

Nonparametric inference on structural breaks

Miguel A. Delgado^a, Javier Hidalgo^{b,*}



^a*Departamento de Estadística y Econometría, Universidad Carlos III, cl Madrid 126-128, Getafe, 8403 Madrid, Spain*

^b*Department of Economics, TGO London School of Economics and Political Science, Houghton Street, London WC2A2AE, UK*

Received 1 January 1998; received in revised form 1 April 1999

Abstract

This paper proposes estimators of location and size of structural breaks in a, possibly dynamic, nonparametric regression model. The structural breaks can be located at given periods of time and/or they can be explained by the values taken by some regressor, as in threshold models. No previous knowledge of the underlying regression function is required. The paper also studies the case in which several regressors explain the breaks. We derive the rate of convergence and provide Central Limit Theorems for the estimators of the location(s) and size(s). A Monte Carlo experiment illustrates the performance of our estimators in small samples. © 2000 Published by Elsevier Science S.A. All rights reserved.

JEL classification: C14; C32

Keywords: Nonparametric regression; Dynamic models; Structural breaks; One-sided kernels

1. Introduction

This paper proposes estimators of the location(s) and size(s) of jump(s) in a, possibly dynamic, nonparametric multiple regression model where the jumps

* Corresponding author.

are located at given periods of time and/or are explained by the values of some regressors, as in threshold models. Regressors can be stochastic and/or fixed. One of the main features of our estimators is that their asymptotic distribution is normal, in contrast to rival parametric ones, which are, in general, not distribution free.

There is a vast literature on testing for the presence of a structural break when the possible timing of the break is unknown. See for instance, Quandt (1960), Hinkley (1969, 1970), Brown et al. (1975), Hawkins (1977), Worsley (1979), Kim and Siegmund (1989), Andrews (1993) and Andrews and Ploberger (1994). When a break exists, there is also some work on the estimation of the location of the break point, e.g. Feder (1975), Yao (1987), Eubank and Speckman (1994) and Bai (1994). The latter are typically based on a quadratic loss function, although robust estimators have been considered by Bai (1995), Antoch and Huskova (1997) and Fiteni (1998), among others. The previous work was done in a parametric framework.

The parametric approach has two potential drawbacks. First, the asymptotic distribution of the estimators (location) typically depend on certain unknown features of the data generating process. Although this problem has been circumvented by assuming that the size of the break shrinks to zero as the sample size increases, it appears that when the regressors are nonstationary (e.g. Bai and Perron, 1998; Hansen, 1998), the estimators are still not distribution free. Second, the asymptotic properties of the estimators depend on a correct specification of the model, for example on the underlying regression function in that a bad specification will induce inconsistent estimates. In addition as Hidalgo (1995) noticed, a poor specification may lead to the conclusion that there is a break when there is no one. Thus the objective of this paper is to propose and examine estimators of the location of the break, which are free from misspecification of the underlying regression model and distribution free without resorting to the artificial device of assuming that the size of the break shrinks to zero as the sample size increases.

Several articles have looked at inferences on changing points in nonparametric trend models. Yin (1988) has proposed strong consistent estimates of the number, location and sizes of jumps in the mean of a random variable, estimating the right and left limits of the regression function by means of uniform kernels non centered at zero (that is, moving averages). Müller (1992) has provided rates of convergence in \mathbb{L}_p and a Central Limit Theorem (CLT) for the estimators of the location and size of the structural break, whereas Chu and Wu (1993) have proposed a test for the number of jumps in a regression model with fixed design, providing CLTs of the estimators of their locations and sizes.

In this paper, in contrast to the previous works, by allowing the regression function to depend on more than one regressor, either stochastic (weakly dependent) or fixed, several additional features are introduced. First, we address the issue of one or more regressors defining the structural break point, see

Müller and Song (1994) for some related work. Second, the asymptotic variance depends on the point at which the function is estimated and, thus, its efficiency hinges on that estimation point. Then, we obtain the asymptotic efficiency bound, and we are able to propose a feasible estimator which achieves such a bound. Finally, the inclusion of lagged dependent variables, when a structural break occurs at a given moment of time, introduces a nontrivial problem on how to obtain efficient estimators of the break(s) and jump(s).

The remainder of the paper is organized as follows. In the next section, we present the estimation method. Section 3 discusses the asymptotic properties of estimators of the location and size of structural breaks when the regressors are strictly stationary. In Section 4, we study the situation where lagged dependent variables are present in the regression model and the break is explained by the regressor “time”. This corresponds to the situation of nonstationary, although stable, regressors. Finally, in Section 5, we show some Monte Carlo simulations. Proofs are confined to the appendices.

2. Estimating the locations and sizes of structural breaks

Let $\{(Y_1, X_1), (Y_2, X_2), \dots, (Y_T, X_T)\}$ be observations of a $(p + 1)$ -dimensional stochastic process where Y_t is scalar and $X_t = (X_{t1}, X_{t2}, \dots, X_{tp})'$ has its support in $\mathcal{X} \subseteq \mathbb{R}^p$, that is, $\Pr (X_t \in \mathcal{X}) = 1$. Define the regression function as $E(Y_t | X_t) = m(Z_t)$, where $Z_t = (Z_{t1}, Z_{t2}, \dots, Z_{tp}, Z_{t(p+1)})'$, $Z_{tr} = X_{tr}$, for $r \leq p$, $Z_{t(p+1)}$ is the regressor “time”, and the regression function $m(\cdot)$ is left unspecified.

Initially, assume that $m(\cdot)$ has N structural breaks explained by the r th regressor. That is,

$$m(z) = g(z) + \sum_{j=1}^N \alpha_0^{(j)} 1(z_r \geq \zeta_0^{(j)}), \quad (1)$$

where $1(A)$ is the indicator function of the event A , $z = (z_1, z_2, \dots, z_{(p+1)})'$ and $g(\cdot)$ is a generic continuous function. As with other problems involving trends, we define the regressor time as $Z_{t(p+1)} = t/T$. This artificial device is commonly introduced to provide justification of asymptotic statistical inference procedures. Like in parametric problems, the statistical properties of the structural break point estimator can only be derived when there is an infinite amount of information before and after the structural break points, $\zeta_0^{(j)}$, $j = 1, \dots, N$. Thus, when the regressor explaining the break is “time”, $\zeta_0^{(j)}$ becomes the proportion of the sample where the j th break have occurred. If there were other fixed regressors, we would assume that they become dense in their domain of definition as the sample size increases, as is the case with t/T . However, in what follows, we will assume, without loss of generality, that time is the only fixed regressor.

Obviously, $\zeta_0^{(j)}$, $j = 1, \dots, N$, are only identified when they are interior points in the domain of the r th regressor. In particular, when “time” is the regressor explaining the break, $\zeta_0^{(j)} = 0$ or $\zeta_0^{(j)} = 1$ are not identifiable. Suppose that we fix all the coordinates except the r th one, and define $z_0(\zeta) = (z_{01}, z_{02}, \dots, z_{0(r-1)}, \zeta, z_{0(r+1)}, \dots, z_{0(p+1)})'$, then consider the objective function

$$\Psi_0(\zeta)^2 = (m_0^+(z_0(\zeta)) - m_0^-(z_0(\zeta)))^2,$$

where $m_0^\pm(z_0(\zeta)) = \lim_{\delta \rightarrow 0^\pm} m(z_0(\zeta) + \delta)$. Thus, $\Psi_0(\zeta) = 0$ for all $\zeta \neq \zeta_0^{(j)}$ and $\Psi_0(\zeta) = \alpha_0^{(j)}$ if $\zeta = \zeta_0^{(j)}$. That clearly motivates our choice of $\Psi_0(\zeta)^2$. Assuming that $z_0^{(j)} \equiv z_0(\zeta_0^{(j)})$, $j = 1, \dots, N$, are interior points of $\mathcal{X} \times [0, 1]$ and, without loss of generality, that $|\alpha_0^{(j)}| > |\alpha_0^{(j+1)}|$, $j = 1, \dots, N - 1$, we have that

$$\zeta_0^{(j)} = \arg \max_{\zeta \in Q^{(j)}} \Psi_0(\zeta)^2,$$

where $Q^{(j)} = Q - \cup_{k=1}^{j-1} [\zeta_0^{(k)} - \varepsilon, \zeta_0^{(k)} + \varepsilon]$, with Q a compact subset in the domain of the r th regressor, and $\varepsilon > 0$ arbitrarily small. That is, the break points $\zeta_0^{(j)}$ are sequentially obtained. It is noteworthy observing that $\arg \max_{\zeta \in Q^{(j)}} \Psi_0(\zeta)^2 = \arg \max_{\zeta \in Q^{(j)}} |\Psi_0(\zeta)|^\beta$ for any $\beta \geq 1$.

Due to the nature of our problem, one sided kernel estimators prove to be very useful for the estimation of $m_0^\pm(\cdot)$. Those kernels were designed to estimate curves at boundary points (see Rice, 1984; Gasser et al., 1985), and are such that all their mass are at the right or left of zero. That is, $m_0^\pm(z)$ are estimated by

$$\hat{m}^\pm(z) = \hat{P}^\pm(z) / \hat{f}^\pm(z),$$

where $\hat{P}^\pm(z) = (Ta^{p+1})^{-1} \sum_{t=1}^T Y_t K^\pm[(Z_t - z)/a]$, $\hat{f}^\pm(z) = (Ta^{p+1})^{-1} \sum_{t=1}^T K^\pm[(Z_t - z)/a]$ and $K^\pm(u) = k^\pm(u_r) \prod_{j \neq r}^p k(u_j)$, with $k(\cdot)$ a symmetric kernel function and $k^\pm(\cdot)$ are one sided kernels. Finally, $a = a(T)$ is a sequence of bandwidths converging to zero as the sample size increases to infinity. More specific conditions on the kernel functions and rates of convergence of a to zero will be given in the next section. Thus, the break points $\zeta_0^{(j)}$, $j = 1, \dots, N$, are sequentially estimated by

$$\zeta^{(j)} = \arg \max_{\zeta \in Q^{(j)}} \hat{\Psi}_0(\zeta)^2, \quad j = 1, \dots, N,$$

where $Q^{(j)} = Q - \cup_{k=1}^{j-1} [\hat{\zeta}^{(k)} - 2a, \hat{\zeta}^{(k)} + 2a]$, with $\hat{\Psi}_0(\zeta) = \hat{m}^+(z_0(\zeta)) - \hat{m}^-(z_0(\zeta))$ and $\alpha_0^{(j)}$ is estimated by

$$\hat{\alpha}^{(j)} = \hat{\Psi}_0(\hat{\zeta}^{(j)}).$$

Next, consider the case where jumps occur simultaneously in several regressors. That is, the regression model (1) becomes

$$m(z) = g(z) + \sum_{\ell=1}^M \alpha_{0\ell} 1(z_\ell \geq \zeta_{0\ell}), \quad (2)$$

assuming, for notational convenience, that the first M regressors explain the break. Since in (2), left and right limits of $m(z)$ are different, in the direction of different coordinates, we can employ the jump functions $\Psi_\ell(\zeta_\ell) = m_\ell^+(z_{0\ell}(\zeta_\ell)) - m_\ell^-(z_{0\ell}(\zeta_\ell))$, where for $\ell = 1, \dots, M$, $m_\ell^\pm(z_{0\ell}(\zeta_\ell)) = \lim_{\delta \rightarrow 0_\pm} m(z_{0\ell}(\zeta_\ell + \delta))$ and $z_{0\ell}(\zeta_\ell) = (z_{01}, z_{02}, \dots, z_{0(\ell-1)}, \zeta_\ell, z_{0(\ell+1)}, \dots, z_{0(p+1)})$. Thus, we estimate $m_\ell^\pm(z_{0\ell}(\zeta_\ell)) = P_\ell^\pm(z_{0\ell}(\zeta_\ell))/f(z_{0\ell}(\zeta_\ell))$ by $\hat{m}_\ell^\pm(z_{0\ell}(\zeta_\ell)) \equiv \hat{P}_\ell^\pm(z_{0\ell}(\zeta_\ell))/\hat{f}_\ell^\pm(z_{0\ell}(\zeta_\ell))$, where $\hat{P}_\ell^\pm(\cdot)$ and $\hat{f}_\ell^\pm(\cdot)$ employ the one-sided kernels $k^\pm(t)$ for the ℓ th coordinate, the symmetric kernel $\tilde{k}(t) = [k^+(t) + k^-(t)]/2$ for the other $M - 1$ coordinates and the kernel $k(t)$ for the remaining $p + 1 - M$ coordinates. This is done to prevent the possible effects that discontinuities in other directions may have on the estimation of $\Psi_\ell(\zeta_\ell)$. Thus, assuming that $z_{0\ell} \equiv z_{0\ell}(\zeta_{0\ell})$, $\ell = 1, \dots, M$, are interior points of $\mathcal{X} \times [0, 1]$, $\zeta_0 = (\zeta_{01}, \dots, \zeta_{0M})'$ is estimated by

$$(\hat{\zeta}_1, \dots, \hat{\zeta}_M)' = \hat{\zeta} = \arg \max_{\zeta = (\zeta_1, \dots, \zeta_M)' \in \mathcal{Q}_{(M)}} \sum_{\ell=1}^M \hat{\Psi}_\ell(\zeta_\ell)^2,$$

where $\hat{\Psi}_\ell(\zeta_\ell) = \hat{m}_\ell^+(z_{0\ell}(\zeta_\ell)) - \hat{m}_\ell^-(z_{0\ell}(\zeta_\ell))$ and $\mathcal{Q}_{(M)} = X_{\ell=1}^M \mathcal{Q}_{\ell(M)}$ is a compact subset in the domain of the first M regressors. Once the estimator of ζ_0 has been obtained, the ℓ th coordinate of $\alpha_0 = (\alpha_{01}, \dots, \alpha_{0M})'$ is estimated by $\hat{\alpha}_\ell = \hat{\Psi}_\ell(\hat{\zeta}_\ell)$.

