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
0165-1765/

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

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

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

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

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

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

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, $E[e^2] < \infty$, $E\left[\mathbf{x}\mathbf{x}'\right] > 0$, $E[\mathbf{x}g_1(x)] \ne E[\mathbf{x}g_2(x)]$, 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 $\text{diag}\{1(q_i \le \gamma)\}$ and $\text{diag}\{1(q_i > \gamma)\}$. 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,

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

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

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

$$= -\left(\frac{1-F(\gamma_0)}{1-F(\gamma)}\right)^2 (b_2-b_1)' E[\mathbf{xx}'] (b_2-b_1) < 0.$$

In other words, $M(\gamma)$ is a strictly decreasing function on $\left[\underline{\gamma}, \gamma_0\right]$. Symmetrically, we can show

$$\frac{dM(\gamma)}{d\gamma}\Big/ f(\gamma) = \left(\frac{F(\gamma_0)}{F(\gamma)}\right)^2 (b_2-b_1)' E[\mathbf{xx}'] (b_2-b_1) > 0$$

on $[\gamma_0, \overline{\gamma}]$. In summary, $M(\gamma)$ achieves its minimum at γ_0. By a standard consistency theorem, e.g., Theorem 2.1 of Newey and McFadden (1994), $\widehat{\gamma}$ is consistent. It is interesting to notice that the left and right derivative of $M(\gamma)$ at γ_0 are not the same, so $M(\gamma)$ has a kink at γ_0. ■

We give a few remarks on this theorem. First, the assumptions in the theorem are quite standard; only $E[\mathbf{x}g_1(x)] \neq E[\mathbf{x}g_2(x)]$ need some explanation. This assumption guarantees that $E[\mathbf{xx}']^{-1}E[\mathbf{x}g_1(x)] \neq E[\mathbf{xx}']^{-1}E[\mathbf{x}g_2(x)]$. In other words, there are threshold effects in the least squares estimation. When $\eta_1(x) = \eta_2(x)$, $E[\mathbf{x}g_1(x)] \neq E[\mathbf{x}g_2(x)]$ is equivalent to $E[\mathbf{xx}']\beta_1 \neq E[\mathbf{xx}']\beta_2$ or $\beta_1 \neq \beta_2$; i.e., there are threshold effects in the original model. Second, this consistency result is related to Chong (2003) and Bai et al. (2008) where the authors show that in the structural change model, the LSE of the structural change point is consistent even if the regressors are misspecified, e.g., contaminated by measurement error. We study a different misspecification of regressors—endogeneity. Also, we point out that the key assumption for such a consistency result to hold is that q is independent of other covariates rather than $q \sim U[0, 1]$ as in the structural change model. Third, the result of this theorem can be extended to accommodate q as a regressor and endogeneity in q. Suppose $\mathbf{x} = (1, x', q)'$,[2] and

$$E[\varepsilon|x,q] = [\eta_1(x) + q\delta_1]\,1(q\le\gamma_0) + [\eta_2(x) + q\delta_2]\,1(q>\gamma_0) \neq 0.$$

In other words, the endogeneity is additively separable in x and q and is linear in q. In this case,

$$E[y|x,q] = \left[g_1(x) + q\left(\beta_{1q}+\delta_1\right)\right]1(q\le\gamma_0) + \left[g_2(x) + q\left(\beta_{2q}+\delta_2\right)\right]1(q>\gamma_0),$$

where β_{1q} and β_{2q} are the coefficients of q in the original model. Now, the specification of $E[y|x,q]$ maintains the additively linear structure of q which is required by the least squares estimation (just as argued in Perron and Yamamoto, forthcoming), so the theorem implies that the LSE of γ is still consistent as long as $q \perp x$.

The result of this theorem can also explain why the two-stage least squares (2SLS) estimator of Caner and Hansen (2004) is generally inconsistent (see Section 2.3 of Yu, 2013) and when the 2SLS estimator would be consistent. Suppose in the first stage regression

$$E[\underline{\mathbf{x}}|z,q] = g(z) + q\delta, \tag{5}$$

where $\underline{\mathbf{x}}$ is $\mathbf{x}$ excluding q, z includes the instruments which cover the constant 1, q is assumed to be exogenous so can serve as an instrumental variable, and $g(z)$ and δ are $\dim(\underline{\mathbf{x}}) \times 1$ vectors. If $q \perp z$, then the linear projection of $\underline{\mathbf{x}}$ on (z, q) is

