

# SEMIPARAMETRIC ESTIMATION OF A BINARY RESPONSE MODEL WITH A CHANGE-POINT DUE TO A COVARIATE THRESHOLD\*

Sokbae Lee  
Department of Economics  
University College London  
London, WC1E 6BT  
United Kingdom

And

Myung Hwan Seo  
Department of Economics  
London School of Economics  
London, WC2A 2AE  
United Kingdom

Discussion paper  
No. EM/2007/516  
February 2007

The Suntory Centre  
Suntory and Toyota International Centres for  
Economics and Related Disciplines  
London School of Economics and Political Science  
Houghton Street  
London WC2A 2AE  
Tel: 020 7955 6679

---

\* Lee thanks the Leverhulme Trust through the funding of the Centre for Microdata Methods and Practice (<http://cemmap.ifs.org.uk>) and the research programme Evidence, Inference and Inquiry (<http://www.evidencescience.org>). Seo thanks the Economic and Social Research Council (ESRC) for financial support.

## Abstract

This paper is concerned with semiparametric estimation of a threshold binary response model. The estimation method considered in the paper is semiparametric since the parameters for a regression function are finite-dimensional, while allowing for heteroskedasticity of unknown form. In particular, the paper considers Manski (1975, 1985)'s maximum score estimator. The model in this paper is irregular because of a change-point due to an unknown threshold in a covariate. This irregularity coupled with the discontinuity of the objective function of the maximum score estimator complicates the analysis of the asymptotic behavior of the estimator. Sufficient conditions for the identification of parameters are given and the consistency of the estimator is obtained. It is shown that the estimator of the threshold parameter is  $n$ -consistent and the estimator of the remaining regression parameters is cube root  $n$ -consistent. Furthermore, we obtain the asymptotic distribution of the estimators. It turns out that a suitably normalized estimator of the regression parameters converges weakly to the distribution to which it would converge weakly if the true threshold value were known and likewise for the threshold estimator.

**Key Words:** Binary response model, maximum score estimation, semiparametric estimation, threshold regression, nonlinear random utility models.

**JEL Classification:** C25

# 1 Introduction

A binary response model is very commonly used in a number of fields. In this model, an observable binary outcome  $Y$  is modelled typically as  $Y = 1(Y^* \geq 0)$ , where  $1(A)$  denotes the indicator function, i.e.,  $1(A) = 1$  if  $A$  is true and zero otherwise, and  $Y^*$  is an unobservable continuous variable that determines the binary outcome  $Y$  (see e.g. Manski, 1988). In most applications,  $Y^*$  has the following form:

$$(1.1) \quad Y^* = X'\beta_0 + U,$$

where  $X$  is a vector of observed random variables including an intercept term,  $\beta_0$  is a vector of unknown parameters, and  $U$  is an unobserved random variable.

In this paper, we consider a threshold regression model for  $Y^*$  instead of the linear regression model (1.1). In particular, the model is

$$(1.2) \quad Y^* = \begin{cases} W_1'\beta_{10} + W_2'\beta_{20} + U, & \text{if } D \leq \gamma_0, \\ W_1'\beta_{10} + W_2'\beta_{30} + U, & \text{if } D > \gamma_0, \end{cases}$$

or equivalently

$$Y^* = W'\beta_0 + Z'\delta_0 1\{D > \gamma_0\} + U,$$

where  $W = (W_1', W_2)'$ ,  $Z = W_2$ ,  $\beta_0 = (\beta_{10}', \beta_{20}')'$ , and  $\delta_0 = \beta_{30} - \beta_{20}$ . In particular,  $D$  is the observed threshold variable and  $\gamma_0$  is the threshold parameter. We denote by  $X$  the set of regressors  $(W', D)'$  and by  $\theta_0 = (\beta_0', \delta_0)'$  the set of regression parameters excluding  $\gamma_0$ . The threshold variable  $D$  can be an element of  $W$ . The literature on the threshold model (also called change-point model, two-phase regression, or sample splitting) is vast. It has been studied for autoregressive models (e.g. Tong, 1990; Chan, 1993), for linear regression models (e.g. Hansen, 2000; Koul and Qian, 2002), for nonparametric models (e.g. Delgado and Hidalgo, 2000), and more recently for transformation models (e.g. Pons, 2003; Kosorok and Song, 2007).

Threshold binary regression models have a wide variety of applications. For example, in biostatistics, dose-response models are modelled with threshold parameters (see, e.g.

Cox, 1987; Schwartz et al., 1995). In epidemiology, logistic regressions with change-points are used to model the relationship between the continuous exposure variable and disease risk (see Pastor and Guallar, 1998; Pastor-Barriuso et al., 2003). In economics, finance, and management, random utility models that are nonlinear in income and/or price are commonly employed (see, e.g. Herriges and Kling, 1999; Dagsvik and Karlstrom, 2005) and Terui and Dahana (2006) adopted asymmetric market response model with the threshold due to price based on consumer behavior theory. The effect of financial contagion (see, e.g. Forbes and Rigobon, 2002) can also be modelled as a discontinuous threshold effect as in Pesaran and Pick (in press). While the threshold model is easier to interpret than more complex nonlinear models, the irregular feature of the model makes estimation of the unknown change-point complicated. To our best knowledge, all existing methods for estimating a change-point according to a covariate threshold assumes that the distribution of  $U$  in (1.2) belongs to a family of parametric distributions.

The purpose of this paper is to develop a method for estimating  $(\theta_0, \gamma_0)$  in (1.2) without imposing the parametric distribution on  $U$ . First, we establish identification of  $(\theta_0, \gamma_0)$ , allowing for conditional heteroskedasticity of unknown form, under the conditional median independence assumption. This is substantially weaker than the typical assumption such that the distribution of  $U$  is parametric and  $U$  and  $X$  are independent. The conditional heteroskedasticity is particularly important in random utility models as shown in Brown and Walker (1989). Manski (1988) has shown that if the distribution of  $X$  has sufficiently rich support, the finite-dimensional parameters in a binary regression model with (1.1) are identified up to scale. As in Manski (1988),  $(\theta_0, \gamma_0)$  in (1.2) is identified (up to scale with respect to  $\theta_0$ ) under some regularity conditions, which will be given in Section 2. Once  $(\theta_0, \gamma_0)$  is identified via the conditional median independence assumption, a natural estimation strategy for  $(\theta_0, \gamma_0)$  in (1.2) is to apply Manski (1975, 1985)'s maximum score estimator to the threshold model (1.2). Section 3 describes the corresponding maximum score estimator and section 4 establishes the consistency of the estimator.