3. Asymptotic properties of estimators of the break(s) and jump(s) with stationary regressors

In this section we will focus on the case where the regressors in model (1) are strictly stationary, leaving the nonstationary case for the next section. We can envisage several situations where this is the case, as thresholds models or when the regressors X_t are exogenous. The following definitions are useful.

Definition 1. Let \mathcal{M}_a^b be the σ -algebra generated by $\{X_t, a \leq t \leq b\}$. $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is a strictly stationary α -mixing stochastic process with mixing coefficients $\zeta(m)$ if $\lim_{m \rightarrow \infty} \zeta(m) = 0$, where $\zeta(m) = \sup_{\mathcal{A} \in \mathcal{M}_{-\infty}^{-m}, \mathcal{B} \in \mathcal{M}_{m+\infty}^{\infty}} |\Pr(\mathcal{A} \cap \mathcal{B}) - \Pr(\mathcal{A}) \Pr(\mathcal{B})|$.

The next three definitions are borrowed from Robinson (1988).

Definition 2. $\mathfrak{F}_r(r \geq 1)$ is the class of real functions satisfying

$$\int_{-\infty}^{\infty} u^i k(u) du = \delta_{i0} \quad (i = 0, 1, \dots, r - 1)$$

$$k(u) = O((1 + |u|^{r+1+\xi})^{-1}) \quad \text{for some } \xi > 0,$$

where δ_{i0} is the Kronecker's delta.

Let u be a generic random variable.

Definition 3. $\Xi^\alpha = \{\phi: \mathbb{R} \rightarrow \mathbb{R}; \mathbb{E} |\phi(u)|^\alpha < \infty\}$.

Definition 4. \mathcal{X}_μ^α , $\alpha > 0$, $\mu > 0$, is the class of $\phi(\cdot)$ functions belonging to Ξ^α such that there exist $h \in \Xi^\alpha$ and some $\gamma > 0$ such that

$$\sup_{S_{\mu\gamma}} \frac{|\phi(u) - \phi(v) - Q(u, v)|}{\|u - v\|^\mu} < h(u) \quad \text{a.e. } (u),$$

where $S_{\mu\gamma} = \{v: v \in \mathbb{R} \text{ and } \|u - v\| < \gamma\}$ for all $u \in \mathbb{R}$, $Q(u, v)$ is the Taylor expansion of $\phi(\cdot)$ up to $m - 1 < \mu \leq m$, and where the coefficients in the Taylor expansion belongs to Ξ^α .

Consider the following assumptions on the data generating process:

- A1. (a) $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is a strictly stationary α -mixing process where the mixing coefficients $\zeta(m)$ satisfy $\sum_{m=q}^\infty \zeta(m)^{\delta/(2+\delta)} = O(q^{-1})$ and (b) $\mathbb{E} \|m(Z_t)\|^{\delta+2} < \infty$ for some $\delta > 0$.
- A2. Let $\varepsilon_t = Y_t - m(Z_t)$. Then, (a) $\mathbb{E}(\varepsilon_t | X_s, s \leq t) = 0$, (b) $\mathbb{E}(\varepsilon_t | \varepsilon_s, s < t) = 0$ and (c) $\mathbb{E}(\varepsilon_t^{2+\nu}) < \infty$ for some $\nu > 0$.
- A3. The pdf of $X_t, f(x)$, belongs to \mathcal{X}_μ^∞ , and $f(x)g(z)$ belongs to \mathcal{X}_μ^α for some $\mu \geq 2$ and $\alpha > 2$.
- A4. $\mathbb{E}(\varepsilon_t^2 | Z_t) = \sigma^2(Z_t)$ with $\sigma^2(z) = s(z) + \rho_0^{(j)} 1(z_r \geq z_r^{(j)})$, where $s(z)f(z) \in \mathcal{X}_\mu^\alpha$ for some $\mu \geq 2$ and $\alpha > 2$, and $\rho_0^{(j)}$ s are fixed numbers.

Assumption A1 is common in nonparametric estimation with α -mixing data and provides minimal conditions on the rate of convergence of $\zeta(m)$ to zero (see Robinson, 1983). Although it allows for dynamic models as threshold models, it does not allow situations where the break is given at a date and we have lagged dependent variables as regressors, since X_t would not be stationary. This case will be deferred to the next section. Assumption A2 can be relaxed, allowing correlation, but then, like in parametric models, no lagged dependent variables can be present in the regression function. However, this assumption does not rule out conditional heteroskedasticity. The smoothness condition, in Assumption A3, is usually required in kernel regression estimation. Assumption A4 explicitly allows for conditional heteroskedasticity, which can be discontinuous at the structural break point. This is a realistic assumption, because it is natural to think that if the first conditional moment has a jump at a given point, the same can happen to the second conditional one.

Define $k_{(m)}^\pm(u) = \partial^m k^\pm(u) / \partial u^m$, $K_{(m)}^\pm(u) = k_{(m)}^\pm(u_r) \prod_{j \neq r}^{p+1} k(u_j)$, $K_{(0)}^\pm(u) = K^\pm(u)$ and $\xi_{(m)} = \int_{\mathbb{R}^{p+1}} K_{(m)}^\pm(u)^2 du$, $m = 0, 1, 2, \dots$. Let us introduce the following assumptions on the kernel functions:

- B1. The kernel function $k(\cdot) \in \mathfrak{F}_2$.

- B2. $k^\pm : \mathbb{R}^\pm \rightarrow \mathbb{R}$, where $k^+(u) = k^-(-u)$, $k_{(1)}^+(u) = -k_{(1)}^-(-u)$ and $k^+(u) \in \mathfrak{S}_2$.
 B3. $k^\pm(0) = k^\pm(\pm \infty) = 0$, with $k_{(1)}^+(0) > 0$.
 B4. $\int_{\mathbb{R}} |u^2 k_{(m)}^\pm(u)| du < \infty$, $\int_{\mathbb{R}} |k_{(m)}^\pm(u)| du < \infty$, $m = 1, 2$.
 B5. $\lim_{T \rightarrow \infty} (Ta^{p+1})^{-1} = 0$ and $\lim_{T \rightarrow \infty} Ta^{p+1+4} < \infty$.

Note that $p + 1$ is the dimension of Z_t . Henceforth, in all the Assumptions, Theorems and Corollaries, when the regressor time is not a component of Z_t , $p + 1$ should be replaced by p . B1 is a common assumption on kernels. Kernels $k^+(\cdot)$, and therefore $k^-(\cdot)$, satisfying Assumptions B2–B4, can be obtained from any function $h(u)$ with domain in \mathbb{R}^+ , as $k^+(u) = u(c_1 + c_2u)h(u)$, where the constants c_1 and c_2 are the solution to $\int_{\mathbb{R}^+} u(c_1 + c_2u)h(u) du = 1$ and $\int_{\mathbb{R}^+} u^2(c_1 + c_2u)h(u) du = 0$. As an example, let $h(u) = \exp(-u)1(u > 0)$, then, $k^+(u) = u(3 - u) \exp(-u)1(u > 0)$. Assumption B5 is satisfied by the bandwidth choice $a = CT^{-1/\beta}$, $p + 1 < \beta \leq p + 5$, where C is a finite constant independent of T .

The following theorem establishes the rate of convergence of $\hat{\zeta}^{(j)}$ to $\zeta_0^{(j)}$.

Theorem 1. Consider model (1). Under Assumptions A1–A4 and B1–B5,

$$\hat{\zeta}^{(j)} - \zeta_0^{(j)} = O_p((Ta^{p-1})^{-1/2}), \quad j = 1, \dots, N.$$

The rate of convergence depends on the bandwidth parameter and it is slower than in the parametric case, where T -consistency is achieved, see Chan (1993) or Bai (1994). However, this is not surprising, due to the local behavior of the statistics.

Next, we establish the asymptotic normality of the estimators, which uses the same strategy of proof as Eddy (1980) and Müller (1992). To that end, introduce $x_0(\zeta) = (x_{01}, x_{02}, \dots, x_{0(r-1)}, \zeta, x_{0(r+1)}, \dots, x_{0p})'$.

Theorem 2. Consider model (1) where $\zeta_0^{(j)}$, $j = 1, \dots, N$, are interior points of $Q^{(j)}$. Under Assumptions A1–A4 and B1–B5,

$$(a) (Ta^{p-1})^{1/2} [(\hat{\zeta}^{(1)} - \zeta_0^{(1)}), (\hat{\zeta}^{(2)} - \zeta_0^{(2)}), \dots, (\hat{\zeta}^{(N)} - \zeta_0^{(N)})]' \xrightarrow{d} N(0, \Sigma_0), \quad \text{and}$$

$$(b) (Ta^{p+1})^{1/2} [(\hat{\alpha}^{(1)} - \alpha_0^{(1)}), (\hat{\alpha}^{(2)} - \alpha_0^{(2)}), \dots, (\hat{\alpha}^{(N)} - \alpha_0^{(N)})]' \xrightarrow{d} N(0, \Omega_0),$$

where $\Sigma_0 = \text{diag}(V_{10}, V_{20}, \dots, V_{N0})$ and $\Omega_0 = \text{diag}(W_{10}, W_{20}, \dots, W_{N0})$, and

$$V_{k0} = \frac{[2s(z_0^{(k)}) + \rho_0^{(k)}] \zeta_{(1)}}{f(x_0^{(k)}) \alpha_0^{(k)2} k_{(1)}^-(0)^2} \quad \text{and} \quad W_{k0} = \frac{[2s(z_0^{(k)}) + \rho_0^{(k)}] \zeta_{(0)}}{f(x_0^{(k)})}, \quad k = 1, \dots, N,$$

where $x_0^{(k)} = x(\zeta_0^{(k)})$.

As was expected, the asymptotic variances of $\hat{\zeta}^{(j)}$, $j = 1, \dots, N$, decrease as the size of the jump increases in absolute value. Observe that, as the pdf of X_t evaluated at $x_0^{(j)}$ increases, the variance decreases and, hence, the asymptotic efficiency of the estimators depends on the choice of $x_0^{(j)}$. We will return to this point in Corollary 2 below. Finally, from Theorem 2(b), we note that the asymptotic distribution of $(Ta^{p+1})^{1/2}(\hat{\alpha}^{(j)} - \alpha_0^{(j)})$ is the same whether $\zeta_0^{(j)}$ or $\hat{\zeta}^{(j)}$ are used to estimate $\alpha_0^{(j)}$.

To construct asymptotically valid confidence intervals, $f(x)$ and $\sigma_{\pm}^2(z(\zeta)) = \lim_{\delta \rightarrow 0_{\pm}} \sigma^2(z(\zeta + \delta))$ are estimated by $\hat{f}(z) = 1/2(\hat{f}^+(z) + \hat{f}^-(z))$ and $\hat{\sigma}_{\pm}^2(z(\zeta)) = \hat{\lambda}^{\pm}(z(\zeta)) - \hat{m}^{\pm}(z(\zeta))^2$ respectively, where $\hat{\lambda}^{\pm}(z) = (Ta^{p+1})^{-1} \sum_{i=1}^T Y_i^{\pm} K^{\pm}[(Z_i - z)/a]/\hat{f}^{\pm}(z)$. Thus, we estimate $f(x_0^{(j)})$ and $2s(z_0^{(j)}) + \rho_0^{(j)} = \sigma_{+}^2(z_0^{(j)}) + \sigma_{-}^2(z_0^{(j)})$ by $\hat{f}(x_0(\hat{\zeta}^{(j)}))$ and $\hat{\sigma}_{+}^2(z_0(\hat{\zeta}^{(j)})) + \hat{\sigma}_{-}^2(z_0(\hat{\zeta}^{(j)}))$ respectively.

Some remarks are in order.

Remark 1. Model (1) can be generalized to situations where the breaks occur simultaneously in both the level and in some derivative. That is, consider

$$m(z) = g(z) + \sum_{j=1}^N \tilde{g}^{(j)}(z) 1(z_r \geq \zeta_0^{(j)}), \quad (3)$$

where the functions $g(z)$ and $\tilde{g}^{(j)}(z)$, $j = 1, \dots, N$, have two continuous derivatives in $\mathcal{X} \times [0, 1]$ and $\alpha_0^{(j)} = \tilde{g}^{(j)}(z_0^{(j)})$. One of the differences, compared with the situation studied in model (1), is due to the bias term $\tilde{g}_{(1)}^{(j)}(z_0^{(j)}) f(x_0^{(j)}) \nu a h^{-1}$, where $\tilde{g}_{(1)}^{(j)}(z) = \partial \tilde{g}^{(j)}(z) / \partial z_r$, which would appear as a third term of $E[\mathbb{P}^+(v)]$ in Proposition 1. Notice that in (1), $\tilde{g}_{(1)}^{(j)}(z) = 0$. Thus, to obtain consistency and asymptotic normality of the estimators, we need to strengthen the range of admissible values of a . In particular B5 should be replaced by $\lim_{T \rightarrow \infty} (Ta^{p+1})^{-1} + Ta^{p+1+2} = 0$. However, it is noteworthy that we can choose $a = CT^{-1/(p+2)}$, say, which is valid for models (1) and (3). The other difference is that the choice of z_0 would be given by $\arg \max_z \tilde{g}^{(j)}(z_0^{(j)})^2 f(x_0^{(j)})$ instead of $\arg \max_{x \in \mathcal{X}} f(x_0^{(j)})$, since here $\alpha_0^{(j)2} f(x_0^{(j)})$ becomes $\tilde{g}^{(j)}(z_0^{(j)})^2 f(x_0^{(j)})$ in the definition of V_{j_0} of Theorem 2.

