h_r = t_r y - \sum_{j=0}^{r} b_{rj} t_j. \tag{9}$$

Here b_{rj} are functions of $\mathsf{E}\,\delta^p$, $p = 0, \dots, r$, and $\mathsf{E}\,\delta^p\varepsilon$, $p = 0, \dots, r$. In particular, $h_0 = y$, $h_1 = xy - \sigma_{\delta\varepsilon}$, and $h_2 = (x^2 - \sigma_\delta^2)y - \mathsf{E}\,\delta^2\varepsilon - 2\sigma_{\delta\varepsilon}x$. Details of the construction of the solutions h_r can be found in [4].

After the estimating function (6) is constructed, the adjusted least squares estimator $\hat\beta_A$ is found from the equation

$$\frac{1}{n}\sum_{i=1}^{n} S_C^{(\beta)}(x_i, y_i; \beta) = 0$$

or

$$\overline{H}\beta = \overline{h}, \qquad \beta \in \mathbb{R}^{k+1}. \tag{10}$$

Here $\overline{H}$ and $\overline{h}$ are averaged matrices $H_{(i)} := H(x_i)$ and vectors $h_{(i)} := h(x_i, y_i)$.

Definition 1. The estimator $\hat{\beta}_A$ is given by

$$\hat{\beta}_A = \overline{H}^{-1}\overline{h} \tag{11}$$

if the matrix $\overline{H}$ is nondegenerate. Otherwise, we put $\hat{\beta}_A = 0$ if the matrix $\overline{H}$ is degenerate.

We also construct an estimator for the variance σ_ε^2. If (ξ_i, y_i), $i = 1, \dots, n$, are observed, then the following estimating function is unbiased:

$$S_{LS}^{(\sigma_\varepsilon^2)}\left(\xi, y; \beta, \sigma_\varepsilon^2\right) := y^2 - y\rho^{\mathsf{T}}(\xi)\beta - \sigma_\varepsilon^2.$$

That is,

$$\mathsf{E}_{\beta,\sigma_\varepsilon^2} S_{LS}^{(\sigma_\varepsilon^2)}\left(\xi, y; \beta, \sigma_\varepsilon^2\right) = 0, \qquad \beta \in \mathbb{R}^{k+1},\ \sigma_\varepsilon^2 > 0.$$

The adjusted estimating function is given by

$$S_C^{(\sigma_\varepsilon^2)}\left(x, y; \beta, \sigma_\varepsilon^2\right) = y^2 - h^{\mathsf{T}}(x, y)\beta - \sigma_\varepsilon^2.$$

In view of equality (5), the latter estimator is unbiased, since

$$\mathsf{E}_{\beta,\sigma_\varepsilon^2} S_C^{(\sigma_\varepsilon^2)}\left(x, y; \beta, \sigma_\varepsilon^2\right) = \mathsf{E}_{\beta,\sigma_\varepsilon^2} S_{LS}^{(\sigma_\varepsilon^2)}\left(\xi, y; \beta, \sigma_\varepsilon^2\right).$$

To estimate σ_ε^2, we use the equation

$$\frac{1}{n}\sum_{i=1}^{n} S_C^{(\sigma_\varepsilon^2)}\left(x_i, y_i; \beta, \sigma_\varepsilon^2\right) = 0,$$

$$\sigma_\varepsilon^2 = \overline{y^2} - \overline{h}^{\mathsf{T}}\beta,$$

in addition to (10).

Definition 2. The estimator $\hat{\sigma}_{\varepsilon,A}^2$ is given by

$$\hat{\sigma}_{\varepsilon,A}^2 = \overline{y^2} - \overline{h}^{\mathsf{T}}\hat{\beta}_A. \tag{12}$$

3. The consistency of the adjusted least squares estimator

Lemma 3. *Let $r > 1$ be a fixed real number and let $\{\eta_k, k \ge 1\}$ be a sequence of independent identically distributed random variables with zero mean and such that $\mathsf{E}\,|\eta_1|^r < \infty$. Let a sequence $\{a_k, k \ge 1\}$ be such that*

$$\overline{|a|^r} = \frac{1}{n}\sum_{k=1}^{n} |a_k|^r \le \text{const}.$$

Then

$$\overline{a\eta} = \frac{1}{n}\sum_{k=1}^{n} a_k\eta_k \xrightarrow{\mathsf{P1}} 0.$$