$$E^*[\underline{\mathbf{x}}|z,q] = \Gamma'z + q\delta,$$

where $\Gamma = E\left[zz'\right]E[zg(z)']$. Now,

$$E[y|z,q] = \left[g(z)'\underline{\beta}_1 + q\left(\delta'\underline{\beta}_1 + \beta_{1q}\right)\right]1(q\le\gamma_0) + \left[g(z)'\underline{\beta}_2 + q\left(\delta'\underline{\beta}_2 + \beta_{2q}\right)\right]1(q>\gamma_0),$$

where $\beta_\ell = \left(\underline{\beta}_\ell', \beta_{\ell q}\right)'$, while we linearly regress y on $\Gamma'z + q\delta$ and q. From the discussion above, the 2SLS estimator of γ need not be consistent even in this case since $\Gamma'z + q\delta$ and q are not independent. Nevertheless, if we regress y on $\Gamma'z$ and q, then the 2SLS estimator of γ is still consistent. Roughly speaking, if we regress $\underline{\mathbf{x}}$ only on z in the first stage and assume (5) and $q \perp z$,[3] then the 2SLS estimator of γ is consistent. In the structural change model $E[\underline{\mathbf{x}}|z,q] = g(z)$ and $E^*[\underline{\mathbf{x}}|z,q] = \Gamma'z$, so the 2SLS estimator of γ is consistent. In the counterexample of Yu (2013), $E[\underline{\mathbf{x}}|z,q]$ is not linear in q so the 2SLS estimator of γ is not consistent.

The result on the consistency of the 2SLS estimator has some similarity with that in treatment effect evaluation without unconfoundedness. Specifically, in the triangular system of equations,

$$Y_i = X_i'\beta + D_i\alpha + u_i,$$
$$D_i = g\left(Z_i'\gamma\right) + v_i,$$

we require $E[v_i|Z_i] = 0$ to make sure the *nonlinear* 2SLS estimator of α to be consistent ($E[Z_iv_i] = 0$ is not enough), where D_i is the treatment status, and the nonlinear 2SLS estimator of α is obtained by regressing Y_i on X_i and $\widehat{D}_i \equiv g\left(Z_i'\widehat{\gamma}\right)$ with $\widehat{\gamma}$ being the first stage estimate of γ based on logit or probit. On the other hand, conventional 2SLS estimates using a linear probability model (i.e., $g\left(Z_i'\gamma\right) = Z_i'\gamma$) are consistent irrespective of $E[D_i|Z_i]$ is linear or not. As already observed by Kelejian (1971) and Heckman (1978, pp. 946–947), it is unnecessary to obtain consistent estimators of the parameters of reduced form equations in order to consistently estimate structural equations. In the threshold model, we need to separate out the effect of the threshold variable to apply such a result; this can explain why we require $q \perp z$.

The referee asked about the consistency of the LSE when $E[\varepsilon|x,q] = \eta_1(x)1(q\le\gamma_1) + \eta_2(x)1(q>\gamma_1)$, where without loss of generality, γ_1 is assumed to be greater than γ_0. This corresponds to the case where the endogenous relation changes somewhere at $\gamma_1 \neq \gamma_0$. In this case, y follows a three-regime threshold model,

$$y = g_1(x)1(q\le\gamma_0) + g_2(x)1(\gamma_0<q\le\gamma_1) + g_3(x)1(q>\gamma_1) + e,$$

where $g_1(x) = \mathbf{x}'\beta_1 + \eta_1(x)$, $g_2(x) = \mathbf{x}'\beta_1 + \eta_2(x)$, $g_3(x) = \mathbf{x}'\beta_2 + \eta_2(x)$, and $E[e|x,q] = 0$. For this model, it turns out that the LSE converges to $\arg\min_{\gamma_0,\gamma_1} M(\gamma)$, where $M(\gamma)$ is the probability limit of the least squares objective function. That is, the LSE is estimating one of the two threshold points which minimizes $M(\gamma)$. This result extends Chong (1995) and Bai (1997) from the linear multiple regime model to the nonlinear counterpart. Also, we do not require q follows $U[0, 1]$, but the assumption that $q \perp x$ cannot be relaxed.

