

# Goodness-of-Fit Test in a Structural Errors-in-Variables Model Based on a Score Function

Alexander Kukush and Andrii Malenko  
Kyiv National Taras Shevchenko University, Ukraine

**Abstract:** A polynomial structural measurement error model is considered. A goodness-of-fit test is constructed based on the quasi-likelihood estimator, which is asymptotically optimal in a large class of estimators. The power of the test is discussed. The test for the linear model with unknown nuisance parameters is studied in more detail. Similar test can be applied to much more general situation, where the estimator is constructed based on a score function.

**Keywords:** Polynomial Measurement Error Model, Linear Measurement Error Model.

## 1 Introduction

Zhu, Song, and Cui (2003), Cheng and Kukush (2004), and Polekha (2006) constructed lack-of-fit tests for a polynomial errors-in-variables model (EIVM). That was a score type test using a weight function.

In this paper we propose a totally different idea. Our test is rather general, it is based on an estimator generated by an unbiased score function and involves its partial derivatives. The test is applied to a polynomial EIVM. We show how to disturb the initial model in order to construct a local alternative, which makes it possible to measure the power of the test. The test in a linear model is studied in more detail.

The paper is organized as follows. Section 2 presents the test in the most generality, where the model of i.i.d. observations is not specified. Section 3 contains a local alternative in rather general nonlinear EIVM. In Section 4 we introduce the quasi-score-like function for the polynomial EIVM, where all the nuisance parameters are unknown. In Sections 5 and 6 we derive the test statistics for the linear and polynomial EIVM, and Section 7 concludes. The main proofs are given in the Appendix.

## 2 General Test

Suppose that a family of  $d$ -dimensional distributions  $\{P_t \mid t \in \Theta\}$  is given. Here  $\Theta$  is a convex compact set in  $\mathbb{R}^d$ . By the i.i.d. observations  $z_1, \dots, z_n$  we want to construct a goodness-of-fit test for the hypothesis

$$H_0 : \mathcal{L}(z_i) \in \{P_t \mid t \in \Theta\}.$$

Now, we suppose that  $H_0$  holds, with a true value  $\theta$ . Let  $q(z, t)$  be a Borel measurable score function valued in  $\mathbb{R}^d$ . The estimator  $\hat{\theta}$  of  $\theta$  is defined as a measurable solution to

the equation

$$S_n(t) = 0, \quad S_n(t) := \frac{1}{n} \sum_{i=1}^n q(z_i, t), \quad t \in \Theta. \quad (1)$$

We need the following assumptions.

- (i)  $\theta$  is an interior point of  $\Theta$ .
- (ii) For each  $z \in \mathbb{R}^d$ , the vector function  $q(z, \cdot) \in C^3(U(\Theta))$ , where  $U(\Theta)$  is a neighborhood of  $\Theta$ .
- (iii)  $\mathbf{E}_\theta \|q(z, \theta)\|^2 < \infty$ , and for each  $t \in \Theta$ ,  $\mathbf{E}_t \sup_{s \in \Theta} \|D_s^j q(z, s)\| < \infty$ ,  $j = 1, 2, 3$ .
- (iv) The vector function  $S_\infty(t, s) := \mathbf{E}_s q(z, t) \in C^2(\Theta^2)$ , and  $S_\infty(t, s) = 0$  iff  $t = s$ .
- (v) The matrix  $V := \partial S_\infty(\theta, \theta) / \partial t^\top$  is nonsingular.