Proof. Without loss of generality one may assume that $1 < r \le 2$. Given random variables $X_n = a_n\eta_n$, $n \ge 1$, consider the series

$$\sum_{n=1}^{\infty} \frac{\mathsf{E}\,|X_n|^r}{n^r} = \mathsf{E}\,|\eta_1|^r \sum_{n=1}^{\infty} \frac{|a_n|^r}{n^r}. \tag{13}$$

Put $A_0 = 0$, $A_n = \frac{1}{n}\sum_{k=1}^n |a_k|^r$, $n \ge 1$. Then $|a_n|^r = nA_n - (n-1)A_{n-1}$, $n \ge 1$. The series on the right hand side of (13) converges since

$$\sum_{n=1}^{\infty} \frac{nA_n - (n-1)A_{n-1}}{n^r} = \sum_{m=1}^{\infty} A_m \, \frac{(m+1)^r - m^r}{m^{r-1}(m+1)^r},$$

and the latter series converges as $\{A_m\}$ is bounded. Since $1 < r \le 2$ and series (13) converges, Theorem 12 of [3, Chapter VI] implies

$$\overline{a\eta} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathsf{P1}} 0. \qquad \square$$

Now we introduce a notion that is useful in the study of the strong consistency of statistical estimators.

Definition 4. Let $\{A_n(\omega), n \ge 1\}$ be a sequence of statements that depend on an elementary random event $\omega \in \Omega$. We say that statements $A_n = A_n(\omega)$ *eventually* hold if there exists a random event Ω_0, $\mathsf{P}(\Omega_0) = 1$, such that, for every $\omega \in \Omega_0$, there exists a positive integer number $n_0(\omega)$ for which the statements $A_n(\omega)$ hold for all $n \ge n_0(\omega)$.

The following conditions are needed to prove the consistency of the estimator $\hat\beta_A$:

(i) $\overline{\xi^{2k}} = \frac{1}{n}\sum_{i=1}^n \xi_i^{2k} \le \text{const}$.
(i′) $\mathsf{E}\,|\delta^{k-1}\varepsilon|^r < \infty$ for some $r > 1$.
(ii) $\underline{\lim}_{n\to\infty} \det(\overline{\rho\rho^\mathsf{T}}) > 0$.

Theorem 5. *Let conditions (i), (i′), and (ii) hold. Then the estimator $\hat\beta_A$ is strongly consistent, that is,*

$$\hat\beta_A \xrightarrow{\mathsf{P1}} \beta, \qquad n \to \infty.$$

Proof. a) First we show that the matrix $\overline{H}$ is eventually nondegenerate (in the sense of Definition 4).

According to equalities (8), the matrix entries are given by $\overline{H}_{ij} = \overline{t}_{i+j}$. The difference $\overline{t}_{i+j} - \mathsf{E}\,\overline{t}_{i+j}$ is a linear combination of terms $z_p := \overline{x^p} - \mathsf{E}\,\overline{x^p}$, $0 \le p \le 2k$. Applying the binomial theorem to $x^p = (\xi+\delta)^p$ we prove that z_p is a linear combination of terms $g_{uv} := \overline{\xi^u\delta^v} - \mathsf{E}\,\overline{\xi^u\delta^v}$, $u, v \ge 0$, $u + v \le p$. If $u = 0$, then the strong law of large numbers implies that $\overline{\delta^v} - \mathsf{E}\,\overline{\delta^v} \xrightarrow{\mathsf{P1}} 0$, since $v \le 2k$ and $\mathsf{E}\,\delta^{2k} < \infty$. If $v = 0$, then $g_{u0} = 0$. Hence

$$g_{uv} = \overline{\xi^u(\delta^v - \mathsf{E}\,\delta^v)} \tag{14}$$

for $u, v \ge 1$, $u + v \le 2k$. Next we use Lemma 3:

