



Threshold regression asymptotics: From the compound Poisson process to two-sided Brownian motion

Ping Yu^{a,*}, Peter C.B. Phillips^{b,c,d,e,1}

^a The University of Hong Kong, Hong Kong

^b Yale University, USA

^c University of Auckland, New Zealand

^d University of Southampton, United Kingdom

^e Singapore Management University, Singapore



HIGHLIGHTS

- We show asymptotic equivalence between joint asymptotics and sequential asymptotics in threshold regression.
- We show how compound Poisson process can be approximated by two-sided Brownian motion.
- We show randomness in the number of summands of compound Poisson process disappears in approximation.

ARTICLE INFO

Article history:

Received 9 June 2018

Received in revised form 24 August 2018

Accepted 28 August 2018

Available online xxx

JEL classification:

C24

Keywords:

Threshold regression

Sequential asymptotics

Doob's martingale inequality

Compound Poisson process

Brownian motion

ABSTRACT

The asymptotic distribution of the least squares estimator in threshold regression is expressed in terms of a compound Poisson process when the threshold effect is fixed and as a functional of two-sided Brownian motion when the threshold effect shrinks to zero. This paper explains the relationship between this dual limit theory by showing how the asymptotic forms are linked in terms of joint and sequential limits. In one case, joint asymptotics apply when both the sample size diverges and the threshold effect shrinks to zero, whereas sequential asymptotics operate in the other case in which the sample size diverges first and the threshold effect shrinks subsequently. The two operations lead to the same limit distribution, thereby linking the two different cases. The proofs make use of ideas involving limit theory for sums of a random number of summands.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Threshold regression (TR) is an important statistical model that has been influential in many fields. There are extensive applications in economics and Hansen (2011) provides a summary of the empirical literature. The typical setup has the following form

$$y = \begin{cases} x' \beta_1 + u_1, & q \leq \gamma; \\ x' \beta_2 + u_2, & q > \gamma; \end{cases} \quad (1)$$

* Correspondence to: Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail addresses: pingyu@hku.hk (P. Yu), peter.phillips@yale.edu (P.C.B. Phillips).

¹ Cowles Foundation for Research in Economics, Yale University, POBox 208281, New Haven, CT, USA.

where u_ℓ satisfies $\mathbb{E}[u_\ell | x, q] = 0$ and may be conditionally heteroskedastic over the two regimes $\ell = 1, 2$,² the variable q governs the threshold trigger γ that splits the sample and q has density $f_q(\cdot)$ and distribution $F_q(\cdot)$, the regressor $x \in \mathbb{R}^k$ may include q as a covariate, and $\beta := (\beta'_1, \beta'_2)' \in \mathbb{R}^{2k}$ is the coefficient vector covering the two regimes. The setup is similar to simple linear regression except that the slope coefficients depend on whether the threshold variable q crosses the threshold point γ . The parameter γ is often of primary interest in applications.

Under the conditional mean independence assumption $\mathbb{E}[u_\ell | x, q] = 0$, the threshold parameter γ can be estimated by nonlinear

² The symbol ℓ is used to indicate the two regimes in (1) and, to simplify notation in what follows, the explicit values " $\ell = 1, 2$ " are often omitted.

least squares regression giving the least squares estimator (LSE)

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} M_n(\gamma),$$

where Γ is the parameter space of γ , which is assumed to be a proper subset of the support of q , the criterion function is

$$M_n(\gamma) := \min_{\beta_1, \beta_2} \sum_{i=1}^n (y_i - x'_i \beta_1 1(q_i \leq \gamma) - x'_i \beta_2 1(q_i > \gamma))^2,$$

and $1(\cdot)$ is the indicator function. Optimization of $M_n(\gamma)$ typically leads to an interval estimate of γ . Common practice in the literature on threshold regression employs the left-endpoint LSE (LLSE) to resolve this uncertainty, although Yu (2012, 2015) has recently shown that the middle-point LSE (MLSE) is more efficient in most cases. The precise definition of the $\arg \min_{\gamma}$ operation or the particular choice (LLSE or MLSE) of practical implementation of the regression estimator $\hat{\gamma}$ do not affect any of the results in this paper.

