GOODNESS-OF-FIT TEST IN A STRUCTURAL ERRORS-IN-VARIABLES MODEL BASED ON THE QUASI-LIKELIHOOD ESTIMATOR

A.G. Kukush and A.L. Malenko
Kyiv National Taras Shevchenko University, Kyiv, Ukraine
e-mails: alexander_kukush@univ.kiev.ua, exipilis@yandex.ru

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 is studied in more detail. Similar test can be applied to much more general situation, where the estimator is constructed by optimization or score equations.

1 Introduction

Zhu et al. (2003) and Cheng and Kukush (2004) 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.

Suppose that a family of d-dimensional distributions $\{P_t |\ t \in \Theta\}$ is given. Here Θ is a convex compact set in $\mathbb{R}^d$. By the i.i.d. observations $x_1, \ldots, x_n$ we want to construct a goodness-of-fit test for the hypothesis

$$H_0: \quad \mathcal{L}(x_i) \in \{P_t |\ t \in \Theta\}.$$

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

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

We need the following assumptions.

(i) θ is an interior point of Θ.

(ii) $q(x, \cdot) \in C^3(U(\Theta))$, $U(\Theta)$ is a neighborhood of Θ.

(iii) For each $\theta \in \Theta$, $\mathbf{E}_\theta \|q(x,\theta)\|^2 < \infty$, $\mathbf{E}_\theta \sup_{t\in\Theta} \|D_t^j q(x,t)\| < \infty$, $j = 1, 2, 3$.

(iv) $S_\infty(t,s) := \mathbf{E}_s\, q(x,t) \in C^2(\Theta^2)$, for each $t, s \in \Theta$, and $S_\infty(t,s) = 0$ iff $t = s$.

(v) $V := \frac{\partial S_\infty}{\partial t^\top}(\theta, \theta)$ is nonsingular.

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

$$f_n = \sqrt{n}\,\mathbf{v}\mathrm{ec}\left(\frac{\partial S_n}{\partial t^\top}(\hat\theta) - \frac{\partial S_\infty}{\partial t^\top}(\hat\theta,\hat\theta)\right).$$

Under (i) to (v), we prove that

$$f_n = A\cdot g_n + o_p(1), \quad n\to\infty, \tag{2}$$

where the matrix

$$A = (I_{k^2};\ A_2), \quad A_2 = -\frac{\partial}{\partial t^\top}\mathbf{v}\mathrm{ec}\left(\frac{\partial}{\partial t^\top}S_\infty\right)(\theta,\theta)\cdot V^{-1},$$

I_s stands for the unit matrix of size s, and the sequence

$$g_n = \sqrt{n}\begin{pmatrix} \mathbf{v}\mathrm{ec}\left(\frac{\partial}{\partial t^\top}S_n(\theta) - \frac{\partial}{\partial t^\top}S_\infty(\theta,\theta)\right) \\ S_n(\theta)\end{pmatrix}.$$

Due to the CLT, $g_n \xrightarrow{d} \mathcal{N}(0,\Sigma)$, where Σ is a dispersion matrix of the vector

$$\begin{pmatrix}\mathbf{v}\mathrm{ec}\left(\frac{\partial}{\partial t^\top}q(x,\theta)\right) \\ q(x,\theta)\end{pmatrix}.$$

Therefore $f_n \xrightarrow{d} \mathcal{N}(0,B)$, $B = A\Sigma A^\top$. We have that a statistic $T_n := \|B^{-1/2}f_n\|^2$ is asymptotically χ^2-distributed with d^2 degrees of freedom which equals the size of B. If B is degenerate then we transform (2) as follows. Let $1 \le r \le \mathbf{r}\mathrm{ank}\,B$, and suppose that we can choose exactly r components of the vector f_n and form the r-dimensional vector $f_n^{(r)}$ in such a way that $B^{(r)}$ be nonsingular matrix. Here $B^{(r)}$ is the asymptotic covariance matrix of $f_n^{(r)}$. Then $T_n^{(r)} = \|{B^{(r)}}^{-1/2}f_n^{(r)}\|^2 \xrightarrow{d} \chi^2_r$. Based on this convergence, a goodness-of-fit test is constructed.