We also develop convergence rates and asymptotic distributions, which are highly non-standard due to mixed irregularities of the model and maximum score estimation. In Section 5, it is shown that  $\hat{\gamma}_n$  is  $n^{-1}$ -consistent and  $\hat{\theta}_n$  is  $n^{-1/3}$ -consistent. It is interesting to find that the fast rate of convergence for the threshold estimator is preserved for the maximum score estimation. Section 6 gives the asymptotic distributions of  $n(\hat{\gamma}_n - \gamma_0)$  and  $n^{1/3}(\hat{\theta}_n - \theta_0)$ . It turns out that both estimators  $\hat{\gamma}$  and  $\hat{\theta}$  are adaptive in that  $n(\hat{\gamma}_n - \gamma_0)$  and  $n^{1/3}(\hat{\theta}_n - \theta_0)$  converge weakly to the distributions to which they would converge weakly if the other parameters were known. In fact, it is shown that the error due to the estimation converges in probability to zero at a rate of  $n^{-1/6}$ . Section 7 discusses the subsampling inference that enables us to carry out inference, although the distributions of  $n(\hat{\gamma}_n - \gamma_0)$  and  $n^{1/3}(\hat{\theta}_n - \theta_0)$  are nonstandard and cannot be tabulated. Concluding remarks are given in Section 8. The proofs of theorems are given in the main text and the proofs of lemmas are in Appendix.

## 2 Identification of $\theta_0$ and $\gamma_0$

This section provides regularity conditions under which  $\theta_0$  and  $\gamma_0$  are identified. Let  $F_{U|X}(u|x)$  and  $f_{U|X}(u|x)$ , respectively, denote the cumulative distribution function and probability density function of  $U$  conditional on  $X = x$ .

**Assumption 1.** *Assume that  $\|\theta_0\| = 1$  and  $\delta_0 \neq 0$ .*

As in the linear binary response model (1.1), the scale of  $\theta_0$  is not identified since the distribution of  $U$  conditional on  $X = x$  is unknown. In this paper, the scale normalization on  $\theta_0$  is given by  $\|\theta_0\| = 1$ . If  $\delta_0$  is zero, then  $\gamma_0$  is unidentified.

**Assumption 2.** *The distribution of  $U$  conditional on  $X = x$  is absolutely continuous with respect to Lebesgue measure and the corresponding conditional density is uniformly continuous and positive everywhere with probability one. In addition,  $F_{U|X}(0|x) = 0.5$  for almost every  $x$ .*

This assumption allows for an arbitrary form of dependence between  $U$  and  $X$  as long

as the conditional median independence assumption is satisfied. For the purpose of identification, any other quantile independence assumption would work as well (see, Assumption QI in Manski, 1988).

**Assumption 3.** *The support of  $W \in \mathbb{R}^q$  is not contained in any proper linear subspace of  $\mathbb{R}^q$ .*

**Assumption 4.** *There exists at least one  $k \in [1, \dots, q]$  such that the  $k$ -th element of  $\beta_0$  is non-zero and such that, for almost every value of  $\tilde{w} \equiv (w_1, \dots, w_{k-1}, w_{k+1}, \dots, w_q)$  and  $d$ ,  $Pr[a_1 < W_k < a_2 | \tilde{W} = \tilde{w}, D = d] > 0$  for all open intervals  $(a_1, a_2) \in \mathbb{R}$ .*

Assumptions 3 and 4 are simply the restatement of Conditions X1 and X3 of Manski (1988), respectively.

**Assumption 5.**  *$D$  is continuously distributed with support containing  $\gamma_0$ .*

This condition ensures the identification of  $\gamma_0$ . If  $\gamma_0$  were not contained in the support of  $D$ , then the threshold model (1.2) would not have a two-phase feature in observed data.

**Theorem 1.** *Let Assumptions 1-5 hold. Then  $\theta_0$  and  $\gamma_0$  are identified.*

*Proof.* Note that by Assumptions 2, 3 and 4, for each fixed  $D = d$ ,  $\beta_0$  and  $\xi_0(d) \equiv \delta_0 1(d > \gamma_0)$  are identified by Corollary 2 of Manski (1988). Thus,  $\delta_0$  is identified as  $\delta_0 = \xi_0(d)$  for any  $d$  such that  $\xi_0(d) \neq 0$ . Finally, given Assumption 5,  $\gamma_0$  is identified as the infimum of  $d$ 's satisfying  $\xi_0(d) \neq 0$  or as the supremum of  $d$ 's satisfying  $\xi_0(d) = 0$ .  $\square$

### 3 The Maximum Score Estimator

We consider estimation with an independent and identically distributed sample  $\{(Y_i, X_i) : i = 1, \dots, n\}$  of  $(Y, X)$ . To describe the maximum score estimator, for any  $\theta = (\beta', \delta)'$ ,  $\gamma$  and  $x = (w, z, d)$ , define

$$(3.1) \quad S_n(\theta, \gamma) = \sum_{i=1}^n (2Y_i - 1) 1 \{G(X_i, \theta, \gamma) \geq 0\},$$

where

$$G(x, \theta, \gamma) = w'\beta + z'\delta 1(d > \gamma).$$

Then Manski (1975, 1985)'s maximum score estimator  $(\hat{\theta}_n, \hat{\gamma}_n)$  of  $(\theta_0, \gamma_0)$  can be obtained in two steps. In the first stage, for fixed  $\gamma$ , obtain

$$\hat{\theta}_n(\gamma) = \operatorname{argmax}_{\theta} S_n(\theta, \gamma)$$

with the scale normalization on  $\theta$  with  $\|\theta\| = 1$ . In the second stage, let  $\hat{S}_n(\gamma) = S_n(\hat{\theta}_n(\gamma), \gamma)$ . Obtain  $\hat{\gamma}_n = \inf\{\gamma_n : \gamma_n = \operatorname{argmax}_{\gamma} \hat{S}_n(\gamma)\}$ . Then the maximum score estimator of  $\theta_0$  is  $\hat{\theta}_n = \hat{\theta}_n(\hat{\gamma}_n)$ . Since there could be infinitely many  $\gamma$ 's at which  $\hat{S}_n(\gamma)$  is maximized,  $\hat{\gamma}_n$  is defined as the infimum of those.

The discontinuity of  $S_n$  with respect to  $(\theta, \gamma)$  complicates the analysis of the asymptotic behavior of the maximum score estimator. There exist two distinct sources of discontinuity: on the one hand, the discontinuity arises because of the presence of the indicator function  $1\{G(X_i, \theta, \gamma) \geq 0\}$  in (3.1); on the other hand, it may occur because of a possible change-point in (1.2) due to an unknown threshold  $\gamma_0$ . We now turn to the asymptotic properties of the maximum score estimator.

## 4 Consistency of the Estimator

This section gives conditions under which the maximum score estimator is consistent.

**Assumption 6.** *Assume that  $\theta_0$  and  $\gamma_0$  are in a compact subset of  $\mathbb{R}^p$ , where  $p$  is the dimension of  $(\theta_0, \gamma_0)$ .*

**Assumption 7.** *Assume that  $\{(Y_i, X_i) : i = 1, \dots, n\}$  is independent and identically distributed.*

These are standard assumptions in the literature. The following theorem gives the consistency result.

