Economics Letters 131 (2015) 41–46

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Consistency of the least squares estimator in threshold regression with endogeneity[☆]

Ping Yu

School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong

HIGHLIGHTS

- The LSE of the threshold point with endogeneity is consistent if the threshold variable is independent of other covariates.
- The LSE of the threshold point with endogeneity is inconsistent if the threshold variable is dependent of other covariates.
- The LSE of the threshold point is inconsistent when endogeneity is not additively linear in the threshold variable.

ARTICLE INFO

Article history:
Received 30 November 2014
Received in revised form
28 February 2015
Accepted 26 March 2015
Available online 2 April 2015

JEL classification:
C13
C14
C21

Keywords:
Threshold regression
Endogeneity
Least squares estimation
Consistency

ABSTRACT

This paper shows that when the threshold variable is independent of other covariates, such as in the structural change model, the least squares estimator of the threshold point is consistent even if endogeneity is present.

The endogeneity problem in threshold regression attracts much attention in the recent econometric practice; see Yu and Phillips (2014) for a summary of the literature in the threshold model and the related structural change model. The usual threshold regression model splits the sample according to the realized value of some observed threshold variable q. The dependent variable y is determined by covariates $\mathbf{x}$ in the split-sample regression

$$y = \mathbf{x}'\beta_1 1(q \le \gamma) + \mathbf{x}'\beta_2 1(q > \gamma) + \varepsilon, \tag{1}$$

where the indicators $1(q \le \gamma)$ and $1(q > \gamma)$ define two regimes in terms of the value of q relative to a threshold point given by the parameter γ, the coefficients β_1 and β_2 are the respective threshold parameters, and ε is a random disturbance which may not follow the same distribution in the two regimes (e.g., $\varepsilon = \sigma_1\epsilon 1(q \le \gamma) + \sigma_2\epsilon 1(q > \gamma)$ with $\sigma_1 \ne \sigma_2$ and ϵ i.i.d.). When there is endogeneity, $E[\varepsilon|\mathbf{x}, q] \ne 0$, and the usual solution to consistently estimate γ is to employ some instrumental variables. However, Perron and Yamamoto (forthcoming) suggest to use the least squares estimator (LSE) to estimate γ in the structural change model when $E[\varepsilon|\mathbf{x}] \ne 0$.[1] Their arguments are as follows. First project ε on $\mathbf{x}$ to get the projection $\mathbf{x}'\delta$, and then y would satisfy

$$y = \mathbf{x}'(\beta_1 + \delta)1(q \le \gamma) + \mathbf{x}'(\beta_2 + \delta)1(q > \gamma) + e,$$

where $e = \varepsilon - \mathbf{x}'\delta$ satisfies $E[\mathbf{x}e] = \mathbf{0}$. Since the linear (in $\mathbf{x}$) structure of the system remains, the LSE of γ is consistent although the LSEs of β_1 and β_2 may not be. Nevertheless, as emphasized in Yu (2013), only if $E[e|\mathbf{x}] = 0$ (rather than $E[\mathbf{x}e] = \mathbf{0}$) the LSE of γ is consistent. Perron and Yamamoto (forthcoming) apply the result of Perron and Qu (2006) to obtain the consistency of the LSE of γ, but Assumption A.4 of Perron and Qu (2006) essentially requires $E[e|\mathbf{x}] = 0$.

[☆] The author thanks Roger Klein for suggesting the problem considered in this paper.

E-mail address: pingyu@hku.hk.

[1] In the structural change model, q is the time index and independent of the rest components of the system.

http://dx.doi.org/10.1016/j.econlet.2015.03.035

In this paper, we show a seemingly surprising result: in Perron and Yamamoto (forthcoming)'s framework, even if $E[e|\mathbf{x}]$ is any *nonlinear* function of $\mathbf{x}$, the LSE of γ is still consistent. The key assumption for this result is that q is the time index and is independent of $\mathbf{x}$ in the structural change model. In the threshold model, this result can be extended to the case where the endogeneity in q takes an additively linear form but the assumption of q independent of other covariates cannot be relaxed in general.

Before our formal discussion, we first define the LSE of γ. Usually, the LSE of γ is defined by a profiled procedure:

$$\widehat{\gamma} = \arg\min_{\gamma} M_n(\gamma),$$

where

$$M_n(\gamma) = \min_{\beta_1,\beta_2} \frac{1}{n}\sum_{i=1}^{n} m(w_i|\theta), \quad (2)$$

with $w_i = (y_i, \mathbf{x}_i', q_i)'$, $\theta \equiv (\beta_1', \beta_2', \gamma)'$, and

$$m(w|\theta) = \left(y - \mathbf{x}'\beta_1 1(q \le \gamma) - \mathbf{x}'\beta_2 1(q > \gamma)\right)^2.$$

Denote $\left(\widehat{\beta}_1(\gamma), \widehat{\beta}_2(\gamma)\right) = \arg\min_{\beta_1,\beta_2} n^{-1}\sum_{i=1}^{n} m(w_i|\theta)$ in (2).