Remark 2. In model (3) we can have that $\alpha_0^{(j)} = \tilde{g}^{(j)}(z_0^{(j)}) = 0$ although $\tilde{g}_{(1)}^{(j)}(z_0^{(j)}) \neq 0$, that is, the change point is in the first derivative of the regression function. Under this framework

$$m_{(1)}(z) = g_{(1)}(z) + \sum_{j=1}^N \tilde{g}_{(1)}^{(j)}(z) 1(z_r \geq \zeta_0^{(j)}),$$

where the functions $g_{(1)}(z)$ and $\tilde{g}_{(1)}^{(j)}(z)$, $j = 1, \dots, N$, have two continuous derivatives in $\mathcal{X} \times [0, 1]$. With $\alpha_0^{(j)} = \tilde{g}_{(1)}^{(j)}(z_0^{(j)})$, $\zeta_0^{(j)}$ is estimated by

$$\hat{\zeta}^{(j)} = \arg \max_{z \in \mathcal{Q}^{(j)}} \hat{\Psi}_{(1)0}(\zeta)^2, \quad j = 1, \dots, N,$$

where $\hat{\Psi}_{(1)0}(\zeta) = \partial \hat{\Psi}_0(\zeta) / \partial z_r$. In this case, see Delgado and Hidalgo (1995), $(\hat{\zeta}^{(j)} - \zeta_0^{(j)}) = O_p((Ta^{p-3})^{-1/6})$ instead of $O_p((Ta^{p-1})^{-1/2})$. That is, the location of a change in the derivative will, not surprisingly, be estimated with lower precision than that for the level. However, it is noteworthy that the same phenomenon is true in parametric models where only *root-T consistency* is achieved, see e.g. Feder (1975).

Remark 3. It seems reasonable to use different bandwidths for each explanatory variable or, at least, a different one for the variable responsible of the break, depending on the smoothness of the regression function with respect to each regressor. In this case condition B5 should read,

$$\left(T \prod_{i=1}^{p+1} a_i \right)^{-1} \rightarrow 0, \left(T \prod_{i=1}^{p+1} a_i \right) \max_{i=1, \dots, p+1} a_i^4 < \infty \text{ and } \max_i a_i \rightarrow 0.$$

The results follow straightforwardly, and if anything, it complicates unnecessarily the already complex notation. Obviously, in this case, the rate of convergence of the break point estimator becomes $(Ta_r^{-2} \prod_{i=1}^{p+1} a_i)^{1/2}$.

Introduce the following assumption.

A5. $E(\varepsilon_t^2 | Z_t) = \sigma^2(Z_t)$, with $\sigma^2(z) = s(z) + \sum_{j=1}^M \rho_{0j} 1(z_j \geq \zeta_{0j})$, where $s(z)f(z) \in \mathcal{X}_\mu^\alpha$ for some $\mu \geq 2$ and $\alpha > 2$, and ρ_{0j} are fixed numbers.

Corollary 1. Consider model (2), where $\zeta_{0\ell}$, $\ell = 1, \dots, M$, are interior points of $Q_{\ell(M)}$. Under Assumptions A1–A3, A5 and B1–B5,

$$(a) (Ta^{p-1})^{1/2} [(\hat{\zeta}_1 - \zeta_{01}), (\hat{\zeta}_2 - \zeta_{02}), \dots, (\hat{\zeta}_M - \zeta_{0M})]' \xrightarrow{d} N(0, \hat{\Sigma}_0), \text{ and}$$

$$(b) (Ta^{p+1})^{1/2} [(\hat{\alpha}_1 - \alpha_{01}), (\hat{\alpha}_2 - \alpha_{02}), \dots, (\hat{\alpha}_M - \alpha_{0M})]' \xrightarrow{d} N(0, \hat{\Omega}_0),$$

where $\hat{\Sigma}_0 = \text{diag}[\hat{V}_{01}, \hat{V}_{02}, \dots, \hat{V}_{0M}]$ and $\hat{\Omega}_0 = \text{diag}[\hat{W}_{01}, \hat{W}_{02}, \dots, \hat{W}_{0M}]$, and

$$\hat{V}_{0i} = \frac{[2s(z_{0i}) + \rho_{0i}] \xi_{(1)}}{f(x_{0i}) \alpha_{0i}^2 k_{(1)}(0)^2}, \hat{W}_{0i} = \frac{[2s(z_{0i}) + \rho_{0i}] \xi_{(0)}}{f(x_{0i})}, \quad i = 1, \dots, M,$$

where $\xi_{(1)} = \int k_{(1)}^+(u)^2 du \int \tilde{k}(u)^2 du)^{M-1} (\int k(u)^2 du)^{p+1-M}$ and $z_{0i} = (x'_{0i}, z_{0(p+1)i})'$.

As was mentioned after Theorem 2, the choice of z_{0j} , for $j \neq r$, $j = 1, \dots, p+1$, affects the efficiency of both $\hat{\zeta}^{(j)}$ and $\hat{\alpha}^{(j)}$. By inspection of that theorem, it is observed that the optimal choice is obtained when $f(x_0^{(j)}) / [2s(z_0^{(j)}) + \rho_0^{(j)}]$ is maximized. For simplicity, consider model (1) with $N = 1$ and homoskedasticity, so that Assumption A4 becomes

$$A4'. E(\varepsilon_t^2 | Z_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty.$$

Under A4', the asymptotic variances given in Theorem 2 become

$$V_{10} = \frac{2\sigma_{\varepsilon}^2 \xi(1)}{f(x_0^{(1)})\alpha_0^{(1)2}k_{(1)}(0)^2} \text{ and } W_{10} = \frac{2\sigma_{\varepsilon}^2 \xi(0)}{f(x_0^{(1)})},$$

and thus, to minimize the above variances is equivalent to maximize $f(x_0^{(1)})$. Since in most of the empirical examples, the regressor time is assumed to be responsible for the structural break, and also for notational simplicity, let us assume that “time” explains the break, that is, $r = p + 1$. Define $x^* = \arg \max_{x \in \mathcal{X}} f(x)$ (the mode of $f(x)$), which is estimated by $\hat{x}^* = \arg \max_{x \in \mathcal{X}} \hat{f}(x)$, where $\hat{f}(x)$ is the kernel density estimator of $f(x)$. Thus, an efficient (feasible) estimator of $\zeta_0^{(1)}$ is given by

$$\hat{\zeta}^* = \arg \max_{\zeta \in Q} \hat{\Psi}_*(\zeta)^2,$$

where $\hat{\Psi}_*(\zeta) = \hat{m}^+((\hat{x}^*, \zeta)Y) - \hat{m}^-((\hat{x}^*, \zeta)Y)$ and $Q = [c, d] \subset (0, 1)$. The following Corollary justifies this (two-step) efficient estimation procedure.

Corollary 2. Consider model (1) with $r = p + 1$ and $N = 1$. Under Assumptions A1–A3, A4' and B1–B5, with $z_0^{(1)} = (x^, \zeta_0^{(1)})'$,*

$$(Ta^{p-1})^{1/2}(\hat{\zeta}^* - \zeta_0^{(1)}) \xrightarrow{d} N(0, V_{10}) \text{ and } (Ta^{p+1})^{1/2}(\hat{x}^* - x_0^{(1)}) \xrightarrow{d} N(0, W_{10}).$$

When $r \neq p + 1$, we fix the values $\hat{x}_1^*, \hat{x}_2^*, \dots, \hat{x}_{r-1}^*, \hat{x}_{r+1}^*, \dots, \hat{x}_p^*$, which maximize $\hat{f}(x(\hat{\zeta}))$ and z_{p+1} is fixed at an arbitrary value, say $1/2$.

4. Regression models with lagged dependent variables with a break at a period of “time”

In this section, we discuss the situation, quite common in econometrics, where X_t contains lagged dependent variables and the break point is given at a point in time. Under this framework, one of the main differences, compared to the situation discussed in the previous section, is that the regressors are not stationary, that is A1 does not hold. In particular, the pdf of the regressors is different before and after the break point. As we will discuss below, this assumption is not vacuous, and to obtain more efficient estimates of the break and jump is more involved, and requires slight changes to the “objective function” in the “two-step” procedure. Given $X_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-q}, \tilde{X}_t, \tilde{X}_{t-1}, \dots, \tilde{X}_{t-\ell})'$, where $p = (\ell + 1)q_1 + q$ and q_1 is the dimension of \tilde{X}_t , consider the model

$$m(x) = g(x) + \alpha_0 1(\tau \geq \tau_0), \tag{4}$$

with $\tau = t/T$. Thus, $[T\tau_0]$ indicates the time of the break. Observe that we have only assumed one structural break and \tilde{X}_t to be stochastic. This is merely to

keep the notation and arguments simpler. Theorems 3 and 4 below are easily generalized to several structural breaks and/or \tilde{X}_t to contain fixed regressors, following the same strategy of Theorem 1 and 2 of Section 3. We need to introduce the following:

Definition 5. Let \mathcal{M}_a^b be the σ -algebra generated by $\{X_t, a \leq t \leq b\}$. $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is an α -mixing stochastic process with mixing coefficients $\zeta(m)$ if $\lim_{m \rightarrow \infty} \zeta(m) = 0$, where $\zeta(m) = \sup_s \sup_{\mathcal{A} \in \mathcal{M}_{-s}^s, \mathcal{B} \in \mathcal{M}_{s+m}^{\infty}} |\Pr(\mathcal{A} \cap \mathcal{B}) - \Pr(\mathcal{A}) \Pr(\mathcal{B})|$.

Consider the following assumptions on the data generating process:

- C1. (a) $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is an α -mixing process where $\zeta(m)$ satisfy $\sum_{m=q}^{\infty} \zeta(m)^{\delta/(2+\delta)} = O(q^{-1})$ and (b) $\sup_t E \|m(X_t)\|^{\delta+2} < \infty$ for some $\delta > 0$.
- C2. Define $\varepsilon_t = Y_t - m(X_t)$. Then, (a) $E(\varepsilon_t | X_s, s \leq t) = 0$, (b) $E(\varepsilon_t | \varepsilon_s, s < t) = 0$ and (c) $E(\varepsilon_t^{2+\nu}) < \infty$ for some $\nu > 0$.
- C3. The pdf's of X_t before and after the break point τ_0 , that is $f_1(x)$ and $f_2(x)$ respectively, belong to $\mathcal{X}_{\mu}^{\infty}$, and $f_i(z)g(z) \in \mathcal{X}_{\mu}^{\alpha}$, $i = 1, 2$, for some $\mu \geq 2$ and $\alpha > 2$.
- C4. $E(\varepsilon_t^2 | Z_t = (X'_t, \tau_t)) = \sigma^2(Z_t)$, with $\sigma^2(z) = s(z) + \rho_0 1(\tau \geq \tau_0)$, where $s(z) f_i(z) \in \mathcal{X}_{\mu}^{\alpha}$ for $i = 1, 2$, for some $\mu \geq 2$ and $\alpha > 2$, and ρ_0 is a fixed number.
- C5. τ_0 is an interior point of Q , a compact set in $(0, 1)$.
- B6. $(Ta^{p+1})^{-1} + Ta^{p+3} \rightarrow 0$ as $T \rightarrow \infty$.

Remark 3 (Cont.) As was said in Remark 3, if we allowed the bandwidth parameter to have different rates of convergence to zero for different coordinates, the theorems below would follow in the same manner. The only difference is that in this case Assumption B6 should read

$$\left(T \prod_{i=1}^{p+1} a_i \right)^{-1} + \left(T \prod_{i=1}^{p+1} a_i \right) a_i^2 + \max_i a_i \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Theorem 3. Consider model (4). Under C1–C5, B1–B4 and B6,

$$\hat{\tau} - \tau_0 = O_p((Ta^{p-1})^{-1/2}).$$

Theorem 4. Consider model (4). Under C1–C5, B1–B4 and B6,

$$(a) (Ta^{p-1})^{1/2}(\hat{\tau} - \tau_0) \xrightarrow{d} (U / (-\alpha_0 k_{(1)}^-(0))) \{dI(U \geq 0) + d^{-1}I(U < 0)\},$$

where $U \sim N(0, \Sigma_0)$, $\Sigma_0 = \xi_{(1)}(s(z_0)) / f_1(x_0) + (\rho_0 + s(z_0)) / f_2(x_0)$ and $d = f_1(x_0) / f_2(x_0)$.

$$(b) (Ta^{p+1})^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \xi_{(0)}(s(z_0)) / f_1(x_0) + (\rho_0 + s(z_0)) / f_2(x_0)).$$

The asymptotic distribution function of $\hat{\tau}$ is continuous with continuous derivative except at zero, where it has a jump reflecting that the *pdfs* of X_t are different before and after τ_0 , although it is easily tabulated. For instance, denote $(\hat{\alpha}k_{(1)}^-(0))^{-1}\hat{\Sigma} = \hat{\varrho}$. Then, an asymptotic 95 confidence interval for τ_0 is given by

$$(\hat{\tau} - (Ta^{p-1})^{-1/2}\hat{d}^{-1}\hat{\varrho}^{1/2}\mathcal{N}_{0,0.025}, \hat{\tau} + (Ta^{p-1})^{-1/2}\hat{d}\hat{\varrho}^{1/2}\mathcal{N}_{0,0.975})$$

where $\mathcal{N}_{0,0.025}$ and $\mathcal{N}_{0,0.975}$ are respectively the 0.025 and 0.975 quantiles of the standard normal and \hat{d} , $\hat{\alpha}$ and $\hat{\Sigma}$ are consistent estimators of d , α_0 and Σ_0 , respectively. Moreover, its first two moments are, respectively,

$$\frac{(\Sigma_0)^{1/2}}{-\alpha_0 k_{(1)}^-(0)\sqrt{2\pi}}(d - d^{-1}) \text{ and } \frac{\Sigma_0}{2(\alpha_0 k_{(1)}^-(0))^2} \left(\left(1 - \frac{1}{\pi}\right)(d^2 + d^{-2}) + \frac{2}{\pi} \right).$$

As discussed after Theorem 2 and before Corollary 2 of the previous section, the efficiencies of $\hat{\tau}$ and $\hat{\alpha}$ depend very much on where the kernel regression function is evaluated. As was done there, one possible way can be to evaluate the kernel function at the mode of the pdf of X_t , but in view of the nonstationarity of the data, this method has not much sense, so that the choice of that point is not so clear. Another possibility could be based on the *maximization* of, say, $f_1(x)$, but clearly this is not a good strategy either, since it might be that $f_2(x)$ is very small at the point $x^* = \arg \max_x f_1(x)$. For instance, suppose that $f_1(\cdot)$ is $N(0, 1)$ while $f_2(\cdot)$ is $N(20, 1)$. Then, $x^* = 0$, while $f_2(0)$ is extremely small and, thus the variances of our estimators will be very large, implying that the bigger the jump, the worse the estimator is. Hence, this method does not appear to be very desirable either. Another possibility can be to evaluate the kernel at a point x where $f_1(x) = f_2(x)$, or perhaps where $f_1(\cdot) + f_2(\cdot)$ is maximized, but as can be observed from the previous example, this method would suffer from the same problem and, thus, it would not render the desired results.