Two approaches have been proposed for inference about γ in the TR model (1). The first is the fixed-threshold-effect framework of Chan (1993) where the break differential $\delta_0 := \beta_{10} - \beta_{20}$ is taken as fixed and where we use the zero subscript to indicate true value. In this framework, $\hat{\gamma}$ is n -consistent, and

$$n(\hat{\gamma} - \gamma_0) \xrightarrow{d} \arg \min_v D(v), \tag{2}$$

where

$$D(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, & \text{if } v \leq 0, \\ \sum_{i=1}^{N_2(v)} z_{2i}, & \text{if } v > 0, \end{cases} \tag{3}$$

$z_{\ell i}$ has an absolutely continuous distribution, $N_{\ell}(\cdot)$ is a Poisson process with intensity $f_q(\gamma_0)$, and $\{z_{1i}, z_{2i}\}_{i \geq 1}$, $N_1(\cdot)$ and $N_2(\cdot)$ are independent of each other. Define the variables

$$\begin{aligned} \bar{z}_{1i} &:= 2x'_i \delta_0 u_{1i} + \delta'_0 x_i x'_i \delta_0, \\ \bar{z}_{2i} &:= -2x'_i \delta_0 u_{2i} + \delta'_0 x_i x'_i \delta_0, \end{aligned}$$

where \bar{z}_{1i} represents the effect on $M_n(\gamma) - M_n(\gamma_0)$ when γ is displaced on the left of γ_0 , and \bar{z}_{2i} represents the converse case. Then $z_{1i} = \lim_{\Delta \uparrow 0} \bar{z}_{1i} 1\{\gamma_0 + \Delta < q_i \leq \gamma_0\}$ is the limiting conditional value of \bar{z}_{1i} given $\gamma_0 + \Delta < q_i \leq \gamma_0$, $\Delta < 0$ with $\Delta \uparrow 0$, and $z_{2i} = \lim_{\Delta \downarrow 0} \bar{z}_{2i} 1\{\gamma_0 < q_i \leq \gamma_0 + \Delta\}$ is the limiting conditional value of \bar{z}_{2i} given $\gamma_0 < q_i \leq \gamma_0 + \Delta$, $\Delta > 0$ with $\Delta \downarrow 0$. It follows that the density of the quantity $z_{\ell i}$ is $f_{\bar{z}_{\ell}, q}(z_{\ell}, \gamma_0) / f_q(\gamma_0)$, the conditional density of \bar{z}_{ℓ} given $q = \gamma_0$. In this framework, the asymptotic distribution of $\hat{\gamma}$ is given as the argmin of the compound Poisson process $D(v)$ in (3).

The second approach is the shrinking-threshold-effect framework of Hansen (2000) which is borrowed from the structural change literature such as Picard (1985) and Bai (1997), where the break differential δ_0 shrinks to zero as $n \rightarrow \infty$ and is therefore denoted by δ_n . As long as $\|\delta_n\| \rightarrow 0$ and $\sqrt{n} \|\delta_n\| \rightarrow \infty$ (i.e., δ_n does not fall in a contiguous neighborhood of the unidentified case $\delta_n = 0$, or in other words, there is sufficient identification information asymptotically in the sample data), then $\hat{\gamma}$ is consistent with the convergence rate $a_n := n \|\delta_n\|^2$, and

$$a_n(\hat{\gamma} - \gamma_0) \xrightarrow{d} \arg \min_v C(v), \tag{4}$$

where

$$C(v) = \begin{cases} 2\sqrt{f_q(\gamma_0)} \Omega_1 W_1(|v|) + f_q(\gamma_0) Q |v|, & \text{if } v \leq 0, \\ 2\sqrt{f_q(\gamma_0)} \Omega_2 W_2(|v|) + f_q(\gamma_0) Q |v|, & \text{if } v > 0, \end{cases} \tag{5}$$

with $Q = \lim_{n \rightarrow \infty} \frac{\delta'_n \mathbb{E}[xx' | q = \gamma_0] \delta_n}{\delta'_n \delta_n}$, $\Omega_{\ell} = \lim_{n \rightarrow \infty} \frac{\delta'_n \mathbb{E}[xxu^2 | q = \gamma_0] \delta_n}{\delta'_n \delta_n}$, and the pair $\{W_{\ell}(v), \ell = 1, 2\}$ being two independent standard Brownian motions defined on $[0, \infty)$. In this framework, the asymptotic

distribution of $\hat{\gamma}$ is given as the argmin of the drifted two-sided Brownian motion $C(v)$ in (5) with different scale parameters in the two directions.