Theorem 2. *Suppose $\{w_i\}_{i=1}^n$ are i.i.d., $\gamma_0, \gamma_1 \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$ which is compact, $E[e^2] < \infty$, $E\left[\mathbf{xx}'\right] > 0$, $M(\gamma_0) < M(\gamma_1)$, 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 $\widehat{\gamma} \xrightarrow{p} \gamma_0$.*

Proof. We use the same notations as in the proof of Theorem 1. It is not hard to show that

$$\widehat{\beta}_1(\gamma) \xrightarrow{p} b_1(\gamma) \equiv \begin{cases} b_1, & \text{if } \gamma\le\gamma_0,\\ \dfrac{F(\gamma_0)}{F(\gamma)}b_1 + \dfrac{F(\gamma)-F(\gamma_0)}{F(\gamma)}b_2, & \text{if } \gamma_0<\gamma\le\gamma_1,\\ \dfrac{F(\gamma_0)}{F(\gamma)}b_1 + \dfrac{F(\gamma_1)-F(\gamma_0)}{F(\gamma)}b_2 + \dfrac{F(\gamma)-F(\gamma_1)}{F(\gamma)}b_3, & \text{if } \gamma>\gamma_1, \end{cases}$$

---

[2] In the structural change model, this means that there is a (linear) trending regressor. See Section 5 of Chong (2003) for some discussion on this case.

[3] Actually, we can use the reduced form directly, i.e., regress y directly on z and q, to consistently estimate γ under these assumptions.

and

$$\widehat{\beta}_2(\gamma) \xrightarrow{p} b_2(\gamma) \equiv \begin{cases} \frac{F(\gamma_0)-F(\gamma)}{1-F(\gamma)}b_1 + \frac{F(\gamma_1)-F(\gamma_0)}{1-F(\gamma)}b_2 \\ + \frac{1-F(\gamma_1)}{1-F(\gamma)}b_3, & \text{if } \gamma \le \gamma_0, \\ \frac{F(\gamma_1)-F(\gamma)}{1-F(\gamma)}b_2 + \frac{1-F(\gamma_1)}{1-F(\gamma)}b_3, & \text{if } \gamma_0 < \gamma \le \gamma_1, \\ b_3, & \text{if } \gamma > \gamma_1, \end{cases}$$

where $b_\ell = E[\mathbf{xx}']^{-1}E[\mathbf{x}g_\ell(x)]$, $\ell = 1, 2, 3$. Similarly as in the proof of Theorem 1, we can show

$$M_n(\gamma) \xrightarrow{p} C - F(\gamma)b_1(\gamma)'E[\mathbf{xx}']b_1(\gamma) - (1-F(\gamma))\, b_2(\gamma)'E[\mathbf{xx}']b_2(\gamma) \equiv M(\gamma).$$

Note that

$$\frac{db_1(\gamma)}{d\gamma} = \begin{cases} 0, \quad \text{if } \gamma \le \gamma_0, \\ -\frac{F(\gamma_0)}{F(\gamma)^2}f(\gamma)b_1 + \frac{F(\gamma_0)}{F(\gamma)^2}f(\gamma)b_2 \\ \quad = \frac{F(\gamma_0)}{F(\gamma)^2}f(\gamma)\,(b_2-b_1), \quad \text{if } \gamma_0 < \gamma \le \gamma_1, \\ -\frac{F(\gamma_0)}{F(\gamma)^2}f(\gamma)b_1 - \frac{F(\gamma_1)-F(\gamma_0)}{F(\gamma)^2}f(\gamma)b_2 \\ \quad + \frac{F(\gamma_1)}{F(\gamma)^2}b_3, \quad \text{if } \gamma > \gamma_1, \end{cases}$$

and

$$\frac{db_2(\gamma)}{d\gamma} = \begin{cases} -\frac{1-F(\gamma_0)}{[1-F(\gamma)]^2}f(\gamma)b_1 + \frac{F(\gamma_1)-F(\gamma_0)}{[1-F(\gamma)]^2}f(\gamma)b_2 \\ \quad + \frac{1-F(\gamma_1)}{[1-F(\gamma)]^2}f(\gamma)b_3, \quad \text{if } \gamma \le \gamma_0, \\ -\frac{1-F(\gamma_1)}{[1-F(\gamma)]^2}b_2 + \frac{1-F(\gamma_1)}{[1-F(\gamma)]^2}f(\gamma)b_3 \\ \quad = \frac{1-F(\gamma_1)}{[1-F(\gamma)]^2}f(\gamma)(b_3-b_2), \quad \text{if } \gamma_0 < \gamma \le \gamma_1, \\ 0, \quad \text{if } \gamma > \gamma_1. \end{cases}$$