Therefore, we propose a two-step procedure with a slight modification of the objective function which will not suffer from the above drawbacks. The following assumptions are useful in order to derive the asymptotic properties of this estimator.

C4'. $E(\varepsilon_t^2|X_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty$.

C6. The pdfs of X_t , *that is* $f_1(\cdot)$ and $f_2(\cdot)$, are unimodal and symmetric around their means μ_1 and μ_2 , respectively.

Consider $\hat{\tau}$ and $\hat{\alpha}$ of Theorem 4, and the objective function

$$\tilde{\Phi}_0(\tau) = \hat{\alpha}(\hat{m}^+(\tilde{z}_1(\tau)) - \hat{m}^-(\tilde{z}_2(\tau))), \tag{5}$$

where $\tilde{z}_i(\tau) = (\tilde{\mu}'_i, \tau)$ for $i = 1, 2$ and where $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are the sample means before and after $[T\hat{\tau}]$, using $[T\hat{\alpha}]$ observations, that is, they are local sample means. Notice that, because $\hat{\tau}$ is consistent, both $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are consistent estimators of

μ_1 and μ_2 , respectively. Moreover, C6 implies that for $i = 1, 2$, μ_i is the mode of $f_i(\cdot)$. Consider

$$\tilde{\tau} = \arg \max_{\tau \in (\hat{\tau} - 2Ta, \hat{\tau} + 2Ta)} \tilde{\Phi}_0(\tau).$$

Before stating the asymptotic results of $\tilde{\tau}$, some comments about our choice of $\tilde{\Phi}_0(\tau)$, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are in place. By Theorems 3 and 4,

$$\hat{m}^+(\tilde{z}_1(\tau)) - \hat{m}^-(\tilde{z}_2(\tau)) \xrightarrow{P} \alpha 1(\tau = \tau_0) + (g(\mu_1) - g(\mu_2)),$$

and thus, $\tilde{\Phi}_0(\tau) \xrightarrow{P} \alpha^2 1(\tau = \tau_0) + \alpha(g(\mu_1) - g(\mu_2))$ which implies that

$$\lim_{n \rightarrow \infty} \Pr(\tilde{\Phi}_0(\tau_0) > \tilde{\Phi}_0(\tau)) = 1$$

for all τ . As for the choice of μ_1 and μ_2 is concerned, under Assumption C4', and if $\hat{m}^+(\cdot)$ and $\hat{m}^-(\cdot)$ are evaluated at two points, say x_1 and x_2 respectively, it is easily observed that the limit distribution in Theorem 4 (see Corollary 3 below) becomes, except constants,

$$\{d_1 I(U \geq 0) + d_2 I(U < 0)\} U \text{ where } U \sim N(0, \Sigma_0),$$

$$\text{with } \Sigma_0 = \sigma_\varepsilon^2 \xi_{(1)} \left(\frac{1}{f_1(x_2)} + \frac{1}{f_2(x_1)} \right) \text{ and } d_1 = \frac{f_1(x_2)}{f_2(x_2)} \text{ and } d_2 = \frac{f_2(x_1)}{f_1(x_1)}.$$

Thus, the choice of $\tilde{z}_i(\tau) = (\tilde{\mu}'_i, \tau)'$ seems natural because the asymptotic variance of the break point estimate becomes

$$2\sigma_\varepsilon^2 \xi_{(1)} \frac{d^2}{f_2(\mu_1)} = 2\sigma_\varepsilon^2 \xi_{(1)} \frac{d^2}{f_1(\mu_2)},$$

where $d = d_1 = d_2 = f_1(\mu_2)/f_2(\mu_2)$, which has the desirable effect because $f_1(\mu_1) = f_2(\mu_2)$ corresponds to a maximum. Then, we have the following corollary:

Corollary 3. Consider model (4). Under C1–C3, C4', C5, C6, B1–B4 and B6,

$$(Ta^{p-1})^{1/2}(\tilde{\tau} - \tau_0) \xrightarrow{d} \frac{d}{-\alpha_0 k_{(1)}^-(0)} U, \text{ with } U \sim N(0, \Sigma_0) \text{ and } \Sigma_0 = 2\sigma_\varepsilon^2 \xi_{(1)} \frac{1}{f_1(\mu_2)}.$$

Observe that in the second step, the search for the maximum is restricted to the subset $(\hat{\tau} - 2Ta, \hat{\tau} + 2Ta)$, since by Theorem 3, $\hat{\tau} - \tau_0 = o_p(a)$. Finally, we observe that the asymptotic distribution in the second step, that is, when the kernels k^+ and k^- are evaluated at the points μ_1 and μ_2 , respectively, is the same regardless of whether the regressors are stationary or not. In the former situation, we have, obviously, that $\mu_1 = \mu_2$. Finally, it is noteworthy to observe that

when the regressors are stationary we could have employed the objective function

$$\arg \max_{\zeta \in Q} \hat{\alpha} \hat{\Psi}_0(\zeta),$$

in the second step of the “two-step” algorithm, which is that employed in (5).

5. Monte Carlo experiments

In the first set of Monte Carlo experiments, we have considered the following models with only one regressor:

$$Y_t = \tau_t + (\beta_1 \tau_t - \beta_0)1(\tau_t \geq \zeta_0) + \varepsilon_t, \quad (6)$$

$$Y_t = X_t + (\beta_1 X_t - \beta_0)1(X_t \geq \zeta_0) + \varepsilon_t \quad (7)$$

with $\varepsilon_t \sim \text{iidN}(0, 0.1)$, $\tau_t = t/T$ and $X_t \sim U(0, 1)$, and where $(\beta_0, \beta_1)'$ takes the values $(0.5, 0)$, $(1, 0)$, $(0, 1)$ and $(1, 1)$. The motivation is to shed some light about the performance of our estimators in two particular situations. First, to examine their performance when the slope of the model is constant and the size of the jump increases, and second when the size of the jump is kept fixed, but the first derivative is not continuous. The former corresponds to $(\beta_0, \beta_1) = (0.5, 0)$ and $(\beta_0, \beta_1) = (1, 0)$ whereas the latter to $(\beta_0, \beta_1) = (0.5, 0)$, $(\beta_0, \beta_1) = (0, 1)$ and $(\beta_0, \beta_1) = (1, 1)$. In all the experiments we have employed the kernel $k^+(u) = u(3 - u) \exp(-u)1(u > 0)$ and $\zeta_0 = 1/2$.

First, we consider the trend model given in (6). The bandwidth parameter was $a = CT^{-1/2}$ where C takes the values 0.3 (0.1), which corresponds to the standard deviation of the Uniform random variable in $(0, 1)$. Notice that the range of bandwidths chosen is quite large, being the ratio between the first and last one equal to 2. We compare our estimator with the least squares estimator (LSE), $\hat{\zeta} = \tilde{k}/T$, where $\tilde{k} = \arg \min_k \{ \sum_{t=1}^k (Y_t - \bar{Y}_k)^2 + \sum_{t=k+1}^T (Y_t - \bar{Y}_k^*)^2 \}$, where \bar{Y}_k and \bar{Y}_k^* are the sample means using the first k and last $T - k$ observations respectively. When the LSE is consistent, it converges to ζ_0 faster than $\hat{\zeta}$. In fact, $\hat{\zeta} - \zeta_0 = O_p(T^{-1})$, which is a typical rate of convergence of change point estimators in parametric models, see Chan (1993) and Bai (1994). However, the LSE can be inconsistent when $\beta^* = E[Y_t 1(\tau_t \geq \zeta_0)] - E[Y_t 1(\tau_t < \zeta_0)] = 0$. In model (6), the performance of the nonparametric estimator depends on $\alpha_0 = 0.5 \beta_1 - \beta_0$, while the performance of the LSE depends on β^* . Table 1 shows that the nonparametric estimator is quite insensitive to the bandwidth choice. It also shows that when $(\beta_0, \beta_1) = (0.5, 0)$ or $(1, 1)$, the LSE is inconsistent, that is, the mean squared error (MSE) does not converge to zero with the sample size and is much bigger than the MSE of the nonparametric estimator. In particular, we observe that when $(\beta_0, \beta_1) = (1, 1)$, the bias of the

Table 1

Biases and MSEs of nonparametric and OLS estimators of structural break points in model

$$Y_t = \tau_t + (\beta_1 \tau_t - \beta_0) 1(\tau_t \geq \zeta_0) + \varepsilon_t, t = 1, \dots, T$$

with $T = 50, 100$ and bandwidth parameter $a = CT^{-1/2}$ based on 5000 replications

	$\beta_0 = 0.5, \beta_1 = 0$				$\beta_0 = 1, \beta_1 = 0$			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>
$C = 0.3$	-0.435	3.100	-0.414	2.405	0.134	0.385	0.062	0.116
$C = 0.4$	-0.610	3.100	-0.250	1.820	0.021	0.253	0.000	0.036
$C = 0.5$	-0.824	3.610	-0.300	1.620	-0.010	0.201	-0.021	0.018
$C = 0.6$	-1.179	4.603	-0.434	1.684	-0.010	0.173	-0.018	0.013
<i>OLS</i>	-0.889	9.182	-0.772	9.835	-1.002	0.017	-0.488	0.002
	$\beta_0 = 0, \beta_1 = 1$				$\beta_0 = 1, \beta_1 = 1$			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>
$C = 0.3$	-0.031	2.731	-0.132	2.396	1.157	3.305	0.000	2.470
$C = 0.4$	0.606	1.948	0.000	1.689	3.745	3.782	0.978	2.058
$C = 0.5$	1.608	1.133	0.314	1.068	7.342	4.806	1.925	2.052
$C = 0.6$	2.095	0.600	0.920	0.639	12.880	6.598	3.610	2.465
<i>OLS</i>	0.837	0.160	0.810	0.074	20.670	7.009	23.83	6.824

Note: All the values in this table must be divided by 100.

LSE converges to 0.25. This appears sensible since β^* increases as τ_t increases to 0.75 and latter decreases. For the other two cases, as the sample size increases, the LSE always performs better as expected, since the MSE of the nonparametric estimator decreases slowly than that of the LSE with the sample size.

The same Monte Carlo experiment was repeated for model (7), with the same kernel and bandwidth choice $a = CT^{-1/2}$, $C = 0.3(0.1)$. The results are provided in Table 2. Qualitatively, the performance of the estimator is similar to the situation of fixed regressor, and the same comments given for model (6) apply here for model (7).

In the second set of Monte Carlo experiments, we have employed the following autoregressive model with a jump in the mean:

$$Y_t = 0.5Y_{t-1} + \beta_0 1(\tau_t \geq \zeta_0) + \varepsilon_t, \quad (8)$$

for two values of β_0 , namely 0.25 and 0.50, which correspond to a jump of 0.5 and 1 respectively, and where ε_t is as in models (6) or (7).

Table 2

Biases and MSEs of nonparametric estimators of structural break points with $a = CT^{-1/2}$ and OLS estimator based on 5000 replications of the threshold model

$$Y_t = X_t + (\beta_1 X_t - \beta_0) 1(X_t \geq \zeta_0) + \varepsilon_t, t = 1, \dots, T$$

and $T = 50, 100$

	$\beta_0 = 0.5, \beta_1 = 0$				$\beta_0 = 1, \beta_1 = 0$			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>
$C = 0.3$	-1.336	4.225	-0.672	3.983	-1.398	3.233	-0.633	2.825
$C = 0.4$	-1.403	3.985	-0.281	3.013	-1.266	2.535	-0.376	1.434
$C = 0.5$	-1.455	4.147	-0.500	2.240	-1.000	2.070	-0.500	0.679
$C = 0.6$	-1.934	4.857	-0.500	1.989	-1.152	1.722	-0.450	0.434
<i>OLS</i>	-1.988	8.968	-1.481	9.639	-1.179	0.125	-0.140	0.020
	$\beta_0 = 0, \beta_1 = 1$				$\beta_0 = 1, \beta_1 = 1$			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>
$C = 0.3$	-0.200	4.679	0.371	4.666	-0.527	4.334	0.395	4.087
$C = 0.4$	-0.100	5.296	0.000	4.788	0.145	4.051	0.271	3.075
$C = 0.5$	2.454	6.032	1.715	5.202	2.780	4.418	1.210	2.366
$C = 0.6$	6.980	5.957	4.906	5.728	7.857	5.627	2.614	2.421
<i>OLS</i>	7.540	1.768	0.475	1.536	19.050	6.364	22.989	6.462

Note: All the values in this table must be divided by 100.