An interesting question that emerges from these two different asymptotic distributions of $\hat{\gamma}$ is how they are related, given that they both arise from the same statistical problem. In particular, why and how does the argmin of a compound Poisson process transition to the argmin of a two-sided Brownian motion as the parameter δ_0 changes from being treated as ‘fixed’ to one that ‘shrinks to zero’. The goal of the present paper is to provide the connection between the two limit theories.

2. Two asymptotic distributions and their connection

This section provides some background on the two different limit forms $D(\cdot)$ and $C(\cdot)$ and some intuition on how they determine the asymptotic distributions of $\hat{\gamma}$ and influence the different convergence rates. From Yu (2014), we have the finite sample formulation

$$n(\hat{\gamma} - \gamma_0) = \arg \min_v D_n(v) + o_p(1), \tag{6}$$

where

$$\begin{aligned} D_n(v) &= \sum_{i=1}^n \bar{z}_{1i} 1\left(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0\right) \\ &\quad + \sum_{i=1}^n \bar{z}_{2i} 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n}\right). \end{aligned}$$

From Hansen (2000), we have the alternate formulation

$$a_n(\hat{\gamma} - \gamma_0) = \arg \min_v C_n(v) + o_p(1), \tag{7}$$

where

$$\begin{aligned} C_n(v) &= \sum_{i=1}^n \bar{z}_{1i} 1\left(\gamma_0 + \frac{v}{n \|\delta_n\|^2} < q_i \leq \gamma_0\right) \\ &\quad + \sum_{i=1}^n \bar{z}_{2i} 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n \|\delta_n\|^2}\right). \end{aligned}$$

Note from these criteria that in estimating γ , we may effectively assume that the parameter vector β is known. The reason is that estimation of γ involves only local information around the threshold value γ_0 while estimation of β involves global information and these two components of the information set are independent – see Yu (2012, 2015).

The difference between the criteria $D_n(\cdot)$ and $C_n(\cdot)$ is that the localizing parameter v in $D_n(\cdot)$ is standardized to $v / \|\delta_n\|^2$ in $C_n(\cdot)$, taking account of the shrinking differential δ_n . As a result, we may write (7) as $\arg \min_v C_n(v) = \|\delta_n\|^2 \arg \min_v D_n(v)$. This re-standardization relating the criteria explains why the convergence rate of $\hat{\gamma}$ changes from n to $a_n = n \|\delta_n\|^2$ in moving from (6) to (7).

To understand the limit theory in which $D_n(\cdot)$ converges to $D(\cdot)$, we may rewrite $D_n(\cdot)$ as

$$D_n(v) = \begin{cases} \sum_{i=1}^{N_{1n}(|v|)} \bar{z}_{1i}, & \text{if } v \leq 0, \\ \sum_{i=1}^{N_{2n}(v)} \bar{z}_{2i}, & \text{if } v > 0, \end{cases}$$

where $N_{1n}(|v|) = \sum_{i=1}^n 1(\gamma_0 + \frac{v}{n} < q_i \leq \gamma_0)$ and $N_{2n}(v) = \sum_{i=1}^n 1(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n})$. Note that $N_{\ell n}(\cdot)$ is a binomial process. For example, for any given $v > 0$,

$$N_{2n}(v) \sim \text{Bin}(n, p_n(v)),$$

with $p_n(v) = F_q(\gamma_0 + \frac{v}{n}) - F_q(\gamma_0)$, and for any $v_2 > v_1 > 0$, the increment $N_{2n}(v_2) - N_{2n}(v_1)$ is independent of $N_{2n}(v_1)$. It is well known that a binomial process will converge to a Poisson

process if $np_n(v)$ converges to a finite number for any v . In our case, $np_n(v) \rightarrow f_q(\gamma_0)v$, where $f_q(\cdot)$ is assumed to be continuous, positive and finite in a neighborhood of γ_0 .³ This explains why $D_n(\cdot) \rightsquigarrow D(\cdot)$, where we use \rightsquigarrow to indicate weak convergence of a stochastic process on the associated probability space.

The reason why the variable $\bar{z}_{\ell i}$ changes to $z_{\ell i}$ is because only those q_i that are local to γ_0 are involved in $D_n(\cdot)$. In $D_n(\cdot)$, $N_{\ell n}(\cdot)$ and $\bar{z}_{\ell i}$ are correlated through the correlation between q_i and $\bar{z}_{\ell i}$. But for two random variables (X, Y) , their joint density $f(X, Y) = f(Y|X)f(X)$ factors so that Y given X is statistically independent of X although $f(Y|X)$ may still be functionally dependent on X – see footnote 10 of Heckman (1997). This explains why $z_{\ell i}$, as a conditional random variable given q_i , is independent of $N_{\ell}(\cdot)$.