So if $\gamma \le \gamma_0$,

$$\begin{aligned} \frac{dM(\gamma)}{d\gamma}\Big/ f(\gamma) &= -b_1'E[\mathbf{xx}']b_1 + b_2(\gamma)'E[\mathbf{xx}']b_2(\gamma) \\ &\quad - 2\,(1-F(\gamma))\left[\frac{db_2(\gamma)}{d\gamma}\Big/ f(\gamma)\right]' E[\mathbf{xx}']b_2(\gamma) \\ &= -b_1'E[\mathbf{xx}']b_1 + \frac{\widetilde{b}_2(\gamma)'E[\mathbf{xx}']\widetilde{b}_2(\gamma)}{[1-F(\gamma)]^2} \\ &\quad - 2\frac{\widetilde{b}_2(\gamma)'E[\mathbf{xx}']L_1}{[1-F(\gamma)]^2} \\ &= -\frac{1}{[1-F(\gamma)]^2}L_1'E[\mathbf{xx}']L_1, \end{aligned}$$

where

$$L_1 = (F(\gamma_1)-F(\gamma_0))\, b_2 + (1-F(\gamma_1))b_3 - (1-F(\gamma_0))b_1,$$

and

$$\begin{aligned} \widetilde{b}_2(\gamma) &= (1-F(\gamma))\, b_2(\gamma) = [F(\gamma_0)-F(\gamma)]\, b_1 \\ &\quad + [F(\gamma_1)-F(\gamma_0)]\, b_2 + [1-F(\gamma_1)]\, b_3 \\ &= L_1 + (1-F(\gamma))b_1. \end{aligned}$$

Similarly, if $\gamma > \gamma_1$,

$$\frac{dM(\gamma)}{d\gamma}\Big/ f(\gamma) = \frac{1}{F(\gamma)^2}L_2'E[\mathbf{xx}']L_2,$$

where

$$L_2 = F(\gamma_0)b_1 + [F(\gamma_1)-F(\gamma_0)]\, b_2 - F(\gamma_1)b_3.$$

For $\gamma \in (\gamma_0, \gamma_1)$, $M(\gamma)$ need not be increasing. To show that $\arg\min_{\gamma\in[\gamma_0,\gamma_1]} M(\gamma) = \gamma_0$, we need to show that $M(\gamma_0) < M(\gamma)$ for all $\gamma \in (\gamma_0, \gamma_1)$. Note that

$$\begin{aligned} &M(\gamma) - M(\gamma_0) \\ &= F(\gamma_0)b_1'E[\mathbf{xx}']b_1' + [1-F(\gamma_0)]\, b_2(\gamma_0)'E[\mathbf{xx}']b_2(\gamma_0) \\ &\quad - F(\gamma)b_1(\gamma)'E[\mathbf{xx}']b_1(\gamma) - [1-F(\gamma)]\, b_2(\gamma)'E[\mathbf{xx}']b_2(\gamma) \\ &= F(\gamma_0)b_1'E[\mathbf{xx}']b_1' + [1-F(\gamma_0)] \\ &\quad \times \left[b_2 + \frac{1-F(\gamma_1)}{1-F(\gamma_0)}(b_3-b_2)\right]' \\ &\quad \times E[\mathbf{xx}']\left[b_2 + \frac{1-F(\gamma_1)}{1-F(\gamma_0)}(b_3-b_2)\right] \\ &\quad - F(\gamma)\left[b_2 + \frac{F(\gamma_0)}{F(\gamma)}(b_1-b_2)\right]' \\ &\quad \times E[\mathbf{xx}']\left[b_2 + \frac{F(\gamma_0)}{F(\gamma)}(b_1-b_2)\right] \\ &\quad - [1-F(\gamma)]\left[b_2 + \frac{1-F(\gamma_1)}{1-F(\gamma)}(b_3-b_2)\right]' \\ &\quad \times E[\mathbf{xx}']\left[b_2 + \frac{1-F(\gamma_1)}{1-F(\gamma)}(b_3-b_2)\right] \\ &= \frac{F(\gamma_0)\,[F(\gamma)-F(\gamma_0)]}{F(\gamma)}(b_1-b_2)'E[\mathbf{xx}'] \\ &\quad \times (b_1-b_2) - \frac{[1-F(\gamma_1)]^2\,[F(\gamma)-F(\gamma_0)]}{[1-F(\gamma)]\,[1-F(\gamma_0)]} \\ &\quad \times (b_2-b_3)'E[\mathbf{xx}']\,(b_2-b_3) \\ &> \frac{F(\gamma_0)\,[F(\gamma)-F(\gamma_0)]}{F(\gamma)}(b_1-b_2)'E[\mathbf{xx}'] \\ &\quad \times (b_1-b_2) - \frac{F(\gamma_1)\,[1-F(\gamma_1)]\,[F(\gamma)-F(\gamma_0)]}{F(\gamma)\,[1-F(\gamma_0)]} \\ &\quad \times (b_2-b_3)'E[\mathbf{xx}']\,(b_2-b_3) \\ &= \frac{F(\gamma_1)\,[F(\gamma)-F(\gamma_0)]}{F(\gamma)\,[F(\gamma_1)-F(\gamma_0)]}\left\{\frac{F(\gamma_0)\,[F(\gamma_1)-F(\gamma_0)]}{F(\gamma_1)}\right. \\ &\quad \times (b_1-b_2)'E[\mathbf{xx}']\,(b_1-b_2) - \frac{[F(\gamma_1)-F(\gamma_0)]\,[1-F(\gamma_1)]}{1-F(\gamma_0)} \\ &\quad \left. \times (b_2-b_3)'E[\mathbf{xx}']\,(b_2-b_3)\right\} \\ &= \frac{F(\gamma_1)\,[F(\gamma)-F(\gamma_0)]}{F(\gamma)\,[F(\gamma_1)-F(\gamma_0)]}[M(\gamma_1)-M(\gamma_0)] > 0, \end{aligned}$$