We have employed the kernels $k^+(u)$ and Gaussian for the “regressor” time and Y_{t-1} respectively, while the bandwidth choices where $a_1 = CT^{-1/2}$ and $a_2 = CT^{-1/5}$ with $C = 0.3$ (0.1). The reason to use two different rates of convergence for the bandwidth comes from the well known nonparametric result that the smoother the function, the optimal bandwidth becomes bigger. The results are given in Table 3.

The results are very encouraging, and the estimator, even in this rather complicated model, appears to work very well. Moreover, we clearly observe the gain obtained using the two step algorithm indicated in Section 4. In the experiment, we have used, as in the initial step, the sample mean of the series Y_t , while in the second step, we employed, as estimators of μ_1 and μ_2 , the sample means, using only $0.5T^{1/2}$ observations, that is local sample means, to provide an even more local character to our estimator. Notice that $0.5T^{1/2}$ corresponds roughly to the value Ta_1 .

Table 3

Biases and MSEs of nonparametric estimators of structural break points based on 5000 replications of model

$$Y_t = 0.5Y_{t-1} + \beta_0 1(\tau_t \geq \zeta_0) + \varepsilon_t, \quad t = 1, \dots, T$$

with $a_1 = CT^{-1/2}$ and $a_2 = CT^{-1/5}$, $T = 50, 100$

Preliminary estimator								
$\beta_0 = 0.25$				$\beta_0 = 0.5$				
$T = 50$		$T = 100$		$T = 50$		$T = 100$		
<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	
$C = 0.3$	- 3.353	4.763	- 2.046	4.842	- 2.711	4.609	- 0.809	4.708
$C = 0.4$	- 3.463	5.139	- 2.224	4.784	- 5.920	5.196	- 3.356	4.660
$C = 0.5$	- 0.010	5.431	- 1.411	4.708	- 5.128	5.900	- 3.704	5.051
$C = 0.6$	4.283	5.482	4.285	4.445	0.388	5.624	- 0.316	4.919
One-step estimator								
$\beta_0 = 0.25$				$\beta_0 = 0.5$				
$T = 50$		$T = 100$		$T = 50$		$T = 100$		
<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	<i>BIAS</i>	<i>MSE</i>	
$C = 0.3$	0.888	2.355	0.135	2.377	3.364	1.932	2.356	1.853
$C = 0.4$	2.194	2.469	0.998	2.340	2.658	1.713	2.052	1.601
$C = 0.5$	5.523	2.897	4.085	2.441	3.164	1.798	1.350	1.636
$C = 0.6$	7.816	3.181	7.016	2.641	4.008	1.917	2.054	1.671

Note: All the values in this table must be divided by 100.

Acknowledgements

Research funded by “Dirección General de Enseñanza Superior” (DGES) of Spain, reference number: PB95-0292. We are grateful to Inmaculada Fiteni for carrying out the numerical work in Section 5.

Appendix A

Let us introduce some notation. From now on, for any function $q: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$, $q_{(m)}(z) = \partial^m q(z) / \partial z_r^m$ and $\int_a^b q(z) dz = \int_a^b [\int_{\mathbb{R}^p} q(z) \prod_{j \neq r}^{p+1} dz_j] dz_r$ where $z = (z_1, z_2, \dots, z_{p+1})'$. Also, $h = (Ta^{p+1})^{1/2}$, $\hat{\delta}_j(v) \equiv \hat{\Psi}_0(\zeta_0^{(j)} + vah^{-1})$ and $f(z)$ as the pdf of $Z_t = (X'_t, Z'_{t(p+1)})'$.

For the sake of presentation we shall prove first Theorem 2, assuming that Theorem 1 holds.

Proof of Theorem 2. By Theorem 1, $\hat{\zeta}^{(j)}$ is a point in Q of the form $\zeta_0^{(j)} + vah^{-1}$, $v \in \mathbb{R}$. By construction, $\hat{v}_j = \arg \max_{v \in \mathbb{R}} \hat{\gamma}_j(v) = (Ta^{p-1})^{1/2} (\hat{\zeta}^{(j)} - \zeta_0^{(j)})$ where $\hat{\gamma}_j(v) = \hat{\delta}_j(v)^2 - \hat{\delta}_j(0)^2$. Define $\hat{\eta}_j(v) = \hat{\delta}_j(v) - \hat{\delta}_j(0)$. Then $\hat{\gamma}_j(v) = \hat{\eta}_j(v)^2 + 2\hat{\eta}_j(v)(\hat{\delta}_j(0) - \alpha_0^{(j)}) + 2\alpha_0^{(j)}\hat{\eta}_j(v)$. Assume, to be shown later, that

$$h^2[\hat{\eta}_1(v_1), \hat{\eta}_2(v_2), \dots, \hat{\eta}_N(v_N)] \stackrel{\text{weakly}}{\Rightarrow} W(v_1, v_2, \dots, v_N) \text{ on } \mathcal{C}(-\infty, \infty)^N \quad (\text{A.1})$$

where $W(v_1, v_2, \dots, v_N) = [\eta_1(v_1), \eta_2(v_2), \dots, \eta_N(v_N)]'$ with $\eta_j(v) = (\alpha_0^{(j)}v^2k_{(1)}^-(0))/2 + U_jv$, $U_j \sim N(0, [2s(z_0^{(j)}) + \rho_0^{(j)}] \xi_{(1)}/f(z_0^{(j)})$ and $z_0^{(j)}$ defined after Eq. (2). Then,

$h^2\hat{\gamma}_j(v) \Rightarrow 2\alpha_0^{(j)}\eta_j(v)$ on $\mathcal{C}(-\infty, \infty)$. Denote v_j^* as the solution to $\partial\eta_j(v)/\partial v = 0$ which is $v_j^* = -U_j/\alpha_0^{(j)}k_{(1)}^-(0)$ and corresponds to a maximum since $\alpha_0^{(j)}\partial^2\eta_j(v)/\partial v^2 < 0$. Therefore, by the continuous mapping theorem, $\hat{v}_j \xrightarrow{d} v_j^*$ if (A.1) holds, which will be shown next. Put $\hat{\eta}_j(v) = R_{1j}^-(v) - R_{1j}^+(v) + R_{2j}(v) + R_{3j}(v) + R_{4j}^-(v) - R_{4j}^+(v) + r_j(v)$, where

$$R_{1j}^\mp(v) = P^\mp(z_0^{(j)})\mathbb{F}_j^\mp(v) \left[\frac{1}{\hat{f}^\mp(z_0^{(j)})\hat{f}^\mp(\omega_j(v))} - \frac{1}{f(z_0^{(j)})f(\omega_j(v))} \right],$$

$$R_{2j}(v) = \mathbb{P}_j^-(v) \left[\frac{1}{\hat{f}^+(z_0^{(j)})} - \frac{1}{\hat{f}^-(z_0^{(j)})} \right],$$

$$R_{3j}(v) = [\mathbb{P}_j^+(v) - \mathbb{P}_j^-(v)] \left[\frac{1}{\hat{f}^+(z_0^{(j)})} - \frac{1}{f(z_0^{(j)})} \right],$$

$$R_{4j}^\mp(v) = \frac{[P^\mp(\omega_j(v)) - P^\mp(z_0^{(j)})]\mathbb{F}_j^\mp(v)}{\hat{f}^\mp(z_0^{(j)})\hat{f}^\mp(\omega_j(v))},$$

$$r_j(v) = \frac{[\mathbb{P}_j^+(v) - \mathbb{P}_j^-(v)]}{f(z_0^{(j)})} - m^+(z_0^{(j)})\frac{\mathbb{F}_j^+(v)}{f(\omega_j(v))} + m^-(z_0^{(j)})\frac{\mathbb{F}_j^-(v)}{f(\omega_j(v))},$$

$$\omega_j(v) = z_0(\zeta_0^{(j)} + vah^{-1}),$$

$$\mathbb{P}_j^\pm(v) = \hat{P}^\pm(\omega_j(v)) - \hat{P}^\pm(z_0^{(j)}),$$

$$\mathbb{F}_j^\pm(v) = \hat{f}^\pm(\omega_j(v)) - \hat{f}^\pm(z_0^{(j)}).$$

By Propositions B.4 and B.5 of Appendix B Theorem 5.1 of Robinson (1983), $R_{1j}^\mp(v)$, $R_{2j}(v)$, $R_{3j}(v)$ and $R_{4j}^\mp(v)$ are $o_p(h^{-2})$ for $j = 1, \dots, N$. Thus, (A.1) follows if $h^2[r_1(v_1), r_2(v_2), \dots, r_N(v_N)] \stackrel{\text{weakly}}{\Rightarrow} W(v_1, v_2, \dots, v_N)$ on the space $\mathcal{C}(-\infty, \infty)^N$.

By Propositions B.1 to B.3, $E[h^2 r_j(v)] = \alpha_0^{(j)} k_{(1)}^-(0) v^2 / 2 + o(v^2)$, while by Propositions B.4–B.7 and the asymptotic uncorrelation at two different points, $\text{Cov}[h^2 r_j(v_{j_1}), h^2 r_i(v_{i_2})] = [2s(z_0^{(j)}) + \rho_0^{(j)}] \xi_{(1)} v_{j_1} v_{i_2} 1(j = i) / f(z_0^{(j)}) + o(v_{j_1} v_{i_2})$. Define $\ddot{r}_j(v) = r_j(v) - E[r_j(v)]$. Then, the weak convergence of $h^2 [r_1(v_1), r_2(v_2), \dots, r_N(v_N)]'$ follows by that of $h^2 \sum_{j=1}^N \lambda_j \ddot{r}_j(v_j)$, for any $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)'$ where $\lambda' \lambda = 1$. By Theorem 8.1 of Billingsley (1968), we need to prove the convergence of the finite dimensional distributions and tightness. For fixed $v_1, v_2, \dots, v_\ell \in [-M, M]^N$, by Lemma 7.1. of Robinson (1983), the asymptotic independence of the kernel regression estimator at two different points and Cr amer–Wold device,

$$h^2 \left\{ \sum_{j=1}^N \lambda_j \ddot{r}_j(v_{j_1}), \sum_{j=1}^N \lambda_j \ddot{r}_j(v_{j_2}), \dots, \sum_{j=1}^N \lambda_j \ddot{r}_j(v_{j_\ell}) \right\} \xrightarrow{d} N_\ell(0, \Gamma)$$

where $\Gamma = [\sigma_{ab}]_{a,b=1, \dots, \ell}$, with $\sigma_{ab} = \sum_{j=1}^N \lambda_j^2 (2s(z_0^{(j)}) + \rho_0^{(j)}) v_{ja} v_{jb} \xi_{(1)} / f(z_0^{(j)})$. Next, tightness. By Bickel and Wichura (1971), it suffices to prove that $\forall \varphi_1, \varphi_2 > 0$, there exists $\beta > 0$ such that

$$Pr \left\{ \sup_{\{v_{j_1}, v_{j_2} \in Q^{(j)}: |v_{j_1} - v_{j_2}| < \beta, j = 1, \dots, N\}} \left| h^2 \sum_{j=1}^N \lambda_j [\ddot{r}_j(v_{j_1}) - \ddot{r}_j(v_{j_2})] \right| > \varphi_1 \right\} < \varphi_2,$$

which, by the triangle inequality, will be sufficient if $\forall j = 1, \dots, N$

$$Pr \left\{ \sup_{\{v_{j_1}, v_{j_2} \in Q^{(j)}: |v_{j_1} - v_{j_2}| < \beta, j = 1, \dots, N\}} \left| h^2 \ddot{r}_j(v_{j_1}) - \ddot{r}_j(v_{j_2}) \right| > \frac{\varphi_1}{|\lambda_j| N} \right\} < \frac{\varphi_2}{N},$$

which holds by Theorem 12.3 of Billingsley (1968), after observing that by Proposition B.8, $\lim_{T \rightarrow \infty} h^4 E(\ddot{r}_j(v_{j_1}) - \ddot{r}_j(v_{j_2}))^2 \leq C |v_{j_1} - v_{j_2}|^2$, $j = 1, \dots, N$. Thus, $h^2 (\hat{\eta}_1(v_1), \hat{\eta}_2(v_2), \dots, \hat{\eta}_N(v_N))' \xRightarrow{\text{weakly}} W(v_1, v_2, \dots, v_N)$ on the space $\mathcal{C}(-M, M)^N$. Then, using Whitt's (1970) metric, we can extend this convergence to the space $\mathcal{C}(-\infty, \infty)^N$, and thus, we conclude the proof of part (a).