In contrast to the case of $D_n(\cdot)$, observe that $n\left(F_q\left(\gamma_0 + \frac{v}{n\|\delta_n\|^2}\right) - F_q(\gamma_0)\right) \approx f_q(\gamma_0) \frac{v}{\|\delta_n\|^2} \rightarrow \infty$, as $\|\delta_n\| \rightarrow 0$, so that infinitely many q_i 's are involved asymptotically in the local neighborhood of γ_0 for any given $v > 0$. As a result, a central limit theorem can be applied for a given v to the sums involved in $C_n(\cdot)$ and under tightness a functional law gives rise to (5). It is not hard to show the following moment limits

$$\mathbb{E}[C_n(v)] = n\mathbb{E}\left[\delta'_n x_i x'_i \delta_n \cdot 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n\|\delta_n\|^2}\right)\right] \rightarrow Q \cdot f_q(\gamma_0)v,$$

and

$$\mathbb{V}(C_n(v)) \approx n\mathbb{V}\left(-2x'_i \delta_n u_{2i} \cdot 1\left(\gamma_0 < q_i \leq \gamma_0 + \frac{v}{n\|\delta_n\|^2}\right)\right) \rightarrow 4\Omega_2 \cdot f_q(\gamma_0)v.$$

Then, because $C_n(\cdot)$ is an independent increments process, it follows by standard empirical process methods that $C_n(\cdot)$ converges to the process $C(\cdot)$ in (5).

3. Sequential asymptotics

Given the different limits described above, the question studied in the present paper is the mechanism by which the convergence $n(\hat{\gamma} - \gamma_0) = \arg \min_v D_n(v) + o_p(1) \xrightarrow{d} \arg \min_v D(v)$ changes to $n\|\delta_n\|^2(\hat{\gamma} - \gamma_0) = \arg \min_v C_n(v) + o_p(1) \xrightarrow{d} \arg \min_v C(v)$ when $\|\delta_n\| \rightarrow 0$. Note that $C_n(v) = D_n\left(\frac{v}{\|\delta_n\|^2}\right)$, so the second limit distribution follows by a joint asymptotic argument in which both $n \rightarrow \infty$ and $\|\delta_n\| \rightarrow 0$. On the other hand, the first limit distribution is obtained by letting $n \rightarrow \infty$ with δ_0 fixed. Given that $\arg \min_v D\left(\frac{v}{\|\delta_0\|^2}\right) = \|\delta_0\|^2 \arg \min_v D(v)$ it follows that $n\|\delta_0\|^2(\hat{\gamma} - \gamma_0)$ is approximately $\arg \min_v D\left(\frac{v}{\|\delta_0\|^2}\right)$ in the first framework. We therefore seek to explain why $\arg \min_v D\left(\frac{v}{\|\delta_0\|^2}\right) \xrightarrow{d} \arg \min_v C(v)$ as $\|\delta_0\| \rightarrow 0$. In other words, the question involves explaining the mechanism by which sequential asymptotics generates the same asymptotic distribution as joint asymptotics, viz., why letting $n \rightarrow \infty$ and $\|\delta_n\| \rightarrow 0$ jointly generates the same limit distribution of $\hat{\gamma}$ as first letting $n \rightarrow \infty$ and then letting $\|\delta_0\| \rightarrow 0$.⁴ The relevant processes are simply represented in the following diagram:

| | | |
|---|--|---|
| $D_n(v)$ | $\xrightarrow{n \rightarrow \infty}$ | $D(v)$ |
| $\downarrow v = \frac{v}{\ \delta_n\ ^2}$ | | $? \downarrow v = \frac{v}{\ \delta_0\ ^2}, \ \delta_0\ \rightarrow 0$ |
| $C_n(v)$ | $\xrightarrow{\ \delta_n\ \rightarrow 0}$ | $C(v)$ |