**Theorem 2.** *Let Assumptions 1-7 hold. Then  $(\hat{\theta}_n, \hat{\gamma}_n) \rightarrow_p (\theta_0, \gamma_0)$ .*

*Proof.* Define

$$H^*(x) = E[1\{G(X, \theta_0, \gamma_0) + U \geq 0\} - 1\{G(X, \theta_0, \gamma_0) + U < 0\} | X = x]$$

and

$$\Delta^*(\theta, \gamma) = n^{-1}E[S_n(\theta, \gamma) - S_n(\theta_0, \gamma_0)].$$

Then it can be shown that

$$\begin{aligned} \Delta^*(\theta, \gamma) = E \left[ H^*(X) \left( 1\{G(X, \theta, \gamma) \geq 0 > G(X, \theta_0, \gamma_0)\} \right. \right. \\ \left. \left. - 1\{G(X, \theta_0, \gamma_0) \geq 0 > G(X, \theta, \gamma)\} \right) \right]. \end{aligned}$$

Note that

$$H^*(x) = 1 - 2F_{U|X}[-G(x, \theta_0, \gamma_0)|x].$$

By the assumption that  $F_{U|X}[0|x] = 0.5$ , note that  $H^*(x) \geq 0$  when  $G(x, \theta_0, \gamma_0) \geq 0$  and that  $H^*(x) < 0$  when  $G(x, \theta_0, \gamma_0) < 0$ . Define

$$\mathcal{Q}(\theta, \gamma) = \left[ x \in \text{supp}(X) : \{G(x, \theta, \gamma) \geq 0 > G(x, \theta_0, \gamma_0)\} \cup \{G(x, \theta_0, \gamma_0) \geq 0 > G(x, \theta, \gamma)\} \right].$$

By arguments identical to those used to prove Proposition 2 of Manski (1988),  $(\theta_0, \gamma_0)$  are identified if and only if  $\Pr(X \in \mathcal{Q}(\theta, \gamma)) > 0$  for any  $(\theta, \gamma) \neq (\theta_0, \gamma_0)$ . Therefore,  $\Delta^*(\theta, \gamma)$  is non-positive everywhere and is equal to zero only when  $(\theta, \gamma) = (\theta_0, \gamma_0)$ . Also, notice that  $G(S, \theta, \gamma)$  is continuous at each  $\theta$  and  $\gamma$  with probability one. Thus,  $S_n(\theta, \gamma)$  converges in probability to  $E[S_n(\theta, \gamma)]$  uniformly over  $(\theta, \gamma)$  by Lemma 2.4 of Newey and McFadden (1994). Then the consistency of the estimator follows from Theorem 2.1 of Newey and McFadden (1994).  $\square$

## 5 Rates of Convergence

This section establishes the rates of convergence in probability of  $(\hat{\theta}_n, \hat{\gamma}_n)$  to  $(\theta_0, \gamma_0)$ . First, we show that  $\hat{\gamma}_n$  is  $n^{-1}$ -consistent for  $\gamma_0$ . To do so, we begin with the following lemma, which is similar to Claim 2 in the proof of Proposition 1 of Chan (1993) and Lemma 3.2 of Koul and Qian (2002). The proof of the lemma is given in Appendix.



**Lemma 1.** For any random variables  $(V, D)$  satisfying  $E[V|D] = 0$  and  $E[V^2|D] < C < \infty$  almost surely, assume that  $\{(V_i, D_i) : i = 1, \dots, n\}$  is a random sample of  $(V, D)$  and that  $D$  is continuously distributed and has a bounded, continuous, positive density in a neighborhood of  $x \in \mathbb{R}$ . Then, for each  $\delta > 0$  and  $\varepsilon > 0$ , there exists a positive constant  $B < \infty$  such that for all  $0 < \Delta < 1$  and for all  $n > B/\Delta$ ,

$$(a) \quad \Pr \left\{ \sup_{\frac{B}{n} < r < \Delta} \left| \frac{1}{n} \sum_{i=1}^n \frac{1 \{x < D_i < x + r\}}{\Pr \{x < D < x + r\}} - 1 \right| > \delta \right\} < \varepsilon$$

$$(b) \quad \Pr \left\{ \sup_{\frac{B}{n} < r < \Delta} \left| \frac{1}{n} \sum_{i=1}^n V_i \frac{1 \{x < D_i < x + r\}}{\Pr \{x < D_i < x + r\}} \right| > \delta \right\} < \varepsilon.$$

We make the following additional assumptions.

**Assumption 8.** Assume that  $D$  is continuously distributed with full support on  $\mathbb{R}$  and its probability density function is strictly positive, bounded and continuous in a neighborhood of  $\gamma_0$ .

This assumption is needed to utilize Lemma 1.

**Assumption 9.** With positive probability,  $Z'\delta_0 \neq 0$  at  $D = \gamma_0$ .

This assumption imposes that the regression function in (1.2) is discontinuous. When the regression function is continuous, then we would have different rates of convergence and asymptotic distributions.

The following theorem gives the  $n^{-1}$ -consistency of  $\hat{\gamma}_n$  to  $\gamma_0$ .

**Theorem 3.** Let Assumptions 1-9 hold. Then,  $\hat{\gamma}_n = \gamma_0 + O_p(n^{-1})$ .

*Proof.* Define

$$\begin{aligned} \Delta_{S_n}(\theta, \gamma) &= S_n(\theta, \gamma) - S_n(\theta, \gamma_0) \\ &= \sum_{i=1}^n (2Y_i - 1) (1 \{G(X_i, \theta, \gamma) \geq 0\} - 1 \{G(X_i, \theta, \gamma_0) \geq 0\}). \end{aligned}$$

Suppose that  $\gamma > \gamma_0$ . Then if  $D_i > \gamma$  or  $D_i \leq \gamma_0$ ,

$$1 \{G(X_i, \theta, \gamma) \geq 0\} - 1 \{G(X_i, \theta, \gamma_0) \geq 0\} = 0.$$

Suppose now that  $\gamma < \gamma_0$ . Then if  $D_i > \gamma_0$  or  $D_i \leq \gamma$ ,

$$1 \{G(X_i, \theta, \gamma) \geq 0\} - 1 \{G(X_i, \theta, \gamma_0) \geq 0\} = 0.$$

Using these, we may write

$$(5.1) \quad \Delta_{S_n}(\theta, \gamma) = \sum_{i=1}^n 1 \{\gamma_0 < D_i \leq \gamma\} \zeta_i^+(\theta) + \sum_{i=1}^n 1 \{\gamma < D_i \leq \gamma_0\} \zeta_i^-(\theta),$$

where  $\zeta_i^+(\theta) = [I_{1i}^+(\theta) - I_{2i}^+(\theta) + I_{3i}^+(\theta) - I_{4i}^+(\theta)]$  with

$$\begin{aligned} I_{1i}^+(\theta) &= 1 \{W_i' \beta \geq 0, W_i' \beta + Z_i' \delta < 0, W_i' \beta_0 + Z_i' \delta_0 \geq -U_i\} \\ I_{2i}^+(\theta) &= 1 \{W_i' \beta \geq 0, W_i' \beta + Z_i' \delta < 0, W_i' \beta_0 + Z_i' \delta_0 < -U_i\} \\ I_{3i}^+(\theta) &= 1 \{W_i' \beta < 0, W_i' \beta + Z_i' \delta \geq 0, W_i' \beta_0 + Z_i' \delta_0 < -U_i\} \\ I_{4i}^+(\theta) &= 1 \{W_i' \beta < 0, W_i' \beta + Z_i' \delta \geq 0, W_i' \beta_0 + Z_i' \delta_0 \geq -U_i\} \end{aligned}$$

and  $\zeta_i^-(\theta) = [I_{1i}^-(\theta) - I_{2i}^-(\theta) + I_{3i}^-(\theta) - I_{4i}^-(\theta)]$  with

$$\begin{aligned} I_{1i}^-(\theta) &= 1 \{W_i' \beta \geq 0, W_i' \beta + Z_i' \delta < 0, W_i' \beta_0 < -U_i\} \\ I_{2i}^-(\theta) &= 1 \{W_i' \beta \geq 0, W_i' \beta + Z_i' \delta < 0, W_i' \beta_0 \geq -U_i\} \\ I_{3i}^-(\theta) &= 1 \{W_i' \beta < 0, W_i' \beta + Z_i' \delta \geq 0, W_i' \beta_0 \geq -U_i\} \\ I_{4i}^-(\theta) &= 1 \{W_i' \beta < 0, W_i' \beta + Z_i' \delta \geq 0, W_i' \beta_0 < -U_i\}. \end{aligned}$$

The following lemma, which is proved in Appendix, is useful to prove the theorem.