$$\overline{\xi^{ru}} \le \text{const}$$

for $r = \frac{2k}{2k-1}$ and

$$\mathsf{E}\,|\delta^v - \mathsf{E}\,\delta^v|^r \le \text{const}\,\mathsf{E}\,|\delta|^{vr} \le \text{const}\left(1 + \mathsf{E}\,|\delta|^{(2k-1)r}\right) < \infty$$

for the same r. Hence Lemma 3 yields the convergence $g_{uv} \xrightarrow{\mathsf{P1}} 0$, $n \to \infty$. Therefore

$$\overline{H} - \mathsf{E}\,\overline{H} = \overline{H} - \overline{\rho\rho^\mathsf{T}} \xrightarrow{\mathsf{P1}} 0, \qquad n \to \infty. \tag{15}$$

Consider the matrix $\Phi_n := \overline{\rho\rho^\mathsf{T}}$. This matrix is nondegenerate for $n \ge n_0$ in view of condition (ii). Moreover, its entries $\overline{\xi^{i+j}}$, $0 \le i, j \le k$, are bounded by condition (i). We represent the inverse matrix Φ_n^{-1} in terms of the determinant and adjoint matrix. Thus the Euclidean matrix norm for $n \ge n_0$ is such that

$$\left\|\Phi_n^{-1}\right\| \le \frac{\text{const}}{\det(\Phi_n)}, \qquad \varlimsup_{n\to\infty} \left\|\Phi_n^{-1}\right\| \le \frac{\text{const}}{\varliminf_{n\to\infty}\det(\Phi_n)} < \infty,$$

$$\left\|\Phi_n^{-1}\right\| \le \text{const}\,. \tag{16}$$

Now convergence (15) and inequality (16) show that eventually $\left\|\overline{H} - \Phi_n\right\| < \left\|\Phi_n^{-1}\right\|^{-1}$. Note that if the latter inequality holds, then the matrix $\overline{H}$ is nondegenerate by the inverse operator perturbation theorem. This proves that $\overline{H}$ is eventually nondegenerate and therefore equality (11) eventually holds for $\hat\beta_A$.

b) The behavior of $\overline{h}$. Equalities (9) for $0 \le r \le k$ imply that

$$h_r = t_r \rho^{\mathsf T}\beta + t_r\varepsilon - \sum_{j=0}^{r} b_{rj} t_j. \tag{17}$$

We are going to check

$$\overline{h}_r - \mathsf{E}_\beta\,\overline{h}_r \xrightarrow{\mathsf{P1}} 0.$$

The further proof is similar to that of part a). Now the term

$$\overline{t_r\varepsilon} - \mathsf{E}\,\overline{t_r\varepsilon}, \qquad 0 \le r \le k,$$

appears, and only this makes a difference. This term is a linear combination of

$$f_{ij} := \overline{\xi^i\delta^j\varepsilon} - \mathsf{E}\,\overline{\xi^i\delta^j\varepsilon}, \qquad i+j \le k. \tag{18}$$

The strong law of large numbers implies for $i = 0$ that $f_{0j} \xrightarrow{\mathsf{P1}} 0$. Now let $i \ge 1,\ j \le k-i$. We have $\overline{\xi^{2i}} \le \text{const}$, and condition (i′) implies

$$\mathsf{E}\left|\delta^j\varepsilon - \mathsf{E}\,\delta^j\varepsilon\right|^r \le \text{const}\cdot\mathsf{E}\left|\delta^j\varepsilon\right|^r < \infty$$

for $r > 1$. Then $f_{ij} \xrightarrow{\mathsf{P1}} 0$ by Lemma 3. Therefore

$$\overline{h} - \mathsf{E}_\beta\,\overline{h} \xrightarrow{\mathsf{P1}} 0. \tag{19}$$

Considering (5) we conclude that

$$h = \overline{\rho\rho^{\mathsf T}}\beta + o(1) = \Phi_n\beta + o(1).$$

c) We derive from equality (11) that eventually

$$\hat\beta_A = (\Phi_n + o(1))^{-1}(\Phi_n\beta + o(1)) = \left(I + \Phi_n^{-1}o(1)\right)^{-1}\left(\beta + \Phi_n^{-1}o(1)\right). \tag{20}$$

Here I denotes the unit matrix. By inequality (16),

$$\left\|\Phi_n^{-1}o(1)\right\| \le \left\|\Phi_n^{-1}\right\|\cdot\|o(1)\| \le \text{const}\cdot o(1) = o(1). \tag{21}$$

Finally, relations (20) and (21) imply the convergence $\hat\beta_A \xrightarrow{\mathsf{P1}} \beta$. □

Consider one more condition.