2 Score function in polynomial model

Consider the polynomial EIVM

$$\begin{cases} y_i = \beta_0 + \beta_1\xi_i + \ldots + \beta_k\xi_i^k + \varepsilon, \\ x_i = \xi_i + \delta_i, \end{cases} \qquad i = \overline{1,n}. \tag{3}$$

Here $\xi \sim \mathcal{N}(\mu,\sigma_\xi^2)$, $\varepsilon \sim \mathcal{N}(0,\varphi)$, $\delta \sim \mathcal{N}(0,\sigma_\delta^2)$, $(\xi_i,\varepsilon_i,\delta_i)$, $i = \overline{1,n}$ are independent vectors, ξ_i, δ_i, and ε_i are independent random variables for each $i = \overline{1,n}$. The parameter $\beta = (\beta_0,\ldots,\beta_k)^\top$ is unknown vector parameter. The nuisance parameters μ, $\sigma^2 := \sigma_\xi^2 + \sigma_\delta^2$, and φ could be unknown, while σ_δ^2 is known. The total vector of unknown parameters is θ.

Denote $m(x,t) = \mathbf{E}_t[y|x]$, $v(x,t) = \mathbf{E}_t[(y - m(x,t))^2|x]$. To estimate θ 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).

3 Linear model

Consider the linear model, $k = 1$. Let $\theta = (\beta^\top, \mu, \sigma, \varphi)^\top$. We have the following result: $\mathbf{r}\mathrm{ank}\, B = 4$, 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 θ, the corresponding matrix $B^{(r)}$ is nonsingular.

Introduce the local alternative to the hypothesis H_0. Let

$$H_{1n}: \quad \begin{cases} \tilde{y}_i = y_i + g(\xi_i)n^{-1/2}, \\ x_i = \xi_i + \delta_i, \end{cases} \qquad i = \overline{1,n}. \tag{4}$$

Here $g(\xi)$ is rather smooth function, $|g(z)| \le c_1 e^{c_2|z|}$ for some positive c_1 and c_2. Then the test vector under H_{1n} has an expansion

$$\tilde{f}_n = f_n + \bar{f} + o_p(1), \quad n \to \infty, \tag{5}$$

where the deviation vector equals

$$\bar{f} = \mathbf{vec}\, \mathbf{E}_\theta\, g(\xi) \frac{\partial^2 q}{\partial y \partial t^\top}(x,y,\theta) + A_2 \cdot \mathbf{E}_\theta\, g(\xi) \frac{\partial q}{\partial y}(x,y,\theta).$$

Denote $K = 1 - \sigma_\delta^2 \sigma^{-2}, \tau^2 = K\sigma_\delta^2$. We can rewrite

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

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

Let $\mathbf{E}_\theta\, g'(\xi) \neq 0$. For $r = 4$ we select a vector $f_n^{(r)}$ 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

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

Here

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

The larger $|E_\theta g'(\xi)|$ the larger the power of the test.

4 Polynomial model

In the polynomial model ($k \geq 2$) it's not easy to compute $\mathbf{r}\text{ank}\, B$ because Σ 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

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

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 Σ_h is a dispersion matrix for the vector $(\frac{\partial}{\partial t^\top}(h^\top q)(\theta), q^\top)^\top$.

For known nuisance parameters (i.e. when $\theta = \beta$) under the condition that the true $\beta_k \neq 0$, it is possible to choose $h \in \mathbb{R}^{k+1}$ such that Σ_h is nonsingular. This implies that B_h is nonsingular as well and

$$T_{h,n} := \|B_h^{-1/2} f_{h,n}\|^2 \xrightarrow{d} \chi^2_{k+1}, \quad n \to \infty.$$

Under the local alternative (4), the modified test vector has an expansion similar to (5),

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

where the modified deviation equals

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

Therefore, under H_{1n} we have

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

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

References

[1] Cheng C.-L., and Kukush A. G. (2004). Goodness-of-fit test in a polynomial errors-in-variables model. *Ukrainian Math. J.*. Vol. **56**, 527-543.

[2] 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 494, SFB 386. Ludwig-Maximilians-University of Munich.

[3] Zhu L. X., Song W. X., and Cui H. J. (2003). Testing lack-of-fit for a plynomial errors-in-variables model. *Acta Math. Appl. Sin. Engl. Ser.* Vol. **19**, 353-362.