Assumption (iii) implies that for each  $t \in \Theta$ ,  $S_n(t) \rightarrow S_\infty(t, \theta)$ , as  $n \rightarrow \infty$   $P_\theta$ -a.s. The sequence  $\{S_n(\cdot) : n \geq 1\}$  is equicontinuous  $P_\theta$ -a.s. due to the boundedness of  $\mathbf{E} D_t q(z, \cdot)$ . Therefore  $P_\theta$ -a.s.

$$S_n(\cdot) \rightarrow S_\infty(\cdot, \theta), \quad \text{as } n \rightarrow \infty,$$

uniformly on  $\Theta$ . Moreover,  $S_\infty(t, t) = 0$  and  $V$  is nonsingular. Due to Kukush and Schneeweiss (2000), the equation (1) eventually has a solution, i.e.,  $\exists \Omega_0$ ,  $P(\Omega_0) = 1$ ,  $\forall \omega \in \Omega_0 \exists N(\omega) \forall n \geq N(\omega)$ : equation (1) has a solution.

Now, we give a more rigorous definition of the estimator  $\hat{\theta}$ . If (1) has no solution then we set  $\hat{\theta} = \theta_f$ , where  $\theta_f$  is a fixed vector from  $\Theta$ ; otherwise we choose an arbitrary solution in such a way that  $\hat{\theta}$  be a measurable function of  $\omega$ . This is possible due to Pfanzagl (1969). Therefore, the equality

$$S_n(\hat{\theta}) = 0 \quad (2)$$

holds eventually. Next,  $S_\infty(t, \theta) = 0$  iff  $t = \theta$ , therefore the estimator  $\hat{\theta}$  is strongly consistent, see Kukush and Schneeweiss (2000) for more detail.

To construct a goodness-of-fit test, introduce the test vector

$$f_n := \sqrt{n} \text{vec} \left( \frac{\partial S_n}{\partial t^\top}(\hat{\theta}) - \frac{\partial S_\infty}{\partial t^\top}(\hat{\theta}, \hat{\theta}) \right).$$

Denote

$$A = [I_{d^2} \ A_2], \quad A_2 = -\frac{\partial}{\partial t^\top} \text{vec} \left( \frac{\partial}{\partial t^\top} S_\infty \right) (\theta, \theta) \cdot V^{-1}, \quad (3)$$

where  $I_k$  stands for the unit matrix of size  $k$ , and in (3)  $k = d^2$ .

**Theorem 1.** Under assumptions (i) to (v),

$$f_n \xrightarrow{d} \mathcal{N}(0, B), \quad B = A \Sigma A^\top,$$

where the matrix  $\Sigma$  is variance-covariance matrix of the vector

$$\begin{pmatrix} \text{vec} \left( \frac{\partial}{\partial t^\top} q(z, \theta) \right) \\ q(z, \theta) \end{pmatrix}. \quad (4)$$

Thus that a statistic

$$T_n := \|B^{-1/2} f_n\|^2$$

is asymptotically  $\chi^2$ -distributed with  $d^2$  degrees of freedom which equals the size of  $B$ .

If  $B$  is degenerate then we transform  $T_n$  as follows. Let  $1 \leq r \leq \text{rank } B$ , and suppose that we can choose exactly  $r$  components of the vector  $f_n$  and form the  $r$ -dimensional subvector  $f_n^{(r)}$  in such a way that  $B^{(r)}$  be nonsingular matrix. Here  $B^{(r)}$  is the asymptotic covariance matrix of  $f_n^{(r)}$ . Then

$$T_n^{(r)} = \|B^{(r)-1/2} f_n^{(r)}\|^2 \xrightarrow{d} \chi_r^2.$$

Based on this convergence, a goodness-of-fit test is constructed.

### 3 Local Alternative in General Errors-in-Variables Model

Suppose that  $\xi$  and  $y$  are random variables and we know the first and second moments of  $y$  given  $\xi$  up to unknown vector parameter  $\beta$ :

$$m^*(\xi, \beta) = \mathbf{E}[y|\xi], \quad v^*(\xi, \beta) = \mathbf{V}[y|\xi]. \quad (5)$$

The model (5) is called *mean-variance* model for the couple  $(\xi, y)$ , see Carroll, Ruppert, Stefanski, and Crainicianu (2006). The  $\xi$  is latent variable. Instead of  $\xi$  we observe a surrogate random variable  $x$ , which is related to  $\xi$  by measurement equation

$$x = \xi + \delta, \quad (6)$$

where  $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$  is independent of  $y$  and  $\xi$ . The variance  $\sigma_\delta^2$  is known and positive.