**Lemma 5.1.** *For any  $\Delta > 0$ , define  $\Theta_\Delta = \{\theta : \|\theta - \theta_0\| < \Delta\}$ . There exists a sufficiently small  $\Delta > 0$  such that*

$$\sup_{\theta \in \Theta_\Delta} E [\zeta_i^+(\theta) | D_i] < 0 \quad \text{and} \quad \sup_{\theta \in \Theta_\Delta} E [\zeta_i^-(\theta) | D_i] < 0$$

*almost surely.*

Now write

$$\begin{aligned}\Delta_{S_n}(\theta, \gamma) &= \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} E[\zeta_i^+(\theta)|D_i] + \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} E[\zeta_i^-(\theta)|D_i] \\ &\quad + \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} \{\zeta_i^+(\theta) - E[\zeta_i^+(\theta)|D_i]\} \\ &\quad + \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} \{\zeta_i^-(\theta) - E[\zeta_i^-(\theta)|D_i]\}.\end{aligned}$$

Then apply Lemmas 1 and 5.1 to obtain  $\Delta_{S_n}(\theta, \gamma) < 0$  with probability approaching one for all  $|\gamma - \gamma_0| > B/n$  and  $\theta \in \Theta_\Delta$  and with some sufficiently large  $B$  and sufficiently small  $\Delta$ . As  $(\hat{\theta}, \hat{\gamma})$  is the maximizer of  $S_n(\theta, \gamma)$ , we conclude that  $\hat{\gamma} = \gamma_0 + O_p(1/n)$ .  $\square$

We now turn to the rates of convergence of  $\hat{\theta}_n$  to  $\theta_0$ . Given the result of Theorem 3 and Kim and Pollard (1990), it is not surprising that we establish the  $n^{-1/3}$ -consistency for  $\hat{\theta}_n$ . To do so, write

$$(5.2) \quad S_n(\theta, \gamma) = S_{n1}(\gamma) + S_{n2}(\theta) + S_{n3}(\theta, \gamma),$$

where  $S_{n1}(\gamma) = S_n(\theta_0, \gamma) - S_n(\theta_0, \gamma_0)$ ,  $S_{n2}(\theta) = S_n(\theta, \gamma_0) - S_n(\theta_0, \gamma_0)$ , and

$$S_{n3}(\theta, \gamma) = [S_n(\theta, \gamma) - S_n(\theta, \gamma_0)] - [S_n(\theta_0, \gamma) - S_n(\theta_0, \gamma_0)].$$

We first give the following lemma, which is proved in Appendix.

**Lemma 5.2.** *For every  $A > 0$ , as  $n \rightarrow \infty$ :*

$$\sup_{\theta, \gamma} \{|S_{n3}(\theta, \gamma)| : n|\gamma - \gamma_0| \leq A\} = O_p(1).$$

**Assumption 10.** *As a function of  $\theta$ ,  $E[(2Y - 1)1\{G(X, \theta, \gamma_0) \geq 0\}]$  has a strictly negative definite second derivative matrix at  $\theta_0$ .*

This assumption is necessary to apply Corollary 4.2 of Kim and Pollard (1990), who also give sufficient conditions for the maximum score estimator (see Section 6.4 in Kim and Pollard, 1990).

**Theorem 4.** *Let Assumptions 1-10 hold. Then,  $\hat{\theta}_n = \theta_0 + O_p(n^{-1/3})$ .*

*Proof.* In view of Corollary 4.2 of Kim and Pollard (1990), it suffices to show that

$$n^{-1}S_n(\hat{\theta}) \geq \sup_{\theta} n^{-1}S_n(\theta) - O_p(n^{-2/3}),$$

which in turn follows from the same Corollary, Lemma 5.2, and the fact that  $S_{n1}$  does not depend on  $\theta$ . □

## 6 The Asymptotic Distribution of the Estimator

Since  $\hat{\gamma}$  is  $n^{-1}$ -consistent and  $\hat{\theta}$  is  $n^{-1/3}$ -consistent, it suffices to consider a  $n^{-1}$ -neighborhood of  $\gamma_0$  and a  $n^{-1/3}$ -neighborhood of  $\theta_0$ . Then Lemma 5.2 can be strengthened by restricting the parameter space in the following way.

**Lemma 6.1.** *For every  $A > 0$ , as  $n \rightarrow \infty$ :*

$$\sup_{\theta, \gamma} \left\{ |S_{n3}(\theta, \gamma)| : n^{1/3} \|\theta - \theta_0\| \leq A, n|\gamma - \gamma_0| \leq A \right\} = O_p(n^{-1/6}).$$

To establish the asymptotic distribution of  $n(\hat{\gamma} - \gamma_0)$ , we now consider the weak convergence of  $S_{n1}$  as a random variable on the space of cadlag functions, denoted by  $\mathcal{D}$ , equipped with the Skorohod topology and on its restrictions to the space of cadlag functions on  $[-A, A]$ , denoted by  $\mathcal{D}_A$ , for any  $A > 0$ . This approach is similar to that taken in the recent literature (for example, Chan, 1993; Koul and Qian, 2002; Pons, 2003; Kosorok and Song, 2007). Define  $\nu = n(\gamma - \gamma_0)$ . Using (5.1), define a rescaled version of  $S_{n1}(\gamma)$  as

$$\tilde{S}_{n1}(\nu) = \tilde{S}_{n1}^+(\nu)1(\nu > 0) + \tilde{S}_{n1}^-(\nu)1(\nu < 0),$$

where

$$\begin{aligned} \tilde{S}_{n1}^+(\nu) &= \sum_{i=1}^n 1 \{ \gamma_0 < D_i \leq \gamma_0 + n^{-1}\nu \} \zeta_i^+(\theta_0), \\ \tilde{S}_{n1}^-(\nu) &= \sum_{i=1}^n 1 \{ \gamma_0 + n^{-1}\nu < D_i \leq \gamma_0 \} \zeta_i^-(\theta_0). \end{aligned}$$