Problems of joint and sequential limit theory have been addressed before in weak convergence theory (Billingsley, 1968, Theorem 4.2), in panel data asymptotics (Phillips and Moon, 1999) and in near-unit root limit theory (Phillips, 1987; Chan and Wei, 1987; Bykhovskaya and Phillips, 2018). The latter work, which bears some similarity to the present context in terms of taking limits to the boundary of the domain of definition of the parameters, deals with the first order autoregression (AR(1)) model $\{y_t = ay_t + u_t; t = 1, 2, \dots, n\}$ with parameter $a = e^{c/n}$, $-\infty < c \leq 0$.⁵ When a is fixed, or c diverges to negative infinity at the rate n so that $c \sim c^*n$ for some fixed $c^* < 0$ then the model is a stable, or asymptotically stable AR(1) model. In this case, we have joint asymptotics where both $|c|$ and n diverge to ∞ , and the asymptotic distribution of the LSE \hat{a} of a is normal. When $n \rightarrow \infty$ and the localizing parameter $c \leq 0$ is fixed, the model has a root that is local to unity since $a \rightarrow 1$, and the asymptotic distribution of \hat{a} involves quadratic functionals of a diffusion process. Theorem 2 of Phillips (1987) shows that the sequential limit distribution of \hat{a} (after normalization and centering) by first letting $n \rightarrow \infty$ and then letting $c \rightarrow -\infty$ is the same as the joint limit distribution obtained by letting $n \rightarrow \infty$ and $c \rightarrow -\infty$ at the simultaneous rate n .

The following theorem provides a rigorous statement of the corresponding asymptotic distributional equivalence that obtains in the threshold regression model.

Theorem 1. If (i) $\frac{\delta_0 \mathbb{E}[xx'|q=\gamma_0]\delta_0}{\delta_0' \delta_0} \rightarrow Q$ and $\frac{\delta_0 \mathbb{E}[xxu_\ell^2|q=\gamma_0]\delta_0}{\delta_0' \delta_0} \rightarrow \Omega_\ell$ as $\|\delta_0\| \rightarrow 0$, $0 < Q < \infty$ and $0 < \Omega_\ell < \infty$; (ii) $0 < f \leq f_q(\gamma_0) \leq \bar{f} < \infty$; (iii) the conditional distributions of $u_\ell|_{x,q=\gamma}$ and $x|_{q=\gamma}$ are continuous at $q = \gamma_0$, $\mathbb{E}[u_\ell|x, q = \gamma_0] = 0$ and $\mathbb{E}[\|x\|^4 | q = \gamma_0] < \infty$, then

$$\hat{v} := \arg \min_v D\left(\frac{v}{\|\delta_0\|^2}\right) \xrightarrow{d} \arg \min_v C(v) \text{ as } \|\delta_0\| \rightarrow 0, \quad (8)$$

where $D(\cdot)$ and $C(\cdot)$ are defined in (3) and (5).

Proof. We apply Theorem 2.7 of Kim and Pollard (1990) to derive this result. We need to check two conditions for that theorem to apply: (i) $\hat{v} = O_p(1)$ as $\|\delta_0\| \rightarrow 0$; and (ii) $D\left(\frac{v}{\|\delta_0\|^2}\right) \rightsquigarrow C(v) \in \mathbf{C}_{\min}(\mathbb{R})$ as $\|\delta_0\| \rightarrow 0$, where $\mathbf{C}_{\min}(\mathbb{R})$ is defined as the subset of continuous functions $g(\cdot) \in \mathbf{B}_{\text{loc}}(\mathbb{R})$ for which (a) $g(t) \rightarrow \infty$ as $|t| \rightarrow \infty$ and (b) $g(t)$ achieves its minimum at a unique point in \mathbb{R} , and $\mathbf{B}_{\text{loc}}(\mathbb{R})$ is the space of all locally bounded real functions on \mathbb{R} endowed with the uniform metric on compacta. Two lemmas in the Appendix are used to establish these conditions. The shelling method (see, e.g., Theorem 3.2.5 of Van der Vaart and Wellner, 1996) is used in Lemma 1 of the Appendix to prove condition (i); and we apply Theorem 2.3 of Kim and Pollard in Lemma 2 to prove condition (ii). These two lemmas then give the convergence result (8) stated in the theorem. ■

An important and novel feature in the above theorem is that as distinct from usual limiting objective functions that arise in extremum estimation problems, the limit function $D(\cdot)$ involves a

³ See Yu and Zhao (2013) for cases where the density $f_q(\cdot)$ need not be continuous at γ_0 and need not be bounded below by zero and bounded above by a positive constant in a neighborhood of γ_0 .

