Economics Letters 172 (2018) 123–126

Contents lists available at ScienceDirect

Economics Letters

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

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 xxxx

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.

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

$$y = \begin{cases} x'\beta_1 + u_1, & q \le \gamma; \\ x'\beta_2 + u_2, & q > \gamma; \end{cases} \tag{1}$$

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 $\mathbb{E}\left[u_\ell|x, q\right] = 0$, the threshold parameter γ can be estimated by nonlinear

* 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.

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.

https://doi.org/10.1016/j.econlet.2018.08.039

least squares regression giving the least squares estimator (LSE)

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

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

$$M_n(\gamma) := \min_{\beta_1,\beta_2} \sum_{i=1}^{n} \left(y_i - x_i'\beta_1 1(q_i \le \gamma) - x_i'\beta_2 1(q_i > \gamma)\right)^2,$$

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 $\widehat{\gamma}$ 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, $\widehat{\gamma}$ is n-consistent, and

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

where

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

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

$$\bar{z}_{1i} := 2x_i'\delta_0 u_{1i} + \delta_0' x_i x_i' \delta_0,$$
$$\bar{z}_{2i} := -2x_i'\delta_0 u_{2i} + \delta_0' x_i x_i' \delta_0,$$

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 $z_{1i} = \lim_{\Delta\uparrow 0} \bar{z}_{1i} 1\{\gamma_0 + \Delta < q_i \le \gamma_0\}$ is the limiting conditional value of $\bar{z}_{1i}$ given $\gamma_0 + \Delta < q_i \le \gamma_0$, $\Delta < 0$ with $\Delta \uparrow 0$, and $z_{2i} = \lim_{\Delta\downarrow 0} \bar{z}_{2i} 1\{\gamma_0 < q_i \le \gamma_0 + \Delta\}$ is the limiting conditional value of $\bar{z}_{2i}$ given $\gamma_0 < q_i \le \gamma_0 + \Delta$, $\Delta > 0$ 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 $\bar{z}_\ell$ given $q = \gamma_0$. In this framework, the asymptotic distribution of $\widehat{\gamma}$ 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 $n \to \infty$ and is therefore denoted by δ_n. As long as $\|\delta_n\| \to 0$ and $\sqrt{n}\|\delta_n\| \to \infty$ (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 $\widehat{\gamma}$ is consistent with the convergence rate $a_n := n\|\delta_n\|^2$, and

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

where

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

with $Q = \lim_{n\to\infty} \frac{\delta_n' \mathbb{E}[xx'|q=\gamma_0]\delta_n}{\delta_n'\delta_n}$, $\Omega_\ell = \lim_{n\to\infty} \frac{\delta_n' \mathbb{E}\left[xxu_\ell^2|q=\gamma_0\right]\delta_n}{\delta_n'\delta_n}$, 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 $\widehat{\gamma}$ 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 $\widehat{\gamma}$ 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 $\widehat{\gamma}$ and influence the different convergence rates. From Yu (2014), we have the finite sample formulation

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

where

$$D_n(v) = \sum_{i=1}^{n} \bar{z}_{1i} 1\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0\right) + \sum_{i=1}^{n} \bar{z}_{2i} 1\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n}\right).$$

From Hansen (2000), we have the alternate formulation

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

where

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

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 $\arg\min C_n(v) = \|\delta_n\|^2 \arg\min D_n(v)$. This restandardization relating the criteria explains why the convergence rate of $\widehat{\gamma}$ 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

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

where $N_{1n}(|v|) = \sum_{i=1}^{n} 1\left(\gamma_0 + \frac{v}{n} < q_i \le \gamma_0\right)$ and $N_{2n}(v) = \sum_{i=1}^{n} 1\left(\gamma_0 < q_i \le \gamma_0 + \frac{v}{n}\right)$. 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 $p_n(v) = F_q\left(\gamma_0 + \frac{v}{n}\right) - F_q(\gamma_0)$, 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