A word on notation: f and F denote the probability distribution function (pdf) and the cumulative distribution function (cdf) of q. $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively. $U[0, 1]$ means the uniform distribution on $[0, 1]$ and $N(0, 1)$ means the standard normal distribution. For any two random vectors x and y, $x \perp y$ means that x is independent of y, and $x \not\perp y$ means that x is not independent of y. plim means the probability limit. ℓ is always used for indicating the two regimes in (1), so it is not written out explicitly as "$\ell = 1, 2$" throughout the paper.

1. Consistency of the LSE when $q \perp x$

We start from a simpler model to get the essence of our arguments. Suppose $\mathbf{x} = (1, x')'$, where x does not include q. In this case, suppose

$$E[\varepsilon|x, q] = \eta_1(x)1(q \le \gamma_0) + \eta_2(x)1(q > \gamma_0) \ne 0, \quad (3)$$

where $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are two smooth functions. Note that we allow the endogeneity to have threshold effects at $q = \gamma_0$; when $\eta_1(x) = \eta_2(x)$, the endogeneity is smooth. Here, we intend to assume q is exogenous as in Caner and Hansen (2004) and Perron and Yamamoto (forthcoming). Notwithstanding, rigorously speaking, q is allowed to be endogenous but only through the threshold indicator $1(q \le \gamma)$; when $\eta_1(x) = \eta_2(x)$, q is exogenous. Under (3), the model can be rewritten as

$$y = g_1(x)1(q \le \gamma_0) + g_2(x)1(q > \gamma_0) + e, \quad (4)$$

where $g_\ell(x) = \mathbf{x}'\beta_\ell + \eta_\ell(x)$ and $e = \varepsilon - E[\varepsilon|x, q]$ satisfies $E[e|x, q] = 0$. Although $E[y|x, q]$ is a nonlinear function of x, we still use the LSE to estimate γ. The following theorem shows that the LSE of γ is consistent when $q \perp x$.

Theorem 1. *Suppose $\{w_i\}_{i=1}^n$ are i.i.d., $\gamma_0 \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$ which is compact, $E[e^2] < \infty$, $E[\mathbf{x}\mathbf{x}'] > 0$, $E[\mathbf{x}g_1(x)] \ne E[\mathbf{x}g_2(x)]$, and $f(\gamma)$ is continuous with $F(\underline{\gamma}) > 0$, $1 - F(\overline{\gamma}) > 0$ and $0 < \underline{f} \le f(\gamma) \le \overline{f} < \infty$ for $\gamma \in \Gamma$. If $q \perp x$, then $\widehat{\gamma}$ is consistent.*

Proof. Define the $n \times 1$ vectors $Y, \mathbf{e}, G_\ell, Q$ by stacking the variables $y_i, e_i, g_\ell(x_i)$, and q_i, the $n \times \dim(\mathbf{x})$ matrix X by stacking the vectors $\mathbf{x}_i'$, and the $n \times n$ matrices $I_{\le\gamma}$ and $I_{>\gamma}$ as $\mathrm{diag}\{1(q_i \le \gamma)\}$ and $\mathrm{diag}\{1(q_i > \gamma)\}$. In this notation system, $Y = I_{\le\gamma_0}G_1 + I_{>\gamma_0}G_2 + \mathbf{e}$, and

$$\widehat{\beta}_1(\gamma) = \left(X'I_{\le\gamma}X\right)^{-1}X'I_{\le\gamma}Y, \qquad \widehat{\beta}_2(\gamma) = \left(X'I_{>\gamma}X\right)^{-1}X'I_{>\gamma}Y.$$

Suppose first $\gamma \le \gamma_0$.

$$\begin{aligned}\widehat{\beta}_1(\gamma) &= \left(X'I_{\le\gamma}X\right)^{-1}X'I_{\le\gamma}\left(I_{\le\gamma_0}G_1 + I_{>\gamma_0}G_2 + \mathbf{e}\right)\\ &\xrightarrow{p} E\left[\mathbf{x}\mathbf{x}'1(q \le \gamma)\right]^{-1}E\left[\mathbf{x}g_1(x)1(q \le \gamma)\right]\\ &= E[\mathbf{x}\mathbf{x}']^{-1}E\left[\mathbf{x}g_1(x)\right] \equiv b_1,\end{aligned}$$