To show (b), note that $(a^{-1} h^2)^{1/2} (\hat{\Psi}_0(\zeta^{(j)}) - \hat{\Psi}_0(\zeta_0^{(j)})) = O_p(1)$, by Theorem 2(a) and the continuous mapping theorem. But, by Robinson (1983), $h(\hat{\Psi}_0(\zeta_0^{(j)}) - \alpha_0^{(j)}) \xrightarrow{d} N(0, W_{j0})$. Thus, by usual arguments, we conclude. \square

Proof of Theorem 1. First, we show that $|\hat{\zeta}^{(j)} - \zeta_0^{(j)}| = O_p(a)$ for all $j = 1, \dots, N$ using the same arguments of Chu and Wu (1993). Define $\hat{Q}_j = [\zeta_0^{(j)} - a, \zeta_0^{(j)} + a]$ for $j = 1, \dots, N$ and $\Phi(\zeta) = E(\hat{P}^-(z_0(\zeta)) - \hat{P}^+(z_0(\zeta))) / f(z_0(\zeta))$. By Theorem 3.3.5 of Gy orfi et al. (1989), $\hat{\Psi}_0(\zeta) = \Phi(\zeta) + O_p(h^{-1} \log a^{-1})$, while by continuity of $m(z)$ in $Q^* = Q - \cup_{j=1}^N \hat{Q}_j$, $\sup_{\zeta \in Q^*} |\Phi(\zeta)| = O(a^2)$. So, the supremum belongs to the set

$\cup_{j=1}^N \tilde{Q}_j$ since $\Phi(\zeta_0^{(j)}) \rightarrow \alpha_0^{(j)}$ for all j . Next, by Propositions B.1 and B.2, with $\Phi(\zeta) = \Phi_j(\zeta)$, for a point $\zeta = \zeta_0^{(j)} + av$,

$$\Phi_j(\zeta) = \alpha_0^{(j)} - \frac{\alpha_0^{(j)} k_{(1)}^-(0) v^2}{2} + \mathcal{O}(a)$$

and thus $\hat{\Psi}_0(\zeta) = \alpha_0^{(j)} - \alpha_0^{(j)} k_{(1)}^-(0) v^2 / 2 + \Xi_j(\zeta)$ where $\sup_j \sup_{\zeta} |\Xi_j(\zeta)| = \mathcal{O}_p(a + h^{-1} \log a^{-1}) = \mathcal{O}_p(a)$ and the subscript j indicates that ζ belongs to \tilde{Q}_j . First, we prove that $|\hat{\zeta}^{(1)} - \zeta_0^{(1)}| = \mathcal{O}_p(a)$. $\Pr \{|\hat{\zeta}^{(1)} - \zeta_0^{(1)}| > a\}$ is

$$\begin{aligned} & \Pr \left\{ \sup_{\zeta \in \tilde{Q}_1} \left| \alpha_0^{(1)} - \frac{\alpha_0^{(1)} k_{(1)}^-(0) v^2}{2} + \Xi_1(\zeta) \right| \right. \\ & \leq \sup_{j=2, \dots, N} \sup_{\zeta \in \tilde{Q}_j} \left| \alpha_0^{(j)} - \frac{\alpha_0^{(j)} k_{(1)}^-(0) v^2}{2} + \Xi_j(\zeta) \right| \left. \right\} \\ & \leq \Pr \left\{ \left| \alpha_0^{(1)} - \frac{\alpha_0^{(1)} k_{(1)}^-(0) v^2}{2} \right| - \left| \alpha_0^{(2)} - \frac{\alpha_0^{(2)} k_{(1)}^-(0) v^2}{2} \right| \right. \\ & \left. \leq \sup_{\zeta \in \tilde{Q}_1} |\Xi_1(\zeta)| + \sup_{j=2, \dots, N} \sup_{\zeta \in \tilde{Q}_j} |\Xi_j(\zeta)| \right\} \end{aligned}$$

which converges to zero since $\sup_j \sup_{\zeta} |\Xi_j(\zeta)| = \mathcal{O}_p(a)$ and $|\alpha_0^{(1)}| > |\alpha_0^{(2)}|$. Thus, $\hat{\zeta}^{(1)}$ is a point in Q of the form $\hat{\zeta}^{(1)} = \zeta_0^{(1)} + a\hat{v}_1$ where $\hat{v}_1 = \arg \max_{|v| \leq 1} \hat{\gamma}_1(v)$ with $\hat{\gamma}_1(v) = \hat{\delta}_1(v)^2 - \hat{\delta}_1(0)^2$ and $\hat{\delta}_1(v) = \hat{\Psi}_0(\zeta_0^{(1)} + av)$ and where, without loss of generality, $|v| \leq 1$. We now give the proof for $\hat{\zeta}^{(2)}$. The proof for the remaining $\hat{\zeta}^{(j)}$ follows similarly. Repeating the same arguments, we have that

$$\Pr \left\{ |\alpha_0^{(2)} + \Xi_1(\zeta_0^{(2)})| \leq \sup_{j=3, \dots, N} \sup_{\zeta \in \tilde{Q}_j} \left| \alpha_0^{(j)} - \frac{\alpha_0^{(j)} k_{(1)}^-(0) v^2}{2} + \Xi_j(\zeta) \right| \right\} \rightarrow 0.$$

Next, because for any arbitrary small $\varepsilon > 0$, $|\zeta_0^{(j)} - \zeta_0^{(\ell)}| > \varepsilon$ for $j \neq \ell$ then $|\zeta_0^{(1)} - \zeta_0^{(2)}| > 3a$, and $\hat{\zeta}^{(1)} \in \tilde{Q}_1$ and $\hat{\zeta}^{(2)} \in \cup_{j=2}^N \tilde{Q}_j$,

$$|\hat{\zeta}^{(2)} - \zeta_0^{(1)}| \geq |\hat{\zeta}^{(1)} - \hat{\zeta}^{(2)}| - |\hat{\zeta}^{(1)} - \zeta_0^{(1)}| \geq 2a - |\hat{\zeta}^{(1)} - \zeta_0^{(1)}|$$

which implies that

$$\Pr \{|\hat{\zeta}^{(2)} - \zeta_0^{(1)}| < a\} \rightarrow 0.$$

Combining these results

$$\Pr |\hat{\zeta}^{(2)} - \zeta_0^{(2)}| > a \rightarrow 0.$$

Thus, it only remains to show that $\hat{v}_j = \mathcal{O}_p(h^{-1})$ which will imply that $(\hat{\zeta}^{(j)} - \zeta_0^{(j)}) = a\hat{v}_j = \mathcal{O}_p(ah^{-1})$. We only give the proof for $\hat{v}_1 = \mathcal{O}_p(h^{-1})$. The proof for the remaining $\hat{v}_j, j = 2, \dots, N$, follows similarly.

Dropping the subscript 1, by standard kernel manipulations and Propositions B.1–B.8, we have that, in $|v| \leq 1$, $\hat{\eta}(v) = r(v) + \mathcal{O}_p(h^{-1})$, and thus

$$\hat{\gamma}(v) = r^2(v) + 2r(v)(\alpha_0^{(1)} + h^{-1}) + 2\alpha_0^{(1)}\mathcal{O}_p(h^{-1}) + \mathcal{O}_p(h^{-2} + ah^{-1}).$$

Denoting $\check{\gamma}(v) = r^2(v) + 2r(v)(\alpha_0^{(1)} + h^{-1})$, we have that

$$|\arg \max_{|v| \leq 1} \hat{\gamma}(v) - \arg \max_{|v| \leq 1} \check{\gamma}(v)| = \mathcal{O}_p(h^{-1}),$$

and hence, it suffices to prove that $\arg \max_{|v| \leq 1} \check{\gamma}(v) = \mathcal{O}_p(h^{-1})$ to conclude.

By Propositions B.1–B.3, and because $a^2 \sim h^{-1}$ by B5,

$$\bar{\eta}(v) = \mathbb{E}(r(v)) = (C_1 + h^{-1})v^2$$

where $C_1 = \alpha_0^{(1)}k_{(1)}^-(0)/2$. Define $\bar{\gamma}(v) = \bar{\eta}(v)^2 + 2(\alpha_0^{(1)} + h^{-1})\bar{\eta}(v)$. The *first order conditions*, $\bar{\gamma}_{(1)}(v) = 2[\bar{\eta}(v) + \alpha_0^{(1)} + h^{-1}]\bar{\eta}_{(1)}(v) = 0$, have two possible solutions v_1 and v_2 in $[-1, 1]$. The first one, that is v_1 , satisfies that $\bar{\eta}(v_1) + (\alpha_0^{(1)} + h^{-1}) = 0$ and corresponds to a minimum since $\bar{\gamma}_{(2)}(v_1) = 2\bar{\eta}_{(1)}(v_1)^2 > 0$, while the second, that is $v_2 = 0$, satisfies that $\bar{\eta}_{(1)}(v_2) = 0$ and corresponds to a maximum since $\bar{\gamma}_{(2)}(v_2) = 2\bar{\eta}_{(1)}(v_2)^2 + 2C_1(\bar{\eta}(v_2) + \alpha_0^{(1)} + h^{-1}) = 2C_1(\alpha_0^{(1)} + h^{-1}) < 0$. Because of the continuity of $\check{\gamma}(v)$, $\check{\gamma}_{(1)}(\dot{v}) = 0$, then $\bar{\gamma}_{(1)}(0) - \check{\gamma}_{(1)}(0) = \check{\gamma}_{(1)}(\dot{v}) - \check{\gamma}_{(1)}(0) = \dot{v}\check{\gamma}_{(2)}(\psi_1\dot{v})$, $\psi_1 \in (0, 1)$ by the mean value theorem (MVT). The theorem is proved since $\check{\gamma}_{(2)}(\psi_1\dot{v}) = \mathcal{O}_p(1)$, and $(\check{\gamma}_{(1)}(0) - \bar{\gamma}_{(1)}(0)) = \mathcal{O}_p(h^{-1})$ by standard kernel manipulations. Then $\dot{v} = \mathcal{O}_p(h^{-1})$, and thus $\hat{v}_1 = \mathcal{O}_p(h^{-1})$. \square

Proof of Corollary 1. (a) Because the additive form of the objective function, that is $\sum_{k=1}^M \hat{\Psi}_k(\zeta)^2$ and we have only changed the kernel function $k(u)$ by $\tilde{k}(u)$ for those coordinates responsible of the jump, then by similar arguments to those employed in Theorem 1, $\hat{\zeta}_k$ is a point in $\mathcal{Q}_{k(M)}$ of the form $\zeta_{0k} + v_k a/h$, where $v_k \in \mathbb{R}$ and $k = 1, \dots, M$. Define $\hat{\delta}_k(v_k) = \hat{m}^+(\omega_k(v_k)) - \hat{m}^-(\omega_k(v_k))$, $k = 1, \dots, M$ where $\omega_k(v_k) = z_0(\zeta_{0k} + v_k a/h)$. By construction, $(\hat{v}_1, \dots, \hat{v}_M)' = \hat{v} = \arg \max_{v \in \mathbb{R}^M} \sum_{k=1}^M \hat{\gamma}_k(v_k)$ where $\hat{\gamma}_k(v_k) = \hat{\delta}_k(v_k)^2 - \hat{\delta}_k(0)^2$. Define $\hat{\eta}_k(v_k) = \hat{\delta}_k(v_k) - \hat{\delta}_k(0)$. Then $\hat{\gamma}_k(v_k) = \hat{\eta}_k(v_k)^2 + 2\hat{\eta}_k(v_k)(\hat{\delta}_k(0) - \alpha_{0k}) + 2\alpha_{0k}\hat{\eta}_k(v_k)$. Thus, as in Theorem 2, we need to show that

$$h^2 \sum_{k=1}^M \lambda_k \hat{\eta}_k(v_k) \stackrel{\text{weakly}}{\Rightarrow} \sum_{k=1}^M \lambda_k \eta_k(v_k) \text{ on } \mathcal{C}^M(-\infty, \infty), \text{ where } \sum_{k=1}^M \lambda_k^2 = 1,$$

$$\eta_k(v_k) = \frac{\alpha_{0k} v_k^2 k_{(1)}^-(0)}{2} + \left\{ \frac{[2s(z_{0k}) + \rho_{0k}]\xi_{(1)}}{f(z_{0k})} \right\}^{1/2} v_k U_k,$$

$z_{0k} = (x'_{0k}, z_{0(p+1)})'$ and $U_k \simeq \text{NID}(0, 1)$, $k = 1, \dots, M$. By similar, if not easier, arguments to those employed in Theorem 2, $\sum_{k=1}^M \lambda_k \hat{\eta}_k(v_k) = \sum_{k=1}^M \lambda_k r_k(v_k) +$

$\text{O}_p(h^{-2})$, where

$$r_k(v_k) = \frac{1}{f(z_{0k})} \{ [\mathbb{P}_k^+(v_k) - \mathbb{P}_k^-(v_k)] - m^+(z_{0k})\mathbb{F}_k^+(v_k) + m^-(z_{0k})\mathbb{F}_k^-(v_k) \},$$

$$\text{and } \mathbb{P}_k^\pm(v_k) = \hat{P}_k^\pm(\omega_k(v_k)) - \hat{P}_k^\pm(z_{0k}), \mathbb{F}_k^\pm(v_k) = \hat{f}_k^\pm(\omega_k(v_k)) - \hat{f}_k^\pm(z_{0k}).$$

The remainder of the proof is identical to that of Theorem 2, and thus is omitted.

(b) The proof is identical to that of Theorem 2(b), and thus is omitted. \square

Proof of Corollary 2. By Prakasa Rao's (1981) Theorem 4.5.6., $\hat{x}^* - x^* = \text{O}_p(g^{-1})$, where $g = (Ta^{p+2})^{1/2}$. Because $\partial f(x^*)/\partial x = 0$, $f(x^*) - f(\hat{x}^*) = \text{O}_p(g^{-2})$ by definition of x^* and $f(\cdot)$ possesses finite second derivatives. Applying the Lemma below, the difference of the objective functions when evaluated at $z^*(\zeta) = (x^{*'}, \zeta)'$ or $\hat{z}^* = (\hat{x}^{*'}, \zeta)'$ is $\text{O}_p(h^{-1})$. From here, the Corollary follows proceeding as in the proof of Theorem 2.