To describe the asymptotic behavior of  $\tilde{S}_{n1}(\nu)$ , let  $f_D(d)$  denote the probability density function of  $D$ , let  $\tilde{\nu}^+$  and  $\tilde{\nu}^-$  denote two independent jump processes on  $\mathbb{R}$  such that  $\tilde{\nu}^+(s)$  is a Poisson random variable with parameter  $sf_D(\gamma_0)$  for  $s > 0$  and  $\tilde{\nu}^+(s) = 0$  for  $s \leq 0$  and  $\tilde{\nu}^-(s)$  is a Poisson random variable with parameter  $-sf_D(\gamma_0)$  for  $s < 0$  and  $\tilde{\nu}^-(s) = 0$  for  $s \geq 0$ . In addition, let  $\{\tilde{\zeta}_k^+ : k = 0, 1, \dots\}$  and  $\{\tilde{\zeta}_k^- : k = 0, 1, \dots\}$  be independent sequences of i.i.d. random variables with characteristic functions

$$\varphi^+(t) = E[\exp(it\zeta^+) | D = \gamma_0^+] \quad \text{and} \quad \varphi^-(t) = E[\exp(it\zeta^-) | D = \gamma_0^-],$$

respectively, and let  $\tilde{\zeta}_0^+ = \tilde{\zeta}_0^- = 0$ . Let  $\tilde{S}_1(s) = \tilde{S}_1^+(s)1\{s > 0\} + \tilde{S}_1^-(s)1\{s < 0\}$  be a right-continuous jump process on  $\mathbb{R}$ , where

$$\tilde{S}_1^+(s) = \sum_{0 \leq k \leq \tilde{\nu}^+(s)} \tilde{\zeta}_k^+ \quad \text{and} \quad \tilde{S}_1^-(s) = \sum_{0 \leq k \leq \tilde{\nu}^-(s)} \tilde{\zeta}_k^-.$$

The following lemma establishes the weak convergence of  $\tilde{S}_{n1}$ .

**Lemma 6.2.** *As  $n \rightarrow \infty$ ,  $\tilde{S}_{n1}$  converges weakly to  $\tilde{S}_1$  in  $\mathcal{D}_A$  for every  $A > 0$ .*

Let  $\nu_{\tilde{S}_1}^* = \inf\{\nu^* : \nu^* = \operatorname{argmax}_\nu \tilde{S}_1(\nu)\}$ . As in the literature (see, e.g., Chan, 1993; Koul and Qian, 2002; Pons, 2003; Kosorok and Song, 2007),  $\tilde{S}_1$  tends to  $-\infty$  almost surely as  $|\nu| \rightarrow \infty$ , thereby implying that the distribution of  $\nu_{\tilde{S}_1}^*$  is tight. Lemma 6.2 coupled with Lemma 6.1 gives the main theorem of this paper.

**Theorem 5.** *Let Assumptions 1-9 hold. Then, as  $n \rightarrow \infty$ ,  $n(\hat{\gamma} - \gamma_0)$  converges weakly to  $\nu_{\tilde{S}_1}^*$ . In addition,  $n^{1/3}(\hat{\theta} - \theta_0)$  and  $\operatorname{argmax}_\mu n^{-1/3}S_{n2}(\theta_0 + n^{-1/3}\mu)$  converge weakly to the same limiting distribution.*

*Proof.* Note that  $n(\hat{\gamma} - \gamma_0)$  is uniformly tight by Theorem 3. Since  $S_{n2}$  depends on only  $\theta$ , Lemmas 6.1 and 6.2 ensure that conditions of the argmax continuous mapping theorem of van der Vaart and Wellner (1996, Theorem 3.2.2, p.286) are satisfied. Then the first conclusion follows from the argmax continuous mapping theorem. Similarly,  $n^{1/3}(\hat{\theta} - \theta_0)$  is uniformly tight by Theorem 4. Since  $S_{n1}$  depends on only  $\gamma$ , by Lemma 6.2 and again by the argmax continuous mapping theorem, the second conclusion follows.  $\square$

Theorem 5 implies that the limiting distributions of  $n(\hat{\gamma} - \gamma_0)$  and  $n^{1/3}(\hat{\theta} - \theta_0)$  are the same as if the other parameters were known. The limiting distribution of  $n^{1/3}(\hat{\theta} - \theta_0)$  is not shown here since it is already given in Kim and Pollard (1990, Example 6.4).

## 7 Inference

The asymptotic distributions obtained in this paper are nonstandard and cannot be tabulated as they depend on nuisance parameters in complex manners. It is known from Abrevaya and Huang (2005) that the bootstrap does not estimate the asymptotic distribution of  $\hat{\theta}$  consistently. While the  $m$  out of  $n$  bootstrap in Lee and Pun (2006) is more general and allows for the standard maximum score estimator, their regularity conditions do not permit threshold models. However, subsampling provides a consistent inferential method for the asymptotic distributions of both  $\hat{\theta}$  and  $\hat{\gamma}$ , as in Delgado et al. (2001) and Gonzalo and Wolf (2005). Confidence intervals can be constructed following the standard subsampling procedure, see e.g. Politis et al. (1999).

As our asymptotic development is based on the discontinuity of the regression function (Assumption 9), the above asymptotics breaks down when the model comes close to a continuous one. In this case, the subsampling confidence intervals may be problematic as their convergences are not uniform around the continuity point, see e.g. Andrews and Guggenberger (2006).

Another interesting inferential approach can be based on the smoothing of the objective function. This can be done by replacing the indicator functions in (3.1) with integrated kernels with smoothing parameters that converge to zero. Horowitz (1992) and Seo and Linton (2007) have developed the smoothing approach for the maximum score estimation and threshold estimation, respectively. This method enables the standard normal inference for both  $\gamma$  and  $\theta$  while it causes the convergence rate of  $\hat{\theta}$  to increase and that of  $\hat{\gamma}$  to decrease. The smoothing, however, demands more stringent smoothness conditions than the ones in this paper as demonstrated in Horowitz (1993).

## 8 Conclusions

This paper has considered the maximum score estimator of a binary response model with a change-point according to the unknown threshold of a covariate, allowing for an arbitrary form of heteroskedasticity. We have obtained the rates of convergence and asymptotic distribution of the estimator. It turns out that the estimator of the threshold parameter is  $n^{-1}$ -consistent, the estimator of the remaining regression parameters is  $n^{-1/3}$ -consistent, and the limiting distributions of the estimators normalized by the rates of convergence are the same as if the other parameters were known. Therefore, an important practical implication of this paper is that the unknown threshold parameter can be estimated precisely with a small or moderate sample size and that there is no loss of efficiency of not knowing other parameters, which are difficult to estimate precisely without a large sample.