⁴ As suggested by the referee, in the sequential asymptotics, δ_0 in $D(\cdot)$ does not depend on n because n has already diverged to ∞ , so we maintain the notation δ_0 in the sequential asymptotics.

⁵ The case where $c > 0$ and $c \rightarrow \infty$ at the upper boundary of the domain of definition is also considered in Phillips (1987). Bykhovskaya and Phillips (2018) consider cases where $c(\cdot) = cg(\cdot)$ is a function and the scale coefficient $c \rightarrow \pm\infty$.

random rather than deterministic number of summands. It is this feature that is decisive in determining the asymptotic equivalence in the theorem.

4. Heuristic arguments and an illustrative example

To provide some intuition on the limit result (8) we use the following simple example. Suppose $x = 1, q \sim U[0, 1], u_1 = u_2 = u$ with variance 1, and u is independent of q in (1). In this simple case, $\|\delta_n\|^2 = \delta_n^2, f_q(\gamma_0) = 1, Q = 1, \Omega_\ell = 1,$

$$z_{1i} = \bar{z}_{1i} = \delta_0^2 + 2\delta_0 u_i \text{ and } z_{2i} = \bar{z}_{2i} = \delta_0^2 - 2\delta_0 u_i,$$

so that

$$\mathbb{E}[z_{\ell i}] = \delta_0^2 \text{ and } \mathbb{V}(z_{\ell i}) = 4\delta_0^2.$$

Let $v > 0$ be given. Then

$$\begin{aligned} \mathbb{E}[D(v)] &= \mathbb{E}[\mathbb{E}[D(v) | N_2(\cdot)]] = \mathbb{E}[N_2(v)\mathbb{E}[z_{2i}]] \\ &= \mathbb{E}[N_2(v)\delta_0^2] = \delta_0^2 v, \end{aligned}$$

and, since $D(v) = \sum_{i=1}^{N_2(v)} z_{2i}$ when $v > 0$, we have

$$\begin{aligned} \mathbb{V}(D(v)) &= \mathbb{E}[\mathbb{V}(D(v) | N_2(\cdot))] + \mathbb{V}(\mathbb{E}[D(v) | N_2(\cdot)]) \\ &= \mathbb{E}[N_2(\cdot)\mathbb{V}(z_{2i})] + \mathbb{V}(N_2(v)\mathbb{E}[z_{2i}]) \\ &= \mathbb{E}[4\delta_0^2 N_2(v)] + \mathbb{V}(N_2(v)\delta_0^2) = 4\delta_0^2 v + \delta_0^4 v \\ &= \delta_0^2 v (4 + \delta_0^2). \end{aligned} \tag{9}$$

By Example 3(i) of Robbins (1948), we have the following CLT for a random number of summands

$$\frac{D(v) - \mathbb{E}[D(v)]}{\mathbb{V}(D(v))} = \frac{\sum_{i=1}^{N_2(v)} z_{2i} - \delta_0^2 v}{\sqrt{\delta_0^2 v (4 + \delta_0^2)}} \xrightarrow{v \rightarrow \infty} N(0, 1).$$

That is, as $v \rightarrow \infty$,

$$\sum_{i=1}^{N_2(v)} z_{2i} \sim_d \delta_0^2 v + \sqrt{\delta_0^2 v (4 + \delta_0^2)} N(0, 1),$$

where the symbol \sim_d is read as ‘approximately distributed as’. Making the change-of-variables $v \mapsto v/\delta_0^2$, we have

$$\sum_{i=1}^{N_2(v/\delta_0^2)} z_{2i} \sim_d v + \sqrt{v(4 + \delta_0^2)} N(0, 1) \tag{10}$$

$$\sim_d v + 2\sqrt{v} N(0, 1) =_{fd} C(v), \tag{11}$$

as $\delta_0 \rightarrow 0$, giving the finite dimensional distribution (represented as ‘ $_{fd}$ ’ in (11)) for given v of $C(v)$ in (5) in the present example.