(iii) $\mathsf{E}\,|\varepsilon|^{2+c} < \infty$ for some $c > 0$.

Theorem 6. *Let conditions* (i), (i′), (ii), *and* (iii) *hold. Then both estimators* $\hat\beta_A$ *and* $\hat\sigma^2_{\varepsilon,A}$ *are strongly consistent. In particular,*

$$\hat\sigma^2_{\varepsilon,A} \xrightarrow{\mathsf{P1}} \sigma^2_\varepsilon, \qquad n\to\infty.$$

Proof. It only remains to prove that the estimator of the variance is strongly consistent. We have

$$\overline{y^2} = \beta^\top \overline{\rho\rho^\top}\beta + 2\beta^\top \overline{\rho\varepsilon} + \overline{\varepsilon^2}.$$

By Lemma 3, $\overline{\rho\varepsilon} \xrightarrow{\mathsf{P1}} 0$, since condition (iii) holds. Moreover, $\overline{\varepsilon^2} \xrightarrow{\mathsf{P1}} \sigma_\varepsilon^2$. Then $\overline{y^2} = \mathsf{E}\,\overline{y^2} + o(1)$, and equality (12) yields the representation

$$\hat\sigma^2_{\varepsilon,A} = \mathsf{E}\,\overline{y^2} + o(1) - \left(\beta^\top \overline{\rho\rho^\top} + o(1)\right)(\beta + o(1)) = \sigma^2_\varepsilon + o(1). \qquad \square$$

Remark 7. Theorem 6 essentially weakens the sufficient conditions for the consistency used in [4]. First, Theorem 6 uses lower moments of errors and, second, it requires that the sampling moment $\overline{\xi^{2k}}$ is bounded and that $\det \Phi_n$ is asymptotically separated from zero instead of the assumption about the existence of a nondegenerate limit of the matrices Φ_n as in [4]. Note that we do not assume the convergence of sampling moments in Theorem 6.

4. Asymptotic normality of $\hat\beta_A$

The following central limit theorem is a special case of Lyapunov's theorem [2, p. 73].

Theorem 8. *Let $\{z_i, i \ge 1\}$ be a sequence of independent random vectors in $\mathbb{R}^d$ with zero means and such that*

$$\overline{\operatorname{cov}(z)} = \frac{1}{n}\sum_{i=1}^n \operatorname{cov}(z_i) \to S, \qquad n \to \infty.$$

Assume that there exists $c > 0$ such that

$$\frac{1}{n}\sum_{i=1}^n \mathsf{E}\,\|z_i\|^{2+c} \le \text{const}\,. \tag{22}$$

Then

$$\sqrt{n}\,\overline{z} = \frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \xrightarrow{\mathrm{d}} N(0, S).$$

The following conditions are needed to prove the asymptotic normality of the estimator.

(iv) For every $r = 0, \dots, 4k-2$, there exists a finite limit

$$\mu_r := \lim_{n\to\infty} \overline{\xi^r}. \tag{23}$$

Moreover, the matrix $H_\infty := (\mu_{i+j})_{i,j=0}^k$ is nondegenerate.

(v) $\mathsf{E}\,\delta^{2k}\varepsilon^2 < \infty$ and

$$\overline{|\xi|^{4k-2+c}} \le \text{const}, \qquad \mathsf{E}\,|\delta|^{4k-2+c} < \infty,$$
$$\mathsf{E}\,|\varepsilon|^{2+c} < \infty, \qquad \mathsf{E}\,|\delta^{k-1}\varepsilon|^{2+c} < \infty$$

for some $c > 0$.