## A Appendix: Proofs of Lemmas

*Proof of Lemma 1.* For any  $r \in (0, \Delta]$ , there exist  $0 < m < M < \infty$  such that

$$(A.1) \quad mr < \Pr \{x < D < x + r\} < Mr.$$

Consider a partition of the interval  $(B/n, \Delta) \equiv \cup_{k=0}^{K-1} J_k$ , where  $K$  is the largest integer such that  $b^{K-1}B/n < \Delta$  for some  $b > 1$ , and  $J_k = (b^k B/n, b^{k+1} B/n]$ ,  $k = 0, \dots, K-2$  and  $J_{K-1} = (b^{K-1} B/n, \Delta)$ . Here we prove part (b) only as part (a) can be proved in the same but simpler way. For a  $r \in J_k$ , we have

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n V_i 1 \{x < D_i < x + r\} \right| &\leq \frac{1}{n} \sum_{i=1}^n |V_i| 1 \{x + b^k B/n < D_i \leq x + b^{k+1} B/n\} \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n V_i 1 \{x < D_i \leq x + b^k B/n\} \right| \end{aligned}$$

and

$$\Pr \{x < D < x + r\} \geq \Pr \{x < D < x + b^k B/n\}.$$

Thus,

$$(A.2) \quad \sup_{\frac{B}{n} < r < \Delta} \left| \frac{1}{n} \sum_{i=1}^n V_i \frac{1 \{x < D_i < x + r\}}{\Pr \{x < D < x + r\}} \right| \leq \max_{0 \leq k < K} \frac{1}{n} \sum_{i=1}^n \frac{|V_i| 1 \{x + b^k B/n < D_i \leq x + b^{k+1} B/n\}}{\Pr \{x < D_i < x + b^k B/n\}} \\ + \max_{0 \leq k < K} \left| \frac{1}{n} \sum_{i=1}^n \frac{V_i 1 \{x < D_i \leq x + b^k B/n\}}{\Pr \{x < D \leq x + b^k B/n\}} \right|.$$

First, consider the second term in (A.2). It follows from the Markov and Cauchy-Schwarz inequalities that for any  $\delta_1 > 0$ ,

$$(A.3) \quad \Pr \left\{ \max_{0 \leq k < K} \left| \frac{1}{n} \sum_{i=1}^n \frac{V_i 1 \{x < D_i \leq x + b^k B/n\}}{\Pr \{x < D \leq x + b^k B/n\}} \right| > \delta_1 \right\} \\ \leq \delta_1^{-2} \sum_{k=0}^{K-1} E \left| \frac{1}{n} \sum_{i=1}^n \frac{V_i 1 \{x < D_i \leq x + b^k B/n\}}{\Pr \{x < D < x + b^k B/n\}} \right|^2 \\ \leq \delta_1^{-2} \sum_{k=0}^{K-1} \frac{E[V^2]}{n \cdot \Pr \{x < D \leq x + b^k B/n\}} \\ \leq \frac{E[V^2]}{m \delta_1^2 B (1 - b^{-1})}.$$

Next, note that due to (A.1),

$$\text{var} \left( |V| 1 \{x + b^k B/n < D \leq x + b^{k+1} B/n\} \right) \leq C b^k (b - 1) B/n,$$

for some constant  $C$ . Thus, for the first term in (A.2), we have that

$$\Pr \left\{ \max_{0 \leq k < K} \frac{1}{n} \sum_{i=1}^n \frac{|V_i| 1 \{x + b^k B/n < D_i \leq x + b^{k+1} B/n\}}{\Pr \{x < D < b^k B/n\}} > \delta_2 \right\} \\ \leq \sum_{k=0}^{K-1} \frac{1}{n} \text{var} \left( \frac{|V| 1 \{x + b^k B/n < D_i \leq x + b^{k+1} B/n\}}{\Pr \{x < D < x + b^k B/n\}} \delta_2 \right) \\ + \max_{0 \leq k < K} E \left( \frac{|V| 1 \{x + b^k B/n < D \leq x + b^{k+1} B/n\}}{\Pr \{x < D < x + b^k B/n\}} \delta_2 \right) \\ \leq \sum_{k=0}^{K-1} \frac{1}{n} \frac{C b^k (b - 1) B/n}{(\delta_2 m b^k B/n)^2} + E[V^2] (b - 1) / \delta_2 m \\ = \frac{C (b - 1)}{(\delta_2 m)^2 (1 - b^{-1}) B} + E[V^2] (b - 1) / \delta_2 m.$$

Therefore, we can choose  $b$  close enough to 1 and then  $B$  large enough so that these two quantities and (A.3) can be made arbitrarily small for any  $\delta_1, \delta_2 > 0$ .  $\square$



*Proof of Lemma 5.1.* We first show that  $E \sup_{\theta \in \Theta_\Delta} [I_{1i}^+(\theta) - I_{2i}^+(\theta) | D_i] < 0$  for  $\Delta$  sufficiently small. Let  $\varepsilon > 0, 1/2 < \xi < 1$ , and  $0 < \eta < (4\varepsilon\xi - 2\varepsilon) / (1 + 4\varepsilon\xi - 2\varepsilon)$ . In view of Assumption 2, Choose  $\Delta$  to be small enough so that  $|W_i'(\beta_0 - \beta) + Z_i'(\delta_0 - \delta)| \leq a < \infty$  for all  $\theta \in \Theta_\Delta$  with probability  $1 - \eta$ , for  $a > 0$  such that  $\Pr\{U_i < a | X_i\} \leq 1/2 + \varepsilon$  and  $\Pr\{U_i > -a | X_i\} \leq 1/2 - \varepsilon$  almost surely and that  $\Pr\{A_{i,a} | A_i, D_i\} = \xi > 1/2$ , where

$$\begin{aligned} A_i &= 1 \{W_i'\beta > 0, W_i'\beta + Z_i'\delta \leq 0\} \\ A_{i,a} &= 1 \{W_i'\beta > 0, W_i'\beta + Z_i'\delta \leq -2a < 0\}. \end{aligned}$$

Then for all  $\theta \in \Theta_\Delta$ , we have

$$\begin{aligned} E(I_{1i}^+(\theta) | A_i, D_i) &\leq \eta + (1 - \eta) ((1/2 - \varepsilon)\xi + (1/2 + \varepsilon)(1 - \xi)) \\ E(I_{2i}^-(\theta) | A_i, D_i) &\geq (1 - \eta) ((1/2 + \varepsilon)\xi + (1/2 - \varepsilon)(1 - \xi)), \end{aligned}$$

almost surely, which in turn yields that

$$\begin{aligned} E[I_{1i}^+(\theta) - I_{2i}^-(\theta) | A_i, D_i] &\leq \eta + (1 - \eta) (-2\varepsilon\xi + 2\varepsilon(1 - \xi)) \\ &= \eta + (1 - \eta) (2\varepsilon - 4\varepsilon\xi) \\ &< 0 \end{aligned}$$

almost surely. Similarly,  $E \sup_{\theta \in \Theta_\Delta} [I_{3i}^+(\theta) - I_{4i}^+(\theta) | D_i] < 0$  almost surely for  $\Delta$  sufficiently small. Therefore,  $\sup_{\theta \in \Theta_\Delta} E[\zeta_i^+(\theta) | D_i] < 0$  almost surely for  $\Delta$  sufficiently small. The proof for  $\zeta_i^-(\theta)$  is similar.  $\square$

*Proof of Lemma 5.2.* It follows from (5.1) that

$$S_{n3}(\theta, \gamma) = \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} [\zeta_i^+(\theta) - \zeta_i^+(\theta_0)] + \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} [\zeta_i^-(\theta) - \zeta_i^-(\theta_0)].$$

Since  $\zeta_i^+(\theta)$  and  $\zeta_i^-(\theta)$  consists of just several indicator functions, there exists a universal constant  $C < \infty$  such that

$$|S_{n3}(\theta, \gamma)| \leq C \left[ \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} + \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} \right]$$

uniformly over  $\theta$ . Thus, it suffices to show that for every  $A > 0$ ,

$$(A.4) \quad \sup_{\gamma: n|\gamma - \gamma_0| \leq A} \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} = O_p(1) \quad \text{and} \quad \sup_{\gamma: n|\gamma - \gamma_0| \leq A} \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} = O_p(1).$$