The latent variable  $\xi$  has a probability density  $\rho(\xi, \alpha)$ , which is known up to a vector parameter  $\alpha$ . We observe  $z = (x, y)^\top$  and want to estimate  $\theta = (\alpha^\top, \beta^\top)^\top \in \mathbb{R}^d$ . This is *errors-in-variables model*, see Kukush, Malenko, and Schneeweiss (2006) for more details.

Here and hereafter we consider a quadratic-in- $y$  unbiased score function only, which is valued in  $\mathbb{R}^d$ :

$$q(x, y, t) = a(x, t) + b(x, t)y + c(x, t)y^2, \quad x \in \mathbb{R}, y \in \mathbb{R}, t \in \Theta.$$

Let  $H_0$  be the hypothesis that the observations  $(y_i, x_i)$ ,  $i = 1, \dots, n$  are i.i.d. copies of this model. Suppose that the local alternative to  $H_0$  is

$$H_{1n} : \tilde{y}_i = y_i + p(\xi_i)n^{-1/2}, \quad i = 1, \dots, n. \quad (7)$$

Here in (7)  $\tilde{y}_i$  is the observed response variable, and  $(y_i, x_i, \xi_i)$  are i.i.d. copies of the above mentioned model (5) and (6). We need the following relations to hold  $P_\theta$ -a.s. as  $n \rightarrow \infty$ :

$$\frac{1}{n} \sum_{i=1}^n \left( \sup_{t \in \Theta} \|D_t^j b(x_i, t)\| \cdot |p(\xi_i)| + \sup_{t \in \Theta} \|D_t^j c(x_i, t)\| \cdot |y_i p(\xi_i)| \right) = o(\sqrt{n}), \quad (8)$$

$$\frac{1}{n} \sum_{i=1}^n \sup_{t \in \Theta} \|D_t^j c(x_i, t)\| \cdot |p^2(\xi_i)| = o(n); \quad j = 0, 1. \quad (9)$$

Denote

$$\tilde{S}_n(t) = \frac{1}{n} \sum_{i=1}^n q(x_i, \tilde{y}_i, t),$$

and let  $\tilde{\theta}$  be a measurable solution to the equation  $\tilde{S}_n(t) = 0$ . Consider

$$\tilde{f}_n := \sqrt{n} \operatorname{vec} \left( \frac{\partial \tilde{S}_n}{\partial t^\top}(\tilde{\theta}) - \frac{\partial S_\infty}{\partial t^\top}(\tilde{\theta}, \tilde{\theta}) \right).$$

Here consistently with (iv),  $S_\infty(t, s) = \mathbf{E}_s q(x, y, t)$ ,  $t, s \in \Theta$ .

**Theorem 2.** Under assumptions (i) to (v), (8), and (9),

$$\tilde{f}_n = f_n + \bar{f} + o_p(1) \xrightarrow{d} \mathcal{N}(\bar{f}, B), \quad (10)$$

where

$$\bar{f} = \operatorname{vec} \mathbf{E}_\theta \frac{\partial^2 q}{\partial y \partial t^\top}(x, y, \theta) p(\xi) - \frac{\partial}{\partial t^\top} \operatorname{vec} \frac{\partial S_\infty}{\partial t^\top}(\theta, \theta) \cdot V^{-1} \mathbf{E}_\theta \frac{\partial q}{\partial y}(x, y, \theta) p(\xi),$$

and matrix  $B$  is given in Theorem 1 with matrix  $A$  given in (3).