where the first inequality follows from the fact that $\frac{1-F(\gamma_1)}{1-F(\gamma)} < \frac{F(\gamma_1)}{F(\gamma)}$, and the last equality is from

$$\begin{aligned} M(\gamma_1) - M(\gamma_0) &= \frac{F(\gamma_0)\,[F(\gamma_1)-F(\gamma_0)]}{F(\gamma_1)} \\ &\quad \times (b_1-b_2)'E[\mathbf{xx}']\,(b_1-b_2) \\ &\quad - \frac{[F(\gamma_1)-F(\gamma_0)]\,[1-F(\gamma_1)]}{1-F(\gamma_0)} \\ &\quad \times (b_2-b_3)'E[\mathbf{xx}']\,(b_2-b_3), \end{aligned}$$

which is implied by the third equality.

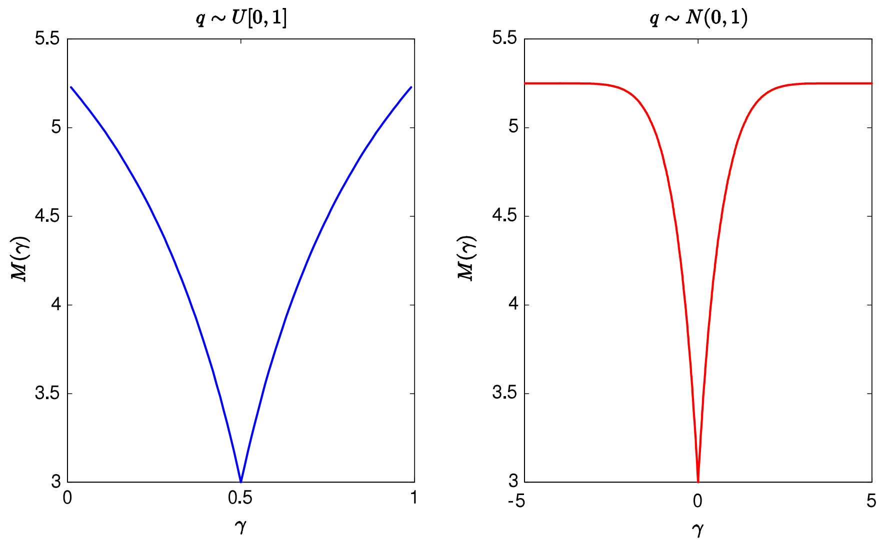

Fig. 1. $M(\gamma)$ when $q \perp x$.

Finally, note that $L_1 = 0$ implies $M(\gamma_0) > M(\gamma_1)$ which is excluded by our assumption. So $M(\gamma)$ is strictly decreasing on $[\underline{\gamma}, \gamma_0]$. Although $M(\gamma_0) < M(\gamma_1)$ does not exclude $L_2 = 0$, given that $M(\gamma_0) < M(\gamma_1)$, $\arg\min_{\gamma\in\Gamma} M(\gamma) = \gamma_0$ even if $M(\gamma)$ is flat on $[\gamma_1, \overline{\gamma}]$. ■

For illustration of the consistency of the LSE of γ, consider the following simple example. Suppose in (4), $g_\ell(x) = x^3\beta_{\ell 0}$ with $\beta_{10} = 0$ and $\beta_{20} = 1$, $x \sim N(0, 1)$, $e \sim N(0, 1)$, $x \perp e$ and $q \perp (x, e)$. We consider two cases of q's distribution. In case one, $q \sim U(0, 1)$, which corresponds to the structural change model, and in case two, $q \sim N(0, 1)$. Although $E[y|x, q]$ is a nonlinear function of x, we fit a linear function in each regime. It can be shown that in case one, the probability limit of $M_n(\gamma)$,