To show the first claim of (A.4), consider a class of functions indexed by  $\gamma$ ,  $\mathcal{M}_A = \{1\{\gamma_0 < D \leq \gamma\} : n|\gamma - \gamma_0| \leq A\}$ . Then  $\mathcal{M}_A$  is a VC class of functions with an envelope function  $M_A = 1\{\gamma_0 < D \leq \gamma_0 + A/n\}$ . By Theorem 2.14.1 of Van der Vaart and Wellner (1996, p.239), there exists a universal constant  $C < \infty$  such that

$$E \left[ \sup_{\gamma: n|\gamma - \gamma_0| \leq A} \left| \sum_{i=1}^n [1\{\gamma_0 < D_i \leq \gamma\} - E1\{\gamma_0 < D \leq \gamma\}] \right| \right] \leq Cn^{1/2} J(1, \mathcal{M}_A) (EM_A^2)^{1/2},$$

where  $J(1, \mathcal{M}_A)$  is the uniform entropy integral defined in Van der Vaart and Wellner (1996, p.239). Since  $J(1, \mathcal{M}_A)$  is bounded for a VC class and  $(EM_A^2)^{1/2} = O(n^{-1/2})$ ,

$$\sup_{\gamma: n|\gamma - \gamma_0| \leq A} \left| \sum_{i=1}^n [1\{\gamma_0 < D_i \leq \gamma\} - E1\{\gamma_0 < D \leq \gamma\}] \right| = O_p(1).$$

The desired result follows since

$$(A.5) \quad \sup_{\gamma: n|\gamma - \gamma_0| \leq A} E1\{\gamma_0 < D \leq \gamma\} = O(n^{-1}).$$

The proof of the second claim of (A.4) is similar. □

*Proof of Lemma 6.1.* The proof of this lemma is similar to that of Lemma 5.2. It suffices to show that

$$(A.6) \quad \sum_{i=1}^n 1\{\gamma_0 < D_i \leq \gamma\} [\zeta^+(\theta) - \zeta^+(\theta_0)] = O_p(n^{-1/6})$$

and

$$(A.7) \quad \sum_{i=1}^n 1\{\gamma < D_i \leq \gamma_0\} [\zeta^-(\theta) - \zeta^-(\theta_0)] = O_p(n^{-1/6}).$$

uniformly over  $(\theta, \gamma)$  satisfying  $n^{1/3}|\theta - \theta_0| \leq A$  and  $n|\gamma - \gamma_0| \leq A$ . We will prove only (A.6) since the proof of (A.7) is similar. Consider a class of functions indexed by  $(\theta, \gamma)$ ,  $\mathcal{L}_A = \{1\{\gamma_0 < D \leq \gamma\} [\zeta^+(\theta) - \zeta^+(\theta_0)] : n^{1/3}|\theta - \theta_0| \leq A \text{ and } n|\gamma - \gamma_0| \leq A\}$ . Then  $\mathcal{L}_A$  is a VC

class of functions with an envelope function  $L_A = \sup_{\theta, \gamma} \{1\{\gamma_0 < D \leq \gamma\} [\zeta^+(\theta) - \zeta^+(\theta_0)] : n^{1/3}|\theta - \theta_0| \leq A \text{ and } n|\gamma - \gamma_0| \leq A\}$ . Then as in the proof of Lemma 5.2, (A.6) follows from Theorem 2.14.1 of Van der Vaart and Wellner (1996, p.239), the fact that  $(EL_A^2)^{1/2} = O(n^{-1/2}n^{-1/6})$ , (A.5), and

$$\sup_{\theta: n^{1/3}|\theta - \theta_0| \leq A} E [\zeta^+(\theta) - \zeta^+(\theta_0)] = O(n^{-1/6}).$$

□

*Proof of Lemma 6.2.* Since  $\tilde{S}_{n1}^+$  and  $\tilde{S}_{n1}^-$  are independent and similarly defined, we prove the convergence of  $\tilde{S}_{n1}^+$ . Define  $\tilde{S}_{n1}^+(\nu) = S_{n1}(\gamma_0 + n^{-1}\nu)$ . We first prove the tightness of  $\tilde{S}_{n1}^+(\nu)$ . By the  $D$ -tightness criterion of Billingsley (1968, equation (15.21)), it suffices to show that there exists a finite, universal constant  $C < \infty$  such that

$$(A.8) \quad E \left| \tilde{S}_{n1}^+(\nu) - \tilde{S}_{n1}^+(\nu_1) \right| \left| \tilde{S}_{n1}^+(\nu_2) - \tilde{S}_{n1}^+(\nu) \right| \leq C(\nu_2 - \nu_1)^2$$

for any  $\nu, \nu_1$  and  $\nu_2$  satisfying  $0 < \nu_1 < \nu < \nu_2$ . Note that

$$\begin{aligned} & E \left| \tilde{S}_{n1}^+(\nu) - \tilde{S}_{n1}^+(\nu_1) \right| \left| \tilde{S}_{n1}^+(\nu_2) - \tilde{S}_{n1}^+(\nu) \right| \\ & \leq 4E \sum_{i=1}^n \sum_{j=1}^n 1 \left\{ \gamma_0 + \frac{\nu_1}{n} < D_i \leq \gamma_0 + \frac{\nu}{n} \right\} 1 \left\{ \gamma_0 + \frac{\nu}{n} < D_j \leq \gamma_0 + \frac{\nu_2}{n} \right\} \\ & \leq 4E \sum_{i=1}^n \sum_{j=1, j \neq i}^n 1 \left\{ \gamma_0 + \frac{\nu_1}{n} < D_i \leq \gamma_0 + \frac{\nu_2}{n} \right\} 1 \left\{ \gamma_0 + \frac{\nu_1}{n} < D_j \leq \gamma_0 + \frac{\nu_2}{n} \right\} \\ & \leq 4n(n-1) \left[ F_D \left( \gamma_0 + \frac{\nu_2}{n} \right) - F_D \left( \gamma_0 + \frac{\nu_1}{n} \right) \right]^2 \\ & \leq 4C_{F_D} \frac{n(n-1)}{n^2} (\nu_2 - \nu_1)^2, \end{aligned}$$

where the first inequality comes from the fact  $|\zeta_i^+(\theta_0)| \leq 2$  for all  $i$ , the second inequality follows since  $1\{\gamma_1 < D_i \leq \gamma\}$  and  $1\{\gamma < D_j \leq \gamma_2\}$  are disjoint, and  $1\{\gamma_1 < D_i \leq \gamma_2\}$  is nondecreasing as  $\gamma_1$  decreases or as  $\gamma_2$  increases, the third inequality is obvious since  $D_i$  are i.i.d. with the common distribution  $F_D$ , and the fourth inequality follows from the assumption that  $F_D$  is Lipschitz continuous with the constant  $C_{F_D}$ . Then (A.8) follows immediately by taking  $C = 4C_{F_D}$ .