Lemma. For any function $q(z)$ twice continuously differentiable in its first p coordinates, if the kernel $k(\cdot)$ is symmetric, then

$$\begin{aligned} & \sup_{\zeta \in \mathcal{Q}} \left| \frac{1}{Ta^{p+1}} \sum_{t=1}^T q(Z_t) \left\{ K^\pm \left(\frac{Z_t - \hat{z}^*(\zeta)}{a} \right) - K^\pm \left(\frac{Z_t - z^*(\zeta)}{a} \right) \right\} \right| \\ &= \sup_{\zeta \in \mathcal{Q}} \left| g^{-1} \left(\sum_{j=1}^p \frac{\partial}{\partial x_j} [E[q(Z_t|Z_t = z^*(\zeta))] f(x^*)] \right) \right| + \\ & \quad \text{O}_p \left(\frac{\log a^{-1}}{g(Ta^{p+3})^{1/2}} + a^2 \right). \end{aligned}$$

The proof is standard after using the MVT of $K^\pm(\cdot)$.

Proof of Theorem 4. By Theorem 3, $\hat{\tau}$ is a point in \mathcal{Q} of the form $\tau_0 + vah^{-1}$, $v \in \mathbb{R}$. By construction, $\hat{v} = \arg \max_{v \in \mathbb{R}} \hat{\gamma}(v) = (Ta^{p-1})^{1/2}(\hat{\tau} - \tau_0)$ where $\hat{\gamma}(v) = \hat{\delta}(v)^2 - \hat{\delta}(0)^2$. Define $\hat{\eta}(v) = \hat{\delta}(v) - \hat{\delta}(0)$. Then $\hat{\gamma}(v) = \hat{\eta}(v)^2 + 2\hat{\eta}(v)(\hat{\delta}(0) - \alpha_0) + 2\alpha_0\hat{\eta}(v)$. Assume, to be shown later, that

$$h^2 \hat{\eta}(v) \stackrel{\text{weakly}}{\Rightarrow} \eta(v) \text{ on } \mathcal{C}(-\infty, \infty) \quad (\text{A.2})$$

where $\eta(v) = \alpha_0 k_{(1)}^-(0)v^2 \{d^{-1}1(v \geq 0) + d1(v < 0)\}/2 + Uv$, and $U \sim N(0, \xi_{(1)}(s(z_0))/f_1(x_0) + (\rho_0 + s(z_0))/f_2(x_0))$, $z_0 = (x'_0, \tau_0)'$. Then, $h^2 \hat{\gamma}(v) \stackrel{\text{weakly}}{\Rightarrow} 2\alpha_0\eta(v)$ on the space $\mathcal{C}(-\infty, \infty)$. Since $\alpha_0 \partial^2 \eta(v)/\partial v^2 < 0$, then

$$v^* = \frac{U}{-\alpha_0 k_{(1)}^-(0)} (dI(U \geq 0) + d^{-1}I(U < 0))$$

corresponds to a maximum where v^* is the solution to $\alpha_0 \partial \eta(v) / \partial v = 0$. Then, $\hat{v} \xrightarrow{d} v^*$ if (A.2) holds. Observe that this agrees with the results of Theorem 2, where $f_1(x) = f_2(x)$, that is $d = 1$. Thus, it only remains to examine (A.2). Put $\hat{\eta}(v) = \sum_{i=1}^6 R_i(v) + r(v)$, where

$$R_1(v) = -\mathbb{P}^+(v) \left(\frac{\hat{f}^+(\omega(v)) - f^+(\omega(v))}{f^+(\omega(v)) f^+(\omega(v))} \right), R_2(v) = -\frac{(\hat{P}^+(z_0) - P^+(z_0)) \mathbb{F}^+(v)}{f^+(\omega(v)) f_2(z_0)}$$

$$R_3(v) = -P^+(z_0) \mathbb{F}^+(v) \left[\frac{1}{\hat{f}^+(\omega(v)) \hat{f}^+(z_0)} - \frac{1}{f^+(\omega(v)) f_2(z_0)} \right],$$

$$R_4(v) = \mathbb{P}^-(v) \left(\frac{\hat{f}^-(\omega(v)) - f^-(\omega(v))}{\hat{f}^-(\omega(v)) f^-(\omega(v))} \right), R_5(v) = \frac{(\hat{P}^-(z_0) - P^-(z_0)) \mathbb{F}^-(v)}{f^-(\omega(v)) f_1(z_0)}$$

$$R_6(v) = \hat{P}^-(z_0) \mathbb{F}^-(v) \left[\frac{1}{\hat{f}^-(\omega(v)) \hat{f}^-(z_0)} - \frac{1}{f^-(\omega(v)) f_1(z_0)} \right]$$

$$r(v) = \frac{\mathbb{P}^+(v)}{f^+(\omega(v))} - \frac{\mathbb{P}^-(v)}{f^-(\omega(v))} - (g(x_0) + \alpha_0) \frac{\mathbb{F}^+(v)}{f^+(\omega(v))} + g(x_0) \frac{\mathbb{F}^-(v)}{f^-(\omega(v))},$$

where $\omega(v) = z_0(\tau_0 + v a h^{-1})$, $\mathbb{P}^\pm(v) = \hat{P}^\pm(\omega(v)) - \hat{P}^\pm(z_0)$, $\mathbb{F}^\pm(v) = \hat{f}^\pm(\omega(v)) - \hat{f}^\pm(z_0)$, $f^-(\omega(v)) = f_1(x_0) + f_0(x_0) k_{(1)}^-(0) a v^2 \mathbf{1}(v > 0) / 2h$ and $f^+(\omega(v)) = f_2(x_0) - f_0(x_0) k_{(1)}^+(0) a v^2 \mathbf{1}(v < 0) / 2h$, with $f_0(x) = f_2(x) - f_1(x)$.

By Propositions B.4 and B.5 and a routine extension of Theorem 5.1 of Robinson (1983), $R_j(v) = o_p(h^{-2})$, $j = 1, \dots, 6$. Thus, (A.2) follows if $h^2 r(v) \Rightarrow \eta(v)$ on $\mathcal{C}(-\infty, \infty)$. But, this is the case by similar, if not easier, arguments to those employed in Theorem 2, after observing that by Propositions B.9–B.12.

$$E[h^2 r(v)] = \frac{\alpha_0 k_{(1)}^-(0) v^2 \{d^{-1} \mathbf{1}(v \geq 0) + d \mathbf{1}(v < 0)\}}{2} + O(av^2),$$

$$\text{Cov}[h^2 r(v_1), h^2 r(v_2)] = v_1 v_2 \xi_{(1)} \left(\frac{s(z_0)}{f_1(z_0)} + \frac{\rho_0 + s(z_0)}{f_2(z_0)} \right) + o(v_1 v_2),$$

by Propositions B.4–B.7. Then, we conclude the proof of part (a).

(b) The proof is identical to that of Theorem 2(b), and thus is omitted. \square

Proof of Theorem 3. As in Theorem 1, it is obvious that the maximum $\hat{\tau}$ is of the form $\hat{\tau} = \tau_0 + a \hat{v}$. Because by the uniform convergence in $|v| \leq 1$, $\hat{\eta}(v) = r(v) + O_p(h^{-1})$,

$$\hat{\gamma}(v) = r^2(v) + 2r(v)(\alpha_0 + h^{-1}) + 2\alpha_0 O_p(h^{-1}) + O_p(h^{-2} + ah^{-1})$$

and thus, with $\check{\gamma}(v) = r^2(v) + 2r(v)(\alpha_0 + h^{-1})$, we have that

$$\left| \arg \max_{|v| \leq 1} \hat{\gamma}(v) - \arg \max_{|v| \leq 1} \check{\gamma}(v) \right| = O_p(h^{-1}).$$

So, it suffices to prove that $\arg \max_{|v| \leq 1} \check{\gamma}(v) = O_p(h^{-1})$ to conclude.

By Propositions B.9–B.12, $\bar{\eta}(v) = E(r(v)) = d^{-1}C_1 v^2 1(v > 0) + dC_1 v^2 1(v \leq 0)$ where $C_1 = \alpha_0 k_{(1)}^-(0)/2$. Define $\bar{\gamma}(v) = \bar{\eta}(v)^2 + 2(\alpha_0 + h^{-1})\bar{\eta}(v)$. Then, by mimicking the arguments of the proof of Theorem 1, we obtain that $\hat{v} = O_p(h^{-1})$. \square

Proof of Corollary 3. The proof follows by identical arguments of those employed in Theorem 4, after one observes that, by Propositions B.9–B.12,

$$E(h^2 r(v)) = \frac{\alpha_0 f_2(\mu_2)}{2f_1(\mu_2)} k_{(1)}^-(0) v^2 + o(1)$$

whereas, by Propositions B.4–B.7,

$$\text{Cov}(h^2 r(v_1), h^2 r(v_2)) = v_1 v_2 \zeta_{(1)} \sigma_\varepsilon^2 \left[\frac{1}{f_1(\mu_2)} + \frac{1}{f_2(\mu_1)} \right] = 2v_1 v_2 \zeta_{(1)} \sigma_\varepsilon^2 \left[\frac{1}{f_1(\mu_2)} \right].$$

\square

Appendix B

In the following propositions we assume that, without loss of generality, $v > 0$, (the case $v < 0$ is similar). Also, we denote by ζ_0 and α_0 the break point and its size respectively. Define $\bar{g}(z) = f(x)g(z)$, $S^\pm(z, iv) = K^\pm(a^{-1}(z - i\omega(v))) - K^\pm(a^{-1}(z - i\omega(0)))$ where $i\omega(v) = iz_0(\zeta_0 + vah^{-1})$ with $iz_0(\zeta) = (iz_{01}, \dots, iz_{0(r-1)}, \zeta, iz_{0(r+1)}, \dots, iz_{0(p+1)})'$ and e_0 is a $p \times 1$ vector of zeroes with a 1 in the r th position. Propositions B.1–B.8 are proved under the assumptions of Theorem 2, while Propositions B.9–B.12 under those of Theorem 4. Dropping the subscripts i ,

Proposition B.1. $E[\mathbb{P}^+(v)] = [\bar{g}_{(1)}(z_0) + \alpha_0 f_{(1)}(x_0)] vah^{-1} + O(a^2 v^2 h^{-2})$.

Proof. According to (1),

$$E[\mathbb{P}^+(v)] = \frac{1}{a^{p+1}} \int_{-\infty}^{\infty} [\bar{g}(z) + \alpha_0 f(x)] S^+(z, v) dz.$$

Then, apply Lemma C.2. \square

Proposition B.2. $E[\mathbb{P}^-(v)] = -\alpha_0 f(x_0) k_{(1)}^-(0) v^2/2h^2 + \bar{g}_{(1)}(z_0) \nu a h^{-1} + O(a^2 v^2 h^{-2})$.

Proof. By Lemma C.2,

$$E[\mathbb{P}^-(v)] = \frac{\alpha_0}{a^{p+1}} \int_{z_0}^{z_0 + \nu a h^{-1}} f(x) K^-\left(\frac{z - \omega(v)}{a}\right) dz + \bar{g}_{(1)}(z_0) \frac{\nu a}{h} + O\left(\frac{v^2 a^2}{h^2}\right).$$

By a change of variable and Bochner's (1955) Theorem, the integral is

$$\begin{aligned} \alpha_0 \int_{-v/h}^0 f\left(a\left(u_0 + \frac{\nu e_0}{h}\right) + x_0\right) K^-(u) du &= \alpha_0 f(x_0) \int_{-v/h}^0 k^-(t) dt \\ &+ O\left(a \int_{-v/h}^0 k^-(t) dt\right), \end{aligned}$$

where $u = (u'_0, u_{p+1})'$. Then, the result follows because by B3,

$$\int_{-v/h}^0 k^-(t) dt = -\frac{v^2}{2h^2} k_{(1)}^-(0) + o\left(\frac{v^2}{h^2}\right). \quad \square$$

Proposition B.3. $E[\mathbb{F}^\pm(v)] = f_{(1)}(x_0) \nu a h^{-1} + O(a^2 v^2 h^{-2})$.

Proof. As in Propositions B.1 and B.2 by continuity of $f(\cdot)$, after replacing $\bar{g}(z)$ and α_0 by $f(x)$ and 0 respectively. \square

Define $\sigma_+^2(z_0) = s(z_0) + \rho_0$ and $\sigma_-^2(z_0) = s(z_0)$.

Proposition B.4.

$$\begin{aligned} \text{Cov}\{\mathbb{P}^\pm(v_1), \mathbb{P}^\pm(v_2)\} &= h^{-4} f(x_0) v_1 v_2 [m^\pm(z_0)]^2 \\ &+ \sigma_\pm^2(z_0) \xi_{(1)} + o(v_1^2 v_2^2 h^{-4}). \end{aligned}$$

Proof. The left side of the above equation is

$$\begin{aligned} &\frac{2}{h^4} \sum_{i=1}^T \sum_{j>i} \text{Cov}\{m(Z_i) S^\pm(Z_i, v_1), m(Z_j) S^\pm(Z_j, v_2)\} \\ &+ \frac{T}{h^4} \text{Cov}\{m(Z_i) S^\pm(Z_i, v_1), m(Z_i) S^\pm(Z_i, v_2)\} \\ &+ \frac{T}{h^4} E\{\sigma^2(Z_i) S^\pm(Z_i, v_1) S^\pm(Z_i, v_2)\}. \end{aligned}$$

Then apply Lemmas C.4, C.6 and C.8. \square

Proposition B.5. $\text{Cov} \{ \mathbb{F}^\pm(v_1), \mathbb{F}^\pm(v_2) \} = h^{-4} f(x_0) v_1 v_2 \xi_{(1)} + o(v_1^2 v_2^2 h^{-4})$.