$$M(\gamma) = \begin{cases} 15\left(\gamma_0\beta_{10}^2 + (1-\gamma_0)\beta_{20}^2\right) \\ \quad -9\left(\gamma\beta_{10}^2 + (1-\gamma)\beta_2^2(\gamma)\right), & \text{if } \gamma \le \gamma_0, \\ 15\left(\gamma_0\beta_{10}^2 + (1-\gamma_0)\beta_{20}^2\right) \\ \quad -9\left(\gamma\beta_1^2(\gamma) + (1-\gamma)\beta_{20}^2\right) & \text{if } \gamma > \gamma_0, \end{cases}$$

where $\beta_1(\gamma) = \beta_{20} + (\beta_{10} - \beta_{20})\frac{\gamma_0}{\gamma}$, $\beta_2(\gamma) = \beta_{10} + (\beta_{20} - \beta_{10})\frac{1-\gamma_0}{1-\gamma}$, and constants are omitted; in case two,

$$\begin{aligned} M(\gamma) = {} & 15\beta_{10}^2\Phi(\gamma_0) + 15\beta_{20}^2(1-\Phi(\gamma_0)) + \beta_1^2(\gamma)\Phi(\gamma) \\ & + \beta_2^2(\gamma)(1-\Phi(\gamma)) \\ & -2\begin{cases} 3\beta_{10}\beta_1(\gamma)\Phi(\gamma) + 3\beta_{10}\beta_2(\gamma)[\Phi(\gamma_0) \\ \quad -\Phi(\gamma)] + 3\beta_{20}\beta_2(\gamma)(1-\Phi(\gamma_0)), \\ \quad \text{if } \gamma \le \gamma_0, \\ 3\beta_{10}\beta_1(\gamma)\Phi(\gamma_0) + 3\beta_{20}\beta_1(\gamma)[\Phi(\gamma) \\ \quad -\Phi(\gamma_0)] + 3\beta_{20}\beta_2(\gamma)(1-\Phi(\gamma)), \\ \quad \text{if } \gamma > \gamma_0, \end{cases} \end{aligned}$$

where

$$\begin{aligned} \beta_1(\gamma) &= \text{plim}\left(\widehat{\beta}_1(\gamma)\right) \\ &= 3\begin{cases} \beta_{10}, & \text{if } \gamma \le \gamma_0, \\ \beta_{10}\dfrac{\Phi(\gamma_0)}{\Phi(\gamma)} + \beta_{20}\dfrac{\Phi(\gamma)-\Phi(\gamma_0)}{\Phi(\gamma)}, & \text{if } \gamma > \gamma_0, \end{cases} \end{aligned}$$

and

$$\begin{aligned} \beta_2(\gamma) &= \text{plim}\left(\widehat{\beta}_2(\gamma)\right) \\ &= 3\begin{cases} \beta_{10}\dfrac{\Phi(\gamma_0)-\Phi(\gamma)}{1-\Phi(\gamma)} + \beta_{20}\dfrac{1-\Phi(\gamma_0)}{1-\Phi(\gamma)}, & \text{if } \gamma \le \gamma_0, \\ \beta_{20}, & \text{if } \gamma > \gamma_0. \end{cases} \end{aligned}$$

Fig. 1 shows the two $M(\gamma)$ functions when $\gamma_0 = 0.5$ and $\gamma_0 = 0$ respectively. Obviously, $\arg\min_\gamma M(\gamma) = \gamma_0$.

2. Inconsistency of the LSE when $q \not\perp x$

To show that the LSE of γ is inconsistent when q is not independent of other covariates, we need only provide a counterexample. Continue to consider the example at the end of the last section, but now assume $x = q$ (i.e., there is perfect correlation between q and x). This example can also serve as a counterexample where $E[\varepsilon|x, q]$ is nonlinear in q. It can be shown that in this case

$$\begin{aligned} M(\gamma) = {} & \beta_{10}^2\left[15\Phi(\gamma_0) - \left(\gamma_0^5 + 5\gamma_0^3 + 15\gamma_0\right)\phi(\gamma_0)\right] \\ & + \beta_{20}^2\left[15(1-\Phi(\gamma_0)) + \left(\gamma_0^5 + 5\gamma_0^3 + 15\gamma_0\right)\phi(\gamma_0)\right] \\ & + \beta_1^2(\gamma)[\Phi(\gamma) - \gamma\phi(\gamma)] \\ & + \beta_2^2(\gamma)[1 - \Phi(\gamma) + \gamma\phi(\gamma)] \\ & -2\begin{cases} \beta_{10}\beta_1(\gamma)\left[3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma)\right] \\ \quad + \beta_{10}\beta_2(\gamma)\left[3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)\right. \\ \quad \left. - \left(3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma)\right)\right] \\ \quad + \beta_{20}\beta_2(\gamma)\left[3(1-\Phi(\gamma_0))\right. \\ \quad \left. + \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)\right], \quad \text{if } \gamma \le \gamma_0, \\ \beta_{10}\beta_1(\gamma)\left[3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)\right] \\ \quad + \beta_{20}\beta_1(\gamma)\left[3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma)\right. \\ \quad \left. - \left(3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)\right)\right] \\ \quad + \beta_{20}\beta_2(\gamma)\left[3(1-\Phi(\gamma))\right. \\ \quad \left. + \left(\gamma^3 + 3\gamma\right)\phi(\gamma)\right], \quad \text{if } \gamma > \gamma_0, \end{cases} \end{aligned}$$