We now consider the weak convergence of the finite-dimensional distributions of  $\tilde{S}_{n1}^+(\nu)$ . To use the Cramér-Wold device, let  $0 < \nu_1 < \dots < \nu_J$  for  $J \in \mathbb{N}$  and  $q_1, \dots, q_J$  be constants. Instead of obtaining the weak convergence of  $\sum_{j=1}^J q_j \tilde{S}_{n1}^+(\nu_j)$ , we will show below the weak convergence of  $\bar{S}_{n1}$ , where  $\bar{S}_{n1}$  is defined as

$$\bar{S}_{n1} = \sum_{j=1}^J q_j \left[ \tilde{S}_{n1}^+(\nu_j) - \tilde{S}_{n1}^+(\nu_{j-1}) \right].$$

As shown in the proof of Theorem 4 of Pons (2003), it is easier to deal with  $\bar{S}_{n1}$ . Specifically, since  $\bar{S}_{n1}$  is a linear combination of  $\{\tilde{S}_{n1}^+(\nu_1), \dots, \tilde{S}_{n1}^+(\nu_J)\}$ , the weak convergence of the finite-dimensional distributions of  $\tilde{S}_{n1}^+(\nu)$  follows if we show that its characteristic function converges to the characteristic function of

$$\sum_{j=1}^J q_j \left[ \tilde{S}_1^+(\nu_j) - \tilde{S}_1^+(\nu_{j-1}) \right].$$

We now consider the characteristic function of  $\bar{S}_{n1}$ , which has the form

$$\varphi_{\bar{S}_{n1}}(t) = \left\{ E \left[ \exp \left( it \sum_{j=1}^J q_j I_{nj}(D) \zeta(\theta_0) \right) \right] \right\}^n.$$

Note that as in the proof of Theorem 4 of Pons (2003), using the fact that for each  $i$ , there is at most one index  $j$  such that  $I_{nj}(D_i) \neq 0$ , write

$$\exp \left( it \sum_{j=1}^J q_j I_{nj}(D) \zeta(\theta_0) \right) = 1 + \sum_{j=1}^J [\exp(itq_j I_{nj}(D) \zeta(\theta_0)) - 1],$$

so that

$$\begin{aligned} & E \left[ \exp \left( it \sum_{j=1}^J q_j I_{nj}(D) \zeta(\theta_0) \right) \right] \\ &= 1 + \sum_{j=1}^J n^{-1} [(\nu_j - \nu_{j-1}) f_D(\gamma_0) \{ E [\exp(itq_j \zeta(\theta_0)) - 1 | D = \gamma_0^+] \} + o(1)]. \end{aligned}$$

Then using the fact that  $A^n = \exp(n \log A)$  for any  $A > 0$  and that  $\log(1 + a_n) = a_n + o(a_n)$  for any  $a_n \rightarrow 0$ , we have

$$\lim_{n \rightarrow \infty} \varphi_{\bar{S}_{n1}}(t) = \sum_{j=1}^J (\nu_j - \nu_{j-1}) f_D(\gamma_0) \{ E [\exp(itq_j \zeta(\theta_0)) - 1 | D = \gamma_0^+] \}.$$

Then the lemma follows from the fact that the characteristic function  $\varphi_{\tilde{S}_1(s)}(t)$  of  $\tilde{S}_1(s)$  is  $\varphi_{\tilde{S}_1(s)}(t) = \exp[sf_D(\gamma_0) \{\varphi^+(t) - 1\}]$ , as in Lemma 5 of Pons (2003).  $\square$

## References

- Jason Abrevaya and Jian Huang. On the bootstrap of the maximum score estimator. *Econometrica*, 73(4):1175–1204, 2005.
- Donald W.K. Andrews and Patrik Guggenberger. The limit of finite sample size and a problem with subsampling. UCLA Economics Online Papers 372, UCLA Department of Economics, January 2006. available at <http://ideas.repec.org/p/cla/uclaol/372.html>.
- Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- Bryan W Brown and Mary Beth Walker. The random utility hypothesis and inference in demand systems. *Econometrica*, 57(4):815–29, July 1989.
- K. S. Chan. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21:520–533, 1993.
- Christopher Cox. Threshold dose-response models in toxicology. *Biometrics*, 43:3, 1987.
- John K. Dagsvik and Anders Karlstrom. Compensating variation and hicksian choice probabilities in random utility models that are nonlinear in income. *Review of Economic Studies*, 72(1):57–76, 2005.
- Miguel A. Delgado and Javier Hidalgo. Nonparametric inference on structural breaks. *Journal of Econometrics*, 96(1):113–144, 2000.
- Miguel A. Delgado, Juan M. Rodriguez-Poo, and Michael Wolf. Subsampling inference in cube root asymptotics with an application to manski’s maximum score estimator. *Economics Letters*, 73(2):241–250, November 2001.

- Kristin J. Forbes and Roberto Rigobon. No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance*, 57(5):2223–2261, 2002.
- Jesus Gonzalo and Michael Wolf. Subsampling inference in threshold autoregressive models. *Journal of Econometrics*, 2005.
- Bruce E. Hansen. Sample splitting and threshold estimation. *Econometrica*, 68:575–603, 2000.
- Joseph A. Herriges and Catherine L. Kling. Nonlinear income effects in random utility models. *Review of Economics and Statistics*, 81:62–72, 1999.
- Joel L. Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531, 1992.
- Joel L. Horowitz. Optimal rates of convergence of parameter estimators in the binary response model with weak distributional assumptions. *Econometric Theory*, 9(1):1–18, 1993.
- Jeankyung Kim and David Pollard. Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219, 1990.
- Michael R. Kosorok and Rui Song. Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*, 35(3), 2007.
- Hira L. Koul and Lianfen Qian. Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *Journal of Statistical Planning and Inference*, 108(1-2): 99–119, 2002.
- Stephen M.S. Lee and M.C. Pun. On m out of n bootstrapping for nonstandard m-estimation with nuisance parameters. *Journal of the American Statistical Association*, 101:1185–1197, 2006.

- Charles F. Manski. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228, 1975.
- Charles F. Manski. Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3):313–333, 1985.
- Charles F. Manski. Identification of binary response models. *Journal of the American Statistical Association*, 83(403):729–738, 1988.
- Whitney K. Newey and Daniel McFadden. Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2, pages 2111–2245. North-Holland, Amsterdam, 1994.
- Roberto Pastor and Eliseo Guallar. Use of tow-segmented logistic regression to estimate change-points in epidemiologic studies. *American Journal of Epidemiology*, 148:7, 1998.
- Roberto Pastor-Barriuso, Eliseo Guallar, and Jeseff Coresh. Transition models for change-point estimation in logistic regression. *Statistics in Medicine*, 22, 2003.
- M. Hashem Pesaran and Andreas Pick. Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control*, in press.
- Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer Series in Statistics. Springer-Verlag, New York, 1999.
- Odile Pons. Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, 31(2):442–463, 2003.
- Pamela F. Schwartz, Chris Gennings, and Vernon M. Chinchilli. Threshold models for combination data from reproductive and development experiments. *Journal of the American Statistical Association*, 90:431, 1995.
- Myunghwan Seo and Oliver Linton. A smoothed least squares estimator for the threshold regression. *Journal of Econometrics*, 2007. forthcoming.

Nobuhiko Terui and Wirawan Dony Dahana. Estimating heterogeneous price thresholds. *Marketing Science*, 25:384–391, 2006.

Howell Tong. *Nonlinear time series*, volume 6 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1990.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Process*. Springer, New York, 1996.