Theorem 9. *Let conditions* (iv)–(v) *hold. Then*

$$\sqrt{n}(\hat\beta_A - \beta) \xrightarrow{\mathrm{d}} N(0, \Sigma), \tag{24}$$

$$\Sigma = H_\infty^{-1} U H_\infty^{-1}, \qquad U = \lim_{n\to\infty} \overline{\operatorname{cov}(h - H\beta)} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \operatorname{cov}\left(h_{(i)} - H_{(i)}\beta\right),$$

where $h_{(i)} = h(x_i, y_i)$ and $H_{(i)} = H(x_i)$.

Proof. Assumptions of Theorem 5 are satisfied, and thus

$$\overline{H} = \mathsf{E}\,\overline{H} + o(1) = \overline{\rho\rho^{\mathsf{T}}} + o(1).$$

Convergence (23) implies $\overline{H} \xrightarrow{\mathsf{P1}} H_\infty$. Note that the latter matrix is positive definite. Hence $\overline{H}$ is also eventually positive definite (see Definition 4). Hence eventually

$$\sqrt{n}(\hat{\beta}_A - \beta) = \overline{H}^{-1} \cdot \sqrt{n} \cdot \overline{h - H\beta}. \tag{25}$$

Now we apply Theorem 8 to the independent random vectors $z_i := h_{(i)} - H_{(i)}\beta$. By relation (17),

$$z := h - H\beta = (t\rho^{\mathsf{T}} - H)\beta - Bt + t\varepsilon, \tag{26}$$

where the entries of the matrix B are equal to

$$B_{rj} = \begin{cases} b_{rj}, & 0 \le j \le r \le k; \\ 0, & j > r,\ 0 \le r \le k-1. \end{cases}$$

By construction of the functions h and H, we have $\mathsf{E}\,z = 0$. Then the matrix

$$U := \lim_{n\to\infty} \overline{\operatorname{cov}(z)} = \lim_{n\to\infty} \mathsf{E}\,\overline{zz^{\mathsf{T}}}$$

is well defined by condition (iv). For example, μ_{4k-2} appears in

$$\lim_{n\to\infty} \mathsf{E}\,\overline{(t\rho^{\mathsf{T}} - H)\beta\beta^{\mathsf{T}}(t\rho^{\mathsf{T}} - H)^{\mathsf{T}}}.$$

It is important that $t_k\rho_k - H_{kk} = t_k\rho_k - t_{2k}$ as a polynomial of ξ involves the term ξ^{2k-1} and does not contain ξ^{2k}. Moreover, the term $t\varepsilon$ contains $\delta^k\varepsilon$, whose variance is finite in view of the first condition in (v).

Lyapunov's condition (22) holds by (v). For example, the term $t_k\rho_k - H_{kk}$ contains $\text{const}\cdot\xi^{2k-1}\delta$. Consider

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left|\xi_i^{2k-1}\delta_i\right|^r = \mathsf{E}\,|\delta|^r \cdot \overline{|\xi|^{(2k-1)r}} < \infty$$

for $r > 2$, where $r = 2 + \frac{c}{2k-1}$ and c is a constant in condition (v). Reasoning as above, we prove that Lyapunov's condition holds for other components of the vector z.

Therefore Theorem 8 implies

$$\sqrt{n}\cdot\overline{h - H\beta} \xrightarrow{\mathrm{d}} N(0, U),$$

whence we conclude that

$$\sqrt{n}(\hat{\beta}_A - \beta) \xrightarrow{\mathrm{d}} N\left(0, H_\infty^{-1} U H_\infty^{-1}\right)$$

by the Slutskiĭ lemma [1] and equality (25). □

Theorem 10. *Let condition* (*iv*) *hold. Assume in addition that*

(vi) $\mathsf{E}\,\delta^{4k} < \infty$, $\mathsf{E}\,\delta^{2k}\varepsilon^2 < \infty$, *and*

$$\mathsf{E}\,|\varepsilon|^{2+c} < \infty, \qquad \mathsf{E}\left|\delta^{2k-1}\varepsilon^2\right|^{1+c} < \infty, \qquad \mathsf{E}\left|\delta^{3k-1}\varepsilon\right|^{1+c} < \infty$$

for some $c > 0$;

(vii) $\overline{|\xi|^{4k}} \le \text{const}$.