where

$$\begin{aligned} \beta_1(\gamma) &= \text{plim}\left(\widehat{\beta}_1(\gamma)\right) \\ &= \begin{cases} \beta_{10}\dfrac{3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma)}{\Phi(\gamma) - \gamma\phi(\gamma)}, \quad \text{if } \gamma \le \gamma_0, \\ \beta_{10}\dfrac{3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)}{\Phi(\gamma) - \gamma\phi(\gamma)} \\ \quad + \beta_{20}\dfrac{3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma) - \left[3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)\right]}{\Phi(\gamma) - \gamma\phi(\gamma)}, \\ \quad \text{if } \gamma > \gamma_0, \end{cases} \end{aligned}$$

and

$$\beta_2(\gamma) = \text{plim}\left(\widehat{\beta}_2(\gamma)\right)$$

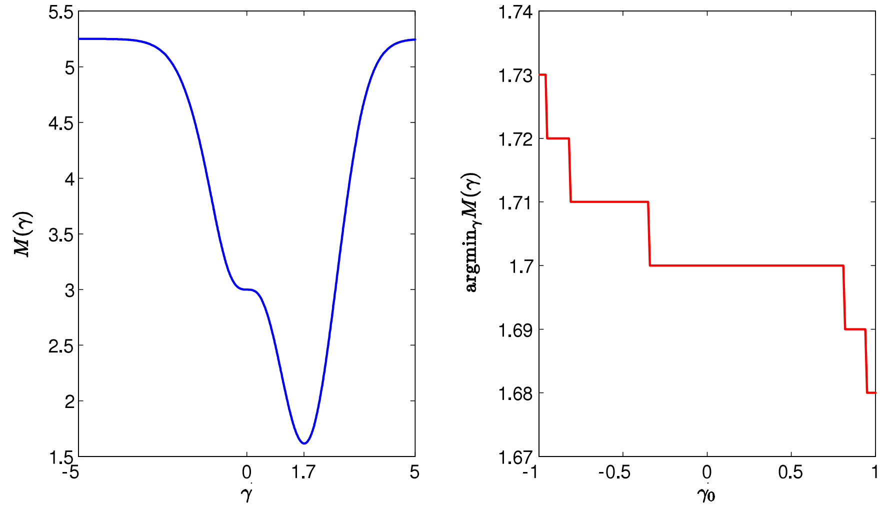

Fig. 2. Inconsistency of the LSE of γ when $q \not\perp x$.

$$= \begin{cases} \beta_{10}\dfrac{3\Phi(\gamma_0) - \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0) - \left[3\Phi(\gamma) - \left(\gamma^3 + 3\gamma\right)\phi(\gamma)\right]}{1 - \Phi(\gamma) + \gamma\phi(\gamma)} \\ \quad + \beta_{20}\dfrac{3\left(1 - \Phi(\gamma_0)\right) + \left(\gamma_0^3 + 3\gamma_0\right)\phi(\gamma_0)}{1 - \Phi(\gamma) + \gamma\phi(\gamma)}, & \text{if } \gamma \le \gamma_0, \\ \beta_{20}\dfrac{3\left(1 - \Phi(\gamma)\right) + \left(\gamma^3 + 3\gamma\right)\phi(\gamma)}{1 - \Phi(\gamma) + \gamma\phi(\gamma)}, & \text{if } \gamma > \gamma_0. \end{cases}$$

The left panel of Fig. 2 shows $M(\gamma)$ when $\gamma_0 = 0$. Obviously, $\arg\min_\gamma M(\gamma) \neq \gamma_0$. The right panel of Fig. 2 shows $\arg\min_\gamma M(\gamma)$ for $\gamma_0 \in [-1, 1]$. For all γ_0's in this interval, $\arg\min_\gamma M(\gamma) \neq \gamma_0$.