As a consequence of Theorem 2 we have that under  $H_{1n}$  the statistic

$$\tilde{T}_n^{(r)} = \|B^{(r)-1/2} \tilde{f}_n^{(r)}\|^2 \xrightarrow{d} \chi_r^2(C), \quad \text{with } C = \|B^{(r)-1/2} \bar{f}^{(r)}\|.$$

Here  $\chi_r^2(C)$  is the noncentral  $\chi_r^2$  distribution with noncentrality  $C$ , i.e.

$$\chi_r^2(C) \sim (\zeta_1 + C)^2 + \sum_{i=2}^r \zeta_i^2, \quad \zeta_i \sim \mathcal{N}(0, 1), \quad \zeta_i \text{ are independent.}$$

The larger  $C$  the larger the power of the test.

## 4 Score Function in Polynomial Model

Consider the polynomial EIVM

$$\begin{cases} y_i = \beta_0 + \beta_1 \xi_i + \cdots + \beta_k \xi_i^k + \varepsilon_i, \\ x_i = \xi_i + \delta_i, \end{cases} \quad i = 1, \dots, n. \quad (11)$$

Here  $\xi \sim \mathcal{N}(\mu, \sigma_\xi^2)$ ,  $\varepsilon \sim \mathcal{N}(0, \varphi)$ ,  $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$ , all the variances are positive, and  $\xi_i, \varepsilon_i, \delta_i$ ,  $i = 1, \dots, n$  are mutually independent. The parameter  $\beta := (\beta_0, \dots, \beta_k)^\top$  is unknown vector parameter. The nuisance parameters  $\mu$ , and  $\sigma^2 := \sigma_\xi^2 + \sigma_\delta^2$ , and  $\varphi$  are unknown, while  $\sigma_\delta^2$  is assumed to be known. The total vector of unknown parameters is  $\theta$ .

Denote  $m(x, t) = \mathbf{E}_t[y|x]$ ,  $v(x, t) = \mathbf{E}_t[(y - m(x, t))^2|x]$ . To estimate  $\theta$  we use the quasi-score-like function  $q(x, t)$  with components

$$\begin{aligned} q^{(\beta)}(x, y, t) &= (y - m(x, t))m_\beta(x, t)v^{-1}(x, t), \\ q^{(\mu, \sigma)}(x, y, t) &= (x - \mu; (x - \mu)^2 - \sigma^2)^\top, \\ q^{(\varphi)}(x, y, t) &= (y - m(x, t))^2 - v(x, t). \end{aligned}$$

This function yields an optimal estimator for a large class of unbiased scores, see Kukush et al. (2006) and (Kukush and Malenko, 2008).

For assumptions (8) and (9) to hold it is enough to have  $\mathbf{E}_\theta p^2(\xi) < \infty$ .

## 5 Test in Linear Model

As a particular case consider the linear model,  $k = 1$ . Let  $\theta = (\beta^\top, \mu, \sigma, \varphi)^\top$ . We have the following result:  $\text{rank } B = 4$ , for all possible values of the parameters. We are able to choose a 4-dimensional vector  $f_n^{(r)}$ ,  $r = 4$ , such that for all possible values of  $\theta$ , the corresponding matrix  $B^{(r)}$  is nonsingular. For example, these are the first, second, third and seventh components of  $f_n$ . The choice is not unique.

We suppose that function  $p(\xi)$  is differentiable and

$$|p(\xi)| \leq c_1 e^{c_2|\xi|}$$

for some positive  $c_1$  and  $c_2$ . Then  $\mathbf{E}_\theta p^2(\xi) < \infty$  and assumptions (8) and (9) hold.