Then $\hat{\Sigma}$ *is a strongly consistent estimator of the matrix* Σ *defined in* (24),

$$\hat{\Sigma} = H^{-1}\hat{U}H^{-1}, \qquad \hat{U} = \frac{1}{n}\sum_{i=1}^{n}\left(h_{(i)} - H_{(i)}\hat{\beta}_A\right)\left(h_{(i)} - H_{(i)}\hat{\beta}_A\right)^{\mathsf{T}},$$

that is, $\hat{\Sigma} \xrightarrow{\mathsf{P1}} \Sigma$ *as* $n \to \infty$.

Proof. All the assumptions of Theorem 9 are satisfied, and thus convergence (24) holds. In particular, the estimator $\hat{\beta}_A$ is strongly consistent.

It is seen from the proof of Theorem 9 that $\overline{H} \xrightarrow{\mathsf{P1}} H_\infty$. It remains to check that

$$\hat{U} \xrightarrow{\mathsf{P1}} U.$$

Put $\Delta\hat{\beta} = \hat{\beta}_A - \beta$. Then $\Delta\hat{\beta} \xrightarrow{\mathsf{P1}} 0$ and

$$\hat{U} = \overline{(h - H\beta)(h - H\beta)^\mathsf{T}} + \text{rest},$$

$$\text{rest} = \overline{H\Delta\hat{\beta} \cdot (\Delta\hat{\beta})^\mathsf{T} H} - \overline{(h - H\beta)(H\Delta\hat{\beta})^\mathsf{T}} - \overline{H\Delta\hat{\beta}(h - H\beta)^\mathsf{T}}.$$

Since $\hat{\beta}_A$ is strongly consistent, rest $\xrightarrow{\mathsf{P1}} 0$. Indeed, the sequence of averages $\{\overline{x^{4k}}, n \geq 1\}$ is bounded almost surely in view of conditions (vi) and (vii), whence we derive that the average $\overline{\|H\|^2}$ remains bounded almost surely as the size of a sample grows. Moreover, the first term of the remainder term "rest" tends to zero almost surely. The almost sure convergence of the second and third terms follows from the Cauchy–Schwarz inequality, since (as shown below) the sequence of averages

$$\overline{\|h - H\beta\|^2} \tag{27}$$

is bounded almost surely. It remains to check that

$$\overline{(h - H\beta)(h - H\beta)^\mathsf{T}} \xrightarrow{\mathsf{P1}} U = \lim_{n\to\infty} \overline{\operatorname{cov}(h - H\beta)} \tag{28}$$

or

$$\overline{(h - H\beta)(h - H\beta)^\mathsf{T}} - \mathsf{E}\,\overline{(h - H\beta)(h - H\beta)^\mathsf{T}} \xrightarrow{\mathsf{P1}} 0. \tag{29}$$

In particular, convergence (28) implies that the averages (27) are bounded almost surely. Expansion (26) holds for $z = h - H\beta$. The terms z involve expressions $\xi^i\delta^j$ for $i + j \leq 2k$, $i \leq 2k - 1$, and $\xi^u\delta^v\varepsilon$ for $u + v \leq k$. When considering zz^T we have to consider some additional terms as explained below.

a) The terms $\xi^i\delta^j$ for $i + j \leq 4k$, $i \leq 4k - 2$. We need to prove the convergence

$$\overline{\xi^i\delta^j} - \mathsf{E}\,\overline{\xi^i\delta^j} \xrightarrow{\mathsf{P1}} 0. \tag{30}$$

Convergence (30) holds for $i = 0$ and $j = 4k$, since $\mathsf{E}\,\delta^{4k} < \infty$ by condition (vi). Now let $j \leq 4k - 1$. We have $\overline{|\xi|^{ri}} \leq$ const for some $r > 1$ by condition (vii). In addition, $\mathsf{E}\,|\delta^j|^q < \infty$, $q = \frac{4k}{j} > 1$, and thus convergence (30) holds by Lemma 3.