More generally, if $E[\varepsilon|x, q] = g(x, q)$ which is not additively linear in q and/or $q \not\perp x$, the LSE of γ is not consistent.

3. Conclusion

We conclude this note by summarizing the consistency results of the LSE based on (2) in different cases. Since endogeneity essentially affects $m(x, q) \equiv E[y|x, q]$, we will consider the identifiability of γ_0 by the LSE in the model $y = m(x, q) + e$ with $E[e|x, q] = 0$ for different specifications of $m(x, q)$. Denote $M(\gamma)$ as the probability limit of the objective function of the LSE as in Theorems 1 and 2. Let $m + 1$ be the number of regimes; when $m = 1$, we get the usual two-regime model. (i) if $m(x, q) = \sum_{j=1}^{m+1} \mathbf{x}'\beta_j 1\left(\gamma_{j-1} < q \le \gamma_j\right)$, where $m \ge 1$, $\mathbf{x} = (1, x', q)'$, $\gamma_0 = -\infty$, and $\gamma_{m+1} = \infty$, then $\text{plim}(\widehat{\gamma}) = \arg\min_{\gamma_1,\ldots,\gamma_m} M(\gamma)$ regardless of $q \perp x$ or not; see, e.g., Gonzalo and Pitarakis (2002). (ii) if $m(x, q) = \sum_{j=1}^{m+1}\left[g_j(x) + q\beta_j\right] 1\left(\gamma_{j-1} < q \le \gamma_j\right)$, where $q \perp x$ and $g_j(x)$ is a generically nonlinear function of x, then $\text{plim}(\widehat{\gamma}) = \arg\min_{\gamma_1,\ldots,\gamma_m} M(\gamma)$. (iii) if $m(x, q) = \sum_{j=1}^{m+1}\left[g_j(x) + q\beta_j\right] 1\left(\gamma_{j-1} < q \le \gamma_j\right)$ with $q \not\perp x$ or $m(x, q) = \sum_{j=1}^{m+1} g_j(x, q) 1\left(\gamma_{j-1} < q \le \gamma_j\right)$ with q being not separably additive in $g_j(x, q)$ (regardless of $q \perp x$ or not), then $\text{plim}(\widehat{\gamma})$ generically does not converge to $\arg\min_{\gamma_1,\ldots,\gamma_m} M(\gamma)$.

The main application of the results in this note is the structural change model with endogeneity. Another possible application is regression discontinuity designs. Suppose the response variable y follows (4), and the treatment status is determined by whether q is greater than a threshold γ_0, where q is randomly designed so is independent of other covariates; then γ_0 or the treatment status can be identified by the LSE. The key point of this note is that in some special cases, the threshold point can be identified by the "parametric" LSE without employing any instrument. In the general case, Yu and Phillips (2014) show that the threshold point can be actually identified by a form of nonparametric LSE without instruments.

References

Bai, J., 1997. Estimating multiple breaks one at a time. Econometric Theory 13, 315–352.
Bai, J., Chen, H., Chong, T.T.-L., Wang, S.X., 2008. Generic consistency of the break-point estimators under specification errors in a multiple-break model. Econom. J. 11, 287–307.
Caner, M., Hansen, B.E., 2004. Instrumental variable estimation of a threshold model. Econometric Theory 20, 813–843.
Chong, T.T.-L., 1995. Partial parameter consistency in a misspecified structural change model. Econom. Lett. 49, 351–357.
Chong, T.T.-L., 2003. Generic consistency of the break-point estimator under specification errors. Econom. J. 6, 167–192.
Gonzalo, J., Pitarakis, J.-Y., 2002. Estimation and model selection based inference in single and multiple threshold models. J. Econometrics 110, 319–352.
Heckman, J.J., 1978. Dummy endogenous variables in a simultaneous equations system. Econometrica 46, 931–960.
Kelejian, H.H., 1971. Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. J. Amer. Statist. Assoc. 66, 373–374.
Newey, W.K., McFadden, D.L., 1994. Large sample estimation and hypothesis testing. In: Eagle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. Vol. 4. Elsevier Science B.V., pp. 2113–2245 (Chapter 36).
Perron, P., Qu, Z., 2006. Estimating restricted structural change models. J. Econometrics 134, 373–399.
Perron, P., Yamamoto, Y., 2013. Using OLS to estimate and test for structural changes in models with endogenous regressors. J. Appl. Econometrics (forthcoming).
Yu, P., 2013. Inconsistency of 2SLS estimators in threshold regression with endogeneity. Econom. Lett. 120, 532–536.
Yu, P., Phillips, P.C.B., 2014. Threshold regression with endogeneity, Mimeo, HKU.