Introduce the reliability ratio  $K = 1 - \sigma_\delta^2 \sigma^{-2}$ , and  $\tau^2 = K \sigma_\delta^2$ . Then under the local alternative  $H_{1n}$  the deviation vector  $\bar{f}$  from Theorem 2 is equal

$$\bar{f} = \frac{\mathbf{E}_\theta p'(\xi)}{v^2} \times$$

$$\times \text{vec} \begin{bmatrix} 2\tau^2\beta_1 & 2\mu\tau^2\beta_1 & (1-K)(v-2\varphi) & 0 & 0 \\ 2\mu\tau^2\beta_1 & 2\tau^2(\mu^2 + \sigma_\xi^2 + \tau^2)\beta_1 & \mu(1-K)(v-2\varphi) & 2\tau^2(v-2\varphi)/\sigma & 0 \\ -(1-K)(v-2\varphi) & \mu(1-K)(v-2\varphi) & -2\varphi(1-K)^2\beta_1 & 0 & 0 \\ 0 & 2\tau^2(v-2\varphi)/\sigma & 0 & -8\varphi(1-K)^2\beta_1 & 0 \\ 0 & -2\tau^2v^2 & 0 & -4(1-K)^2\sigma v^2\beta_1 & 0 \end{bmatrix}.$$

Let  $\mathbf{E}_\theta p'(\xi) \neq 0$ . For  $r = 4$  we select a subvector  $f_n^{(r)}$  in such a way that the corresponding  $B^{(r)}$  is nonsingular, and then the corresponding deviation vector  $\bar{f}^{(r)}$  will not vanish. We will have

$$T_n^{(r)} = \|B^{(r)-1/2} f_n^{(r)}\|^2 \xrightarrow{d} \chi_r^2(\|B^{(r)-1/2} \bar{f}^{(r)}\|).$$

We mention that the proposed test is very sensitive: even for a linear disturbance function  $p(\xi)$ , the noncentrality parameter in the limit distribution of the  $T_n^{(r)}$  will not be equal zero.

## 6 Test in Polynomial Model

In the polynomial model of order  $k \geq 2$ , it is not easy to compute  $\text{rank } B$ , because the matrix  $\Sigma$  is always degenerate. Therefore we propose a modified test vector

$$f_{h,n} := \sqrt{n} \left( \frac{\partial S_n}{\partial t^\top}(\hat{\theta}) - \frac{\partial S_\infty}{\partial t^\top}(\hat{\theta}, \hat{\theta}) \right) h,$$

where  $h$  is fixed nonzero vector from  $\mathbb{R}^d$ . Then

$$f_{h,n} \xrightarrow{d} \mathcal{N}(0, B_h), \quad B_h = A_h \Sigma_h A_h^\top,$$

where

$$A_h = [I_d \ A_{h,2}], \quad A_{h,2} = -\frac{\partial^2 (h^\top S_\infty)}{\partial t \partial t^\top}(\theta, \theta) \cdot V^{-1},$$

and  $\Sigma_h$  is variance-covariance matrix for the vector  $[\partial h^\top q(\theta) / \partial t^\top; q^\top]^\top$ .

For known nuisance parameters (i.e. when  $\theta = \beta$ ) under the condition that the true  $\beta_k \neq 0$ , for any  $h \in \mathbb{R}^{k+1}$  such that  $h_k \neq 0$  we have that  $\Sigma_h$  is nonsingular. This implies that  $B_h$  is nonsingular as well and

$$T_{h,n} := \|B_h^{-1/2} f_{h,n}\|^2 \xrightarrow{d} \chi_{k+1}^2, \quad \text{as } n \rightarrow \infty.$$

Under the local alternative  $H_{1n}$  given in (7), the modified test vector has an expansion similar to (10),

$$\tilde{f}_{h,n} = f_{h,n} + \bar{f}_h + o_p(1), \quad \text{as } n \rightarrow \infty,$$

where the modified deviation equals

$$\bar{f}_h = \mathbf{E}_\theta p(\xi) \frac{\partial^2 (h^\top q)}{\partial y \partial t^\top}(x, y, \theta) + A_{h,2} \cdot \mathbf{E}_\theta p(\xi) \frac{\partial q}{\partial y}(x, y, \theta).$$

Therefore, under  $H_{1n}$  we have

$$\tilde{T}_{h,n} := \|B_h^{-1/2} \tilde{f}_{h,n}\|^2 \xrightarrow{d} \chi_{k+1}^2(\|B_h^{-1/2} \bar{f}_h\|).$$

The larger  $\|B_h^{-1/2} \bar{f}_h\|$  the larger the power of the test.