b) The terms $\xi^i\delta^j\varepsilon^2$ for $i + j \leq 2k$. We need to prove the convergence

$$\overline{\xi^i\delta^j\varepsilon^2} - \mathsf{E}\,\overline{\xi^i\delta^j\varepsilon^2} \xrightarrow{\mathsf{P1}} 0. \tag{31}$$

We have $\overline{\delta^j\varepsilon^2} - \mathsf{E}\,\delta^j\varepsilon^2 \xrightarrow{\mathsf{P1}} 0$ for $i = 0$ and $j \leq 2k$, since $\mathsf{E}\,|\delta^j\varepsilon^2| < \infty$ by condition (vi). Moreover, $\overline{|\xi|^{ri}} \leq$ const for some $r > 1$ if $1 \leq i \leq 2k$ and $j \leq 2k - 1$ in view of condition (vii). In addition, $\mathsf{E}\,|\delta^j\varepsilon^2|^{1+c} < \infty$, where the number c appears in condition (vi). Convergence (31) holds by Lemma 3.

c) The terms $\xi^i\delta^j\varepsilon$ for $i \leq 3k - 1$, $i + j \leq 3k$. We need to prove the convergence

$$\overline{\xi^i\delta^j\varepsilon} - \mathsf{E}\,\overline{\xi^i\delta^j\varepsilon} \xrightarrow{\mathsf{P1}} 0. \tag{32}$$

We have $\overline{\delta^j\varepsilon} - \mathsf{E}\,\delta^j\varepsilon \xrightarrow{\mathsf{P1}} 0$ for $i = 0$ and $j \leq 3k$, since $\mathsf{E}\,|\delta^j\varepsilon| < \infty$ by condition (vi). Moreover, $\overline{|\xi|^{ri}} \leq$ const for some $r > 1$ by condition (vii) if $1 \leq i \leq 3k-1$ and $j \leq 3k-1$. Here $\mathsf{E}\,|\delta^j\varepsilon|^{1+c} < \infty$, where the number c appears in condition (vi). Convergence (32) holds by Lemma 3.

Cases a)–c) considered above prove convergence (29), and this completes the proof of Theorem 10. ☐

Note that modified estimators $\hat{\beta}_M$ and $\hat{\sigma}^2_{\varepsilon,M}$ are constructed in the paper [5] for the model of polynomial regression under consideration. These estimators are more stable than the adjusted least squares estimator in the case of small samples, since the matrix $\overline{H}$ (which can be nearly singular) is changed by a certain positive definite matrix when evaluating $\hat{\beta}_M$ (this procedure is similar to (11)). In addition, the estimator $\hat{\sigma}^2_{\varepsilon,M}$ is always positive in contrast to $\hat{\sigma}^2_{\varepsilon,A}$. If the covariance matrix Ω in (1) is nondegenerate as well as the limit matrix H_∞ in (iv), then the estimator $\hat{\beta}_M$ is asymptotically equivalent to $\hat{\beta}_A$, that is, $\sqrt{n}(\hat{\beta}_M - \hat{\beta}_A) \xrightarrow{\mathsf{P}} 0$ (see Theorem 2 in the paper [5]). Then the Slutskiĭ lemma [1] implies that Theorem 9 remains true for $\hat{\beta}_M$ if Ω is nondegenerate. Also the latter condition implies an analogue of Theorem 10 where we change the estimator $\hat{\beta}_A$ by $\hat{\beta}_M$ in the expression for $\hat{U}$.

5. Concluding remarks

We obtained weaker conditions for the consistency of adjusted least squares estimators as compared to those used in the paper [4] for the polynomial measurement error functional model. We also found sufficient conditions for the asymptotic normality of the estimator $\hat{\beta}_A$ (this property is not studied in [4]). Modified estimators $\hat{\beta}_M$ and $\hat{\sigma}^2_{\varepsilon,M}$ introduced in the paper [5] are more useful for small or moderate samples. Note that $\hat{\beta}_M$ possesses the same asymptotic covariance matrix as $\hat{\beta}_A$. The estimator $\hat{\beta}_M$ is asymptotically normal under the same conditions as those for $\hat{\beta}_A$ if matrix (1) is nondegenerate.