In the final line of approximation above, the quantity $\delta_0^4 v$, which is present in the variance of $\sum_{i=1}^{N_2(v)} z_{2i}$ in (9) and which appears after the transformation $v \mapsto v/\delta_0^2$ as $\delta_0^2 v$ in (10), is negligible when $\delta_0 \rightarrow 0$. Since this term $\delta_0^4 v$ arises from the randomness of the number of summands $N_2(\cdot)$, it follows from (11) that it can be neglected asymptotically if $\delta_0 \rightarrow 0$ as it is of smaller order. In other words, if $N_2(\cdot)$ is replaced by its mean or if the random number of summands is changed to a deterministic summation, the final approximation is unchanged. In effect, upon rescaling so that $v \mapsto v/\delta_0^2 \rightarrow \infty$ as $\delta_0 \rightarrow 0$, the random sum CLT argument leads directly to the finite dimensional distribution of $C(v)$.

While the CLT of Robbins (1948, Theorem 1) cannot strictly be applied as it stands in the present example, it provides a powerful measure of intuition.⁶ Moreover, because $D(v/\delta_0^2)$ is an

independent increments process, we naturally expect tightness to hold and then, in view of (11), we have weak convergence of the random sum process $\sum_{i=1}^{N_2(v/\delta_0^2)} z_{2i} \rightsquigarrow v + 2W(v)$ as a stochastic process on $[0, \infty)$. In sum, these heuristics deliver the sequence of approximations

$$\begin{aligned} n\delta_0^2 (\hat{\gamma} - \gamma_0) &\sim_d \delta_0^2 \arg \min_v \sum_{i=1}^{N_2(v)} z_{2i} \\ &= \arg \min_v \sum_{i=1}^{N_2(v/\delta_0^2)} z_{2i} \\ &\sim_d \arg \min_v \{v + 2W(v)\} \\ &= \arg \min_v C(v), \end{aligned}$$

which are rigorously justified in the general setting by Theorem 1 above.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.econlet.2018.08.039>.

References

Bai, J., 1997. Estimation of a change point in multiple regression models. *Rev. Econ. Stat.* 79, 551–563.
 Billingsley, P., 1968. *Convergence of Probability Measures*. J. Wiley, New York.
 Bykhovskaya, A., Phillips, P.C.B., 2018. Boundary limit theory for functional local to unity regression. *J. Time Ser. Anal.* 39, 523–562.
 Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* 21, 520–533.
 Chan, N.H., Wei, C.Z., 1987. Asymptotic Inference for Nearly Nonstationary AR(1) Processes. *Ann. Statist.* 15, 1050–1063.
 Hansen, B.E., 2000. Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
 Hansen, B.E., 2011. Threshold autoregression in economics. *Stat. Interface* 4, 123–127.
 Heckman, J.J., 1997. Instrumental variables - a study of implicit behavioral assumption used in making program evaluations. *J. Hum. Resour.* 32, 441–462.
 Kim, J., Pollard, D., 1990. Cube root asymptotics. *Ann. Statist.* 18, 191–219.
 Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
 Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.
 Picard, D., 1985. Testing and estimating change-points in times series. *Adv. Appl. Probab.* 17, 841–867.
 Robbins, H., 1948. The asymptotic distribution of the sum of a random Number of Random Variables. *Bull. Amer. Math. Soc.* 115, 1–1161.
 Van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.
 Yu, P., 2012. Likelihood estimation and inference in threshold regression. *J. Econometrics* 167, 274–294.
 Yu, P., 2014. The bootstrap in threshold regression. *Econ. Theory* 30, 676–714.
 Yu, P., 2015. Adaptive estimation of the threshold point in threshold regression. *J. Econometrics* 189, 83–100.
 Yu, P., Zhao, Y.Q., 2013. Asymptotics for threshold regression under general conditions. *Econom. J.* 16, 430–462.

(10) in Robbins (1948), we note that $\lambda = v/\delta_0^2 \rightarrow \infty, \sigma^2 = \text{Var}(\sum_{i=1}^{N_2(v/\delta_0^2)} z_{2i}) = v(4 + \delta_0^2)$ does not diverge to infinity, and $\gamma = \sqrt{\text{Var}(N_2(v/\delta_0^2))} = \sqrt{v/\delta_0^2} \neq o(\sigma^2)$. In other words, the Robbins (1948) theorems and corollaries cannot be strictly applied in the present case but are nonetheless highly intuitive in the illustrative example.

⁶ The key point is that as $\delta_0 \rightarrow 0$, the mean $\mathbb{E}[z_{2i}^2] = \delta_0^2 (4 + \delta_0^2)$ depends on δ_0^2 in the present case and is not fixed as in Robbins (1948). If we check the condition