## 7 Conclusion

We constructed a very general goodness-of-fit test, which works for any unspecified model of observations, where the estimator is generated by an unbiased score function. In errors-in-variables setup, we proposed a local alternative for the test. In a forthcoming paper we will compare the power of this test and other tests known in the literature.

## Acknowledgements

Alexander Kukush is supported by the Swedish Institute grant SI-01424/2007.

## References

- Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainicianu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.). New York: Chapman & Hall, CRC.
- Cheng, C.-L., and Kukush, A. (2004). Goodness-of-fit test in a polynomial errors-in-variables model. *Ukrainian Mathematical Journal*, 56, 527-543.
- Kukush, A., and Malenko, A. (2008). Optimal joint estimator of regression and dispersion parameters in nonlinear measurement error models. *Probability Theory and Mathematical Statistics*. (To appear)
- Kukush, A., Malenko, A., and Schneeweiss, H. (2006). *Optimality of the quasi-score estimator in a mean-variance model with applications to measurement error models* (Discussion Paper No. 494). SFB 386. Ludwig-Maximilians-University of Munich.
- Kukush, A., and Schneeweiss, H. (2000). *A comparison of asymptotic covariance matrices of adjusted least squares and structural least squares in error ridden polynomial regression* (Discussion Paper No. 218). SFB 386. Ludwig-Maximilians-University of Munich.
- Pfanzagl, G. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika*, 14, 249-273.
- Polekha, M. Y. (2006). Goodness-of-fit test with polynomial weight function in polynomial EIVM. *Applied Statistics. Actuarial and Financial Math.*, 1-2, 141-148.
- Zhu, L. X., Song, W. X., and Cui, H. J. (2003). Testing lack-of-fit for a polynomial errors-in-variables model. *Acta Math. Appl. Sin. Engl. Ser.*, 19, 353-362.

## Appendix

### Proof of Theorem 1

1°. First prove the asymptotic normality of  $\hat{\theta}$ .

Denote  $\Delta\theta = \hat{\theta} - \theta$ . From (2) and the consistency of  $\hat{\theta}$  we have that eventually

$$S_n(\theta) + \frac{\partial S_n}{\partial t^\top}(\theta)\Delta\theta + r_n = 0, \quad (12)$$

where

$$\|r_n\| \leq \frac{1}{2} \sup_{t \in \Theta} \left\| \frac{\partial^2 S_n}{\partial t^\top \partial t}(t) \right\| \cdot \|\Delta\theta\|^2.$$

But due to (iii),  $\mathbf{E}_\theta \sup_{t \in \Theta} \|D_t^2 q(z, t)\| < \infty$ . Then  $P_\theta$ -a.s.

$$\sup_{t \in \Theta} \left\| \frac{\partial^2 S_n}{\partial t^\top \partial t}(t) \right\| \leq \frac{1}{n} \sum_{i=1}^n \sup_{t \in \Theta} \left\| \frac{\partial^2 q}{\partial t^\top \partial t}(z, t) \right\| \rightarrow \mathbf{E}_\theta \sup_{t \in \Theta} \left\| \frac{\partial^2 q}{\partial t^\top \partial t}(z, t) \right\| < \infty,$$

as  $n \rightarrow \infty$ , and  $P_\theta$ -a.s.

$$\sup_{n \geq 1} \sup_{t \in \Theta} \left\| \frac{\partial^2 S_n}{\partial t^\top \partial t}(t) \right\| < \infty,$$

implying

$$\|r_n\| = O(\|\Delta\theta\|^2).$$

From (12) we have

$$\begin{aligned} \frac{\partial S_n}{\partial t^\top}(\theta)\sqrt{n}\Delta\theta &= -\sqrt{n}S_n(\theta) - \sqrt{n}r_n, \\ \sqrt{n}\Delta\theta &= -V^{-1} \cdot \sqrt{n}S_n(\theta) + o_p(1). \end{aligned} \quad (13)$$