In a forthcoming research work we plan to investigate whether or not the estimators $\hat{\sigma}^2_{\varepsilon,A}$ and $\hat{\sigma}^2_{\varepsilon,M}$ are asymptotically equivalent. It would be interesting to consider the structure normal model (2) where ξ_i, ε_i, and δ_i are normal random variables. Naturally, the adjusted least squares estimator is useful for this model as well, but one can construct the quasi-likelihood estimator $\hat{\beta}_Q$. We would like to compare the asymptotic effectiveness of the estimators $\hat{\beta}_A$ and $\hat{\beta}_Q$. Such a comparison for noncorrelated errors ($\sigma_{\delta\varepsilon} = 0$) is done in [9] (also see survey [10] where some other regression measurement error models are mentioned). It is proved in [8] that the adjusted least squares estimator is asymptotically normal in the linear functional measurement error model. We are interested in obtaining such a result for the polynomial model.

Acknowledgement

The authors are indebted to Professor H. Schneeweiss (München, Germany) who turned our attention to the polynomial model with errors in observations and to S. V. Shklyar (Kyiv) for fruitful discussions.

Bibliography

1. P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York–London–Sydney, 1968. MR0233396
2. V. S. Korolyuk, N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, *A Manual on Probability Theory and Mathematical Statistics*, second edition, "Nauka", Moscow, 1985. (Russian) MR828455
3. Valentin V. Petrov, *Limit theorems of probability theory*, Sequences of independent random variables, Oxford Studies in Probability, vol. 4, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1995. MR1353441
4. C.-L. Cheng and H. Schneeweiss, *Polynomial regression with errors in the variables*, J. Royal Statist. Soc. B **60** (1998), no. 1, 189–199. MR1625632
5. C.-L. Cheng, H. Schneeweiss, and M. Thamerus, *A small sample estimator for a polynomial regression with errors in the variables*, J. Royal Statist. Soc. **62** (2000), no. 4, 699–709. MR1796286

6. C.-L. Cheng and A. Kukush, *Goodness-of-fit test in a polynomial errors-in-variables model*, Ukr. Mat. Zh. **56** (2004), no. 4, 527–543; English transl. in Ukrain. Math. J. **56** (2004), no. 4, 641–661. MR2105906
7. P. Hall and Y. Ma, *Testing the suitability of polynomial models in errors-in-variables problems*, Ann. Statist. **35** (2007), no. 6, 2620–2638. MR2382660
8. A. Kukush and E.-O. Maschke, *The efficiency of adjusted least squares in the linear functional relationship*, J. Multivariate Anal. **87** (2003), no. 2, 261–274. MR2016938
9. A. Kukush, A. Malenko, and H. Schneeweiss, *Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models*, J. Stat. Plann. Inference **139** (2009), no. 10, 3461–3472. MR2549095
10. H. Schneeweiss and A. Kukush, *Comparing the efficiency of structural and functional methods in measurement error models*, Teor. Ĭmovir. Mat. Stat. **80** (2009), 117–127; English transl. in Theory Probab. Math. Statist. **80** (2010), 131–142. MR2541958
11. I. O. Sen'ko, *Consistency of an adjusted least-squares estimator in a vector linear model with measurement errors*, Ukr. Mat. Zh. **64** (2012), no. 11, 1536–1546; English transl. in Ukrain. Math. J. **64** (2013), no. 11, 1739–1751. MR3104843
12. I. O. Sen'ko, *The asymptotic normality of an adjusted least squares estimator in a multivariate vector errors-in-variables regression model*, Teor. Ĭmovir. Mat. Stat. **88** (2013), 157–170; English transl. in Theory Probab. Math. Statist. **88** (2014), 175–190. MR3112643

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

Department of Mathematical Analysis, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue, 6, Kyiv 03127, Ukraine
E-mail address: 777Tsar777@mail.ru

Received 23/DEC/2014
Translated by S. KVASKO