and

$$\begin{aligned}\widehat{\beta}_2(\gamma) &= \left(X'I_{>\gamma}X\right)^{-1}X'I_{>\gamma}\left(I_{\le\gamma_0}G_1 + I_{>\gamma_0}G_2 + \mathbf{e}\right)\\ &\xrightarrow{p} E\left[\mathbf{x}\mathbf{x}'1(q > \gamma)\right]^{-1}\{E\left[\mathbf{x}g_1(x)1(\gamma < q \le \gamma_0)\right]\\ &\quad + E\left[\mathbf{x}g_2(x)1(q > \gamma_0)\right]\}\\ &= E[\mathbf{x}\mathbf{x}']^{-1}E\left[\mathbf{x}g_1(x)\right]\frac{F(\gamma_0) - F(\gamma)}{1 - F(\gamma)}\\ &\quad + E[\mathbf{x}\mathbf{x}']^{-1}E\left[\mathbf{x}g_2(x)\right]\frac{1 - F(\gamma_0)}{1 - F(\gamma)}\\ &\equiv b_1\frac{F(\gamma_0) - F(\gamma)}{1 - F(\gamma)} + b_2\frac{1 - F(\gamma_0)}{1 - F(\gamma)}\\ &= b_1 + (b_2 - b_1)\frac{1 - F(\gamma_0)}{1 - F(\gamma)} \equiv b_2(\gamma),\end{aligned}$$

uniformly for $\gamma \in \left[\underline{\gamma}, \gamma_0\right]$ by a Glivenko–Cantelli theorem, where the second equalities use the assumption that $q \perp x$. Given that $E[\mathbf{x}g_1(x)] \ne E[\mathbf{x}g_2(x)]$ and $E\left[\mathbf{x}\mathbf{x}'\right] > 0$, $b_1 \ne b_2$. Now,

$$\begin{aligned}M_n(\gamma) &= \frac{1}{n}\big\|I_{\le\gamma_0}G_1 + I_{>\gamma_0}G_2\\ &\quad + \mathbf{e} - I_{\le\gamma}X\widehat{\beta}_1(\gamma) - I_{>\gamma}X\widehat{\beta}_2(\gamma)\big\|^2\\ &= \frac{1}{n}\big\{G_1'I_{\le\gamma_0}G_1 + G_2'I_{>\gamma_0}G_2\\ &\quad + \widehat{\beta}_1(\gamma)'X'I_{\le\gamma}X\widehat{\beta}_1(\gamma) + \widehat{\beta}_2(\gamma)'X'I_{>\gamma}X\widehat{\beta}_2(\gamma)\\ &\quad - 2\widehat{\beta}_1(\gamma)'X'I_{\le\gamma}Y - 2\widehat{\beta}_2(\gamma)'X'I_{>\gamma}Y\big\} + \xi(\mathbf{e})\\ &\xrightarrow{p} b_1'E[\mathbf{x}\mathbf{x}']b_1F(\gamma) + b_2(\gamma)'E[\mathbf{x}\mathbf{x}']b_2(\gamma)(1 - F(\gamma))\\ &\quad - 2b_1'E[\mathbf{x}\mathbf{x}']b_1F(\gamma)\\ &\quad - 2b_2(\gamma)'E[\mathbf{x}\mathbf{x}']b_2(\gamma)(1 - F(\gamma)) + C\\ &= C - b_1'E[\mathbf{x}\mathbf{x}']b_1F(\gamma)\\ &\quad - b_2(\gamma)'E[\mathbf{x}\mathbf{x}']b_2(\gamma)(1 - F(\gamma))\\ &\equiv M(\gamma),\end{aligned}$$

where $\xi(\mathbf{e})$ is a function of $\mathbf{e}$ whose probability limit is a constant and does not depend on γ, and C is a constant. Note that

$$\frac{db_2(\gamma)}{d\gamma} = \frac{[1 - F(\gamma_0)]f(\gamma)}{[1 - F(\gamma)]^2}(b_2 - b_1),$$

so

$$\begin{aligned}\frac{dM(\gamma)}{d\gamma}\Big/f(\gamma) &= -b_1'E[\mathbf{x}\mathbf{x}']b_1 + b_2(\gamma)'E[\mathbf{x}\mathbf{x}']b_2(\gamma)\\ &\quad - 2\left[\frac{db_2(\gamma)}{d\gamma}\Big/f(\gamma)\right]'E[\mathbf{x}\mathbf{x}']b_2(\gamma)(1 - F(\gamma))\\ &= -b_1'E[\mathbf{x}\mathbf{x}']b_1 - 2\frac{1 - F(\gamma_0)}{1 - F(\gamma)}(b_2 - b_1)'\\ &\quad \times E[\mathbf{x}\mathbf{x}']\left[b_1 + (b_2 - b_1)\frac{1 - F(\gamma_0)}{1 - F(\gamma)}\right]\\ &\quad + \left[b_1 + (b_2 - b_1)\frac{1 - F(\gamma_0)}{1 - F(\gamma)}\right]'\\ &\quad \times E[\mathbf{x}\mathbf{x}']\left[b_1 + (b_2 - b_1)\frac{1 - F(\gamma_0)}{1 - F(\gamma)}\right]\end{aligned}$$