By (iii) and (v)

$$\sqrt{n}S_n(\theta) \xrightarrow{d} N(0, \Sigma_1), \quad \Sigma_1 := \mathbf{E}_\theta q(z, \theta)q^\top(z, \theta).$$

Therefore, we get

$$\sqrt{n}\Delta\theta \xrightarrow{d} \mathcal{N}(0, V^{-1}\Sigma_1V^{-\top}).$$

2°. The next step is to prove the asymptotic normality of  $f_n$ . Consider the  $(i, j)$ -th element of the matrix, which forms  $f_n$ :

$$\begin{aligned} s_{ij} &= \frac{\partial S_{ni}}{\partial t_j}(\hat{\theta}) - \frac{\partial S_{\infty i}}{\partial t_j}(\hat{\theta}, \hat{\theta}) = \\ &= \frac{\partial S_{ni}}{\partial t_j}(\theta) + \frac{\partial^2 S_{ni}}{\partial t_j \partial t}(\theta)\Delta\theta - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) - L_{ij}\Delta\theta + O(\|\Delta\theta\|^2), \end{aligned}$$

where

$$L_{ij} = \frac{\partial}{\partial t_j} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) S_{\infty i}(t, s) \Big|_{(\theta, \theta)} = 0,$$

due to the condition  $S_\infty(t, t) = 0$ . Here assumptions (iii) and (iv) were used together with the same reasoning as in the first part of the proof. Applying (13) we have

$$\begin{aligned} \sqrt{n}s_{ij} &= \sqrt{n} \left( \frac{\partial S_{ni}}{\partial t_j}(\theta) - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) \right) + \frac{\partial^2 S_{ni}}{\partial t_j \partial t}(\theta)\sqrt{n}\Delta\theta + o_p(1) = \\ &= \sqrt{n} \left( \frac{\partial S_{ni}}{\partial t_j}(\theta) - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) \right) - \frac{\partial^2 S_{\infty i}}{\partial t_j \partial t}(\theta, \theta)V^{-1}\sqrt{n}S_n(\theta) + o_p(1) = \\ &= \left[ 1 \quad - \frac{\partial^2 S_{\infty i}}{\partial t_j \partial t}(\theta, \theta)V^{-1} \right] \cdot \sqrt{n} \left[ \frac{\partial S_{ni}}{\partial t_j}(\theta) - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) \quad S_n^\top(\theta) \right]^\top + o_p(1). \end{aligned}$$

Therefore,

$$f_n = A \cdot g_n + o_p(1),$$

where  $A$  is defined in (3) and the sequence

$$g_n := \sqrt{n} \begin{pmatrix} \text{vec} \left( \frac{\partial}{\partial t^\top} S_n(\theta) - \frac{\partial}{\partial t^\top} S_\infty(\theta, \theta) \right) \\ S_n(\theta) \end{pmatrix}$$

satisfies the Central Limit Theorem.  $g_n \xrightarrow{d} \mathcal{N}(0, \Sigma)$  with the matrix  $\Sigma$  given in (4). We have  $f_n \xrightarrow{d} \mathcal{N}(0, A\Sigma A^\top)$ .



## Proof of Theorem 2

Consider the difference between  $\tilde{S}_n(t)$  and  $S_n(t)$ . Assumptions (8) and (9) with  $j = 0$  imply that

$$\begin{aligned} & \sup_{t \in \Theta} \|\tilde{S}_n(t) - S_n(t)\| \leq \\ & \leq \frac{1}{n\sqrt{n}} \sum_{i=1}^n |p(\xi_i)| \sup_{t \in \Theta} \left\| \frac{\partial q}{\partial y}(x_i, y_i, t) \right\| + \frac{1}{n^2} \sum_{i=1}^n |p^2(\xi_i)| \sup_{t \in \Theta} \left\| \frac{\partial^2 q}{\partial y^2}(x_i, y_i, t) \right\| \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$   $P_\theta$ -a.s. Then  $\tilde{S}_n(t) \rightarrow S_\infty(t, \theta)$  uniformly in  $t \in \Theta$   $P_\theta$ -a.s. (here  $S_\infty$  is the same as in assumption (iv)). Then eventually a solution to the equation  $\tilde{S}_n(t) = 0$  exists and the equality  $\tilde{S}_n(\tilde{\theta}) = 0$  holds implying  $\tilde{\theta} \rightarrow \theta$  as  $n \rightarrow \infty$   $P_\theta$ -a.s. We have a.s.

$$\begin{aligned} \tilde{S}_n(\theta) + \frac{\partial \tilde{S}_n}{\partial t^\top}(\theta)(\tilde{\theta} - \theta) &= O(\|\tilde{\theta} - \theta\|^2), \\ \sqrt{n}(\tilde{\theta} - \theta) &= - \left( \frac{\partial \tilde{S}_n}{\partial t^\top}(\theta) \right)^{-1} \sqrt{n} \tilde{S}_n(\theta) + o_p(1). \end{aligned}$$

Under assumptions (iii), (iv), (8), and (9) with  $j = 1$ ,

$$\frac{\partial \tilde{S}_n}{\partial t^\top}(\theta) \rightarrow \frac{\partial S_\infty}{\partial t^\top}(\theta, \theta) = V, \quad \sqrt{n} \tilde{S}_n(t) = \sqrt{n} S_n(t) + \mathbf{E}_\theta \frac{\partial q}{\partial y}(x, y, \theta) p(\xi) + o_p(1).$$

Thus

$$\sqrt{n}(\tilde{\theta} - \theta) = -V^{-1} \sqrt{n} \tilde{S}_n(\theta) + o_p(1),$$

and

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} \mathcal{N}(\lambda, \Sigma_1), \quad \lambda = -V^{-1} \mathbf{E}_\theta \frac{\partial q}{\partial y}(x, y, \theta) p(\xi), \quad \Sigma_1 := \mathbf{E}_\theta q(z, \theta) q^\top(z, \theta).$$

Investigating the  $\tilde{f}_n$  we repeat the reasoning from the proof of Theorem 1. Denote by  $\tilde{s}_{ij}$  the  $(i, j)$ -th element of the matrix, which forms  $\tilde{f}_n$ . Then

$$\begin{aligned} \tilde{s}_{ij} &= \sqrt{n} \left( \frac{\partial \tilde{S}_{ni}}{\partial t_j}(\theta) - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) \right) + \frac{\partial^2 \tilde{S}_{ni}}{\partial t_j \partial t}(\theta) \sqrt{n}(\tilde{\theta} - \theta) + o_p(1) = \\ & \sqrt{n} \left( \frac{\partial \tilde{S}_{ni}}{\partial t_j}(\theta) - \frac{\partial S_{\infty i}}{\partial t_j}(\theta, \theta) \right) - \frac{\partial^2 S_{\infty i}}{\partial t_j \partial t}(\theta, \theta) V^{-1} \sqrt{n} \tilde{S}_n(\theta) + o_p(1) = \\ & s_{ij} + \mathbf{E}_\theta \frac{\partial^2 q_i}{\partial y \partial t_j}(x, y, \theta) p(\xi) + \frac{\partial^2 S_{\infty i}}{\partial t_j \partial t}(\theta, \theta) \cdot \lambda + o_p(1). \end{aligned}$$

Taking vec operation we obtain the expansion (10) and then the statement of the theorem.

## Corresponding Author's Address:

Alexander Kukush  
Kyiv National Taras Shevchenko University  
Volodymyrska st. 64  
01034, Kyiv  
Ukraine  
E-mail: [alexander\\_kukush@univ.kiev.ua](mailto:alexander_kukush@univ.kiev.ua)