Proof. By Lemmas C.6 and C.8 as in Proposition B.4, since $f(\cdot)$ is twice continuously differentiable, $m^\pm(z) = 1$ and $\sigma_\pm^2(z) = 0$ in this case. \square

Proposition B.6. $\text{Cov} \mathbb{P}^\pm(v_1), \mathbb{F}^\pm(v_2) = h^{-4} f(x_0) v_1 v_2 m^\pm(z_0) \xi_{(1)} + o(v_1^2 v_2^2 h^{-4})$.

Proof. The left side of the above equation is

$$\begin{aligned} & \frac{2}{h^4} \sum_{i=1}^T \sum_{j>i} \text{Cov} \{ m(Z_i) S^\pm(Z_i, v_1), S^\pm(Z_j, v_2) \} \\ & + \frac{T}{h^4} \text{Cov} \{ m(Z_i) S^\pm(Z_i, v_1), S^\pm(Z_i, v_2) \}. \end{aligned}$$

Then, by Lemmas C.6 and C.8 as in Proposition B.4 or B.5. \square

Proposition B.7. For $i, j = 1, 2$, $\text{Cov} \{ \mathbb{P}^\pm(i v_1), \mathbb{P}^\mp(j v_2) \} = o(i v_1^2 j v_2^2 h^{-4})$, $\text{Cov} \mathbb{F}^\pm(i v_1), \mathbb{F}^\mp(j v_2) = o(i v_1^2 j v_2^2 h^{-4})$, $\text{Cov} \mathbb{P}^\pm(i v_1), \mathbb{F}^\mp(j v_2) = o(i v_1^2 j v_2^2 h^{-4})$.

Proof. By Lemmas C.7 and C.9. \square

Dropping the superscripts i ,

Proposition B8. $\lim_{T \rightarrow \infty} h^4 \text{Var}[\mathbb{P}^\pm(v_1) - \mathbb{P}^\pm(v_2)] \leq C |v_1 - v_2|^2$ and $\lim_{T \rightarrow \infty} h^4 \text{Var}[\mathbb{F}^\pm(v_1) - \mathbb{F}^\pm(v_2)] \leq C |v_1 - v_2|^2$.

Proof. By Propositions B.4 and B.5. \square

For the next four Propositions, let us introduce the following notation. For $i = 1, 2$, denote $g(z) f_i(x_0)$ as $\bar{g}_i(z)$.

Proposition B9. $\mathbb{E}[\mathbb{P}^+(v)] = -\frac{1}{2} [\bar{g}_2(z_0) + \alpha_0 f_2(x_0)] k_{(1)}^-(0) v^2 h^{-2} + O(a^2 v^2 h^{-2})$.

Proof. According to (8), by Lemma C.3

$$\mathbb{E}[\mathbb{P}^+(v)] = \frac{1}{a^{p+1}} \int_{-\infty}^{\infty} [\bar{g}_2(z) + \alpha_0 f_2(x)] S^+(z, v) dz,$$

then apply Lemma C.2 as in Proposition B.1. \square

Proposition B.10.

$$\begin{aligned} \mathbb{E}[\mathbb{P}^-(v)] &= -\frac{(\alpha_0 f_2(x_0) + f_0(x_0) g(z_0)) k_{(1)}^-(0) v^2}{2h^2} - \frac{\bar{g}_1(z_0) k_{(1)}^-(0) v^2}{2h^2} \\ &+ O\left(\frac{v^2 a^2}{h^2}\right). \end{aligned}$$

Proof. Using the same arguments of Proposition B.2, by Lemma C.3, and recalling that $f_0(x) = f_2(x) - f_1(x)$,

$$\begin{aligned} \mathbb{E}[\mathbb{P}^-(v)] &= \frac{1}{a^{p+1}} \int_{\tau_0}^{\tau_0 + \nu a h^{-1}} (\alpha_0 f_2(x) + g(z) f_0(x)) K^-\left(\frac{z - \omega(v)}{a}\right) dz \\ &\quad - \frac{\bar{g}_1(x_0) k_{(1)}^-(0) v^2}{2h^2} + \mathcal{O}\left(\frac{v^2 a^2}{h^2}\right). \quad \square \end{aligned}$$

Proposition B.11. $\mathbb{E}[\mathbb{F}^+(v)] = -f_2(x_0) k_{(1)}^-(0) v^2 h^{-2} / 2 + \mathcal{O}(a^2 v^2 h^{-2})$.

Proof. By Lemma C.3, as in Propositions B.9 and B.10, by continuity of $f_2(\cdot)$ after observing that $\bar{g}_2(z)$ is $f_2(x)$ and $\alpha_0 = 0$. \square

Proposition B.12. $\mathbb{E}[\mathbb{F}^-(v)] = -f_0(x_0) k_{(1)}^-(0) v^2 / 2h^2 - f_1(x_0) k_{(1)}^-(0) v^2 / 2h^2 + \mathcal{O}(a^2 v^2 h^{-2})$.

Proof. By Lemma C.3, as in Propositions B.9 and B.10, by continuity of both $f_1(x)$ and $f_0(x)$ after replacing $\bar{g}_1(z)$ and α_0 by $f_1(x)$ and 0 respectively there. \square

Appendix C

The following lemmas are based on the same assumptions as those of Theorem 2. It is also worthy to mention that the reason in Lemmas C.5, C.7 and C.9 to have, say ${}_i v_1$ and ${}_j v_2$, is because in Corollary 3, we evaluate the kernels $K^+(\cdot)$ and $K^-(\cdot)$ at two different points.

Lemma C.1. Let $q: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ be a continuous function. Then,

$$\frac{1}{a^{p+1}} \int_{-\infty}^{\infty} q(z) S^\pm(z, v) dz = o(1).$$

Proof. Using a change of variable, the left side of the above equation is equal to

$$\int_{-\infty}^{\infty} \left\{ q\left(\left(au + \frac{av e_0}{h}\right) + z_0\right) - q(au + z_0) \right\} K^\pm(u) du = o(1),$$

by Bochner's (1955) Theorem. \square

Lemma C.2. Let $q: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ be twice continuously differentiable function. If the kernel functions $K^\pm(\cdot)$ are symmetric, then

$$\frac{1}{a^{p+1}} \int_{-\infty}^{\infty} q(z) S^\pm(z, v) dz = q_{(1)}(z_0) \frac{av}{h} + \mathcal{O}\left(\frac{a^2 v^2}{h^2}\right).$$

Proof. By Taylor's expansion of $q(z)$

$$\begin{aligned}
\frac{1}{a^{p+1}} \int_{-\infty}^{-\infty} q(z) S^{\pm}(z, v) dz &= \frac{av}{h} \int q_{(1)}(au + z_0) K^{\pm}(z, v) dz \\
&\quad + \frac{a^2 v^2}{h^2} \int q_{(2)}(au\gamma + z_0) K^{\pm}(z, v) dz \\
&= \frac{av}{h} q_{(1)}(z_0) + \frac{a^3 v}{h} + \mathcal{O}\left(\frac{a^2 v^2}{h^2}\right) \\
&= \frac{av}{h} q_{(1)}(z_0) + \mathcal{O}\left(\frac{a^2 v^2}{h^2}\right).
\end{aligned}$$

by Lemma C.1 and where the second equality comes because the function $q_{(1)}(\cdot)$ is twice continuously differentiable for all coordinates different than the r th one, and the third equality by assumption B5. \square

The next lemma will be useful in the case where the time is responsible of the break but it is not a regressor of the model, that is the model considered in Section 4.

Lemma C.3. Let $q: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ be twice continuously differentiable function. If the kernel functions $K^{\pm}(\cdot)$ are symmetric, then

$$\frac{1}{a^{p+1}} \int_{-\infty}^{\infty} q(z) S^{\pm}(z, v) dz = -q(x_0) \frac{k_{(1)}^-(0)v^2}{2h^2} (1 + \mathcal{O}(a^2)).$$

Proof. The left side of the above equation is

$$\begin{aligned}
\frac{1}{a^{p+1}} \int_{-\infty}^{\infty} q(x) \prod_{j=1}^p k\left(\frac{x_j - x_{0j}}{a}\right) &\left(\int_{-\infty}^{\infty} \left(k^{\pm}\left(\frac{\tau - (\tau_0 + avh^{-1})}{a}\right) \right. \right. \\
&\quad \left. \left. - k^{\pm}\left(\frac{\tau - \tau_0}{a}\right) \right) d\tau \right) dx.
\end{aligned}$$

Consider the case with $k^-(\cdot)$, being the case with $k^+(\cdot)$ identical. By a change of variables, it is

$$\int_{-\infty}^{\infty} q(au + x_0) \prod_{j=1}^p k(u_j) \left(\int_{-vh^{-1}}^0 k^-(t) dt \right) du = -q(x_0) \frac{k_{(1)}^-(0)v^2}{2h^2} (1 + \mathcal{O}(a^2))$$

because the properties of $k^-(t)$ and $k(u_j)$. \square

Lemma C.4. Let $q: \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ be a twice continuously differentiable function. Then,

$$\frac{T}{h^4} \int_{-\infty}^{\infty} q(z) S^{\pm}(z, v_1) S^{\pm}(z, v_2) dz = \frac{1}{h^4} q(z_0) v_1 v_2 \xi_{(1)} + o\left(\frac{1}{h^4}\right).$$

Proof. Without loss of generality, assume that $0 < v_1 < v_2$. After a change of variable

$$\frac{T}{h^4} \int_{-\infty}^{\infty} q(z) S^+(z, v_1) S^+(z, v_2) dz = D_1 - D_2 + D_3,$$

where

$$\begin{aligned} D_1 &= \theta_0 h^{-2} \int_0^{v_1/h} q(aue_0 + z_0) k^+(u)^2 du, \\ D_2 &= \theta_0 h^{-2} \int_{v_1/h}^{v_2/h} q(aue_0 + z_0) k^+(u) \left\{ k^+\left(u - \frac{v_1}{h}\right) - k^+(u) \right\} du, \\ D_3 &= \theta_0 h^{-2} \int_{v_2/h}^{\infty} q(aue_0 + z_0) \left\{ k^+\left(u - \frac{v_1}{h}\right) - k^+(u) \right\} \left\{ k^+\left(u - \frac{v_2}{h}\right) - k^+(u) \right\} du, \end{aligned}$$

where $\theta_0 = (\int k(t)^2 dt)^p$. For constants $\varphi_1, \varphi_2, \varphi_3 \in (0, 1)$,

$$\begin{aligned} D_1 &= \theta_0 h^{-2} q(z_0) \int_0^{v_1/h} k^+(t)^2 dt + O\left(ah^{-2} \int_0^{v_1/h} k^+(t)^2 dt\right), \\ D_2 &= \theta_0 h^{-3} q(z_0) v_1 \int_{v_1/h}^{v_2/h} k^+(t) k_{(1)}^+\left(t - \frac{\varphi_1 v_1}{h}\right) dt \\ &\quad + O\left(ah^{-3} \int_{v_1/h}^{v_2/h} k^+(t) k_{(1)}^+\left(t - \frac{\varphi_1 v_1}{h}\right) dt\right), \\ D_3 &= \theta_0 h^{-4} q(z_0) v_1 v_2 \int_{v_1/h}^{\infty} k_{(1)}^+\left(t - \frac{\varphi_2 v_1}{h}\right) k_{(1)}^+\left(t - \frac{\varphi_3 v_2}{h}\right) dt \\ &\quad + O\left(ah^{-4} \int_{v_1/h}^{\infty} k_{(1)}^+\left(t - \frac{\varphi_2 v_1}{h}\right) k_{(1)}^+\left(t - \frac{\varphi_3 v_2}{h}\right) dt\right), \end{aligned}$$

by MVT and Bochner's (1955) theorem. The lemma follows because $\xi_{(1)} = \theta_0 \int_0^{\infty} k_{(1)}^+(t)^2 dt$ and

$$\begin{aligned} \int_{v_1/h}^{v_2/h} k^+(t) k_{(1)}^+\left(t - \frac{\varphi_1 v_1}{h}\right) dt &= \frac{1}{2} h^{-2} (v_2^2 - v_1^2) k_{(1)}^+(0) \int_{v_1/h}^{v_2/h} k_{(1)}^+\left(t - \frac{\varphi_1 v_1}{h}\right) dt \\ &\quad + o\left(\frac{1}{h^2}\right), \end{aligned}$$

$$\int_0^{v_1/h} k^+(t)^2 dt = \frac{1}{3} v_1^3 h^{-3} k_{(1)}^+(0)^2 + o\left(\frac{1}{h^3}\right),$$

$$\int_{v_1/h}^{\infty} k_{(1)}^+\left(t - \frac{\varphi_2 v_1}{h}\right) k_{(1)}^+\left(t - \frac{\varphi_3 v_2}{h}\right) dt = \int_0^{\infty} k_{(1)}^+(t)^2 dt + o(1),$$

because $k^\pm(0) = 0$. Also by similar, if not easier, arguments

$$\frac{T}{h^4} \int_{-\infty}^{\infty} q(z) S^-(z, v_1) S^-(z, v_2) = \frac{1}{h^4} q(z_0) v_1 v_2 \zeta_{(1)} + o\left(\frac{1}{h^4}\right). \quad \square$$

Lemma C.5. Let $q(\cdot)$ be as in Lemma C.4. Then, for $i, j = 1, 2$,

$$\frac{T}{h^4} \int_{-\infty}^{\infty} q(z) S^\pm(z, i v_1) S^\mp(z, j v_2) dz = o\left(\frac{1}{h^4}\right).$$

Proof. Without loss of generality, assume that $0 < i v_1 < j v_2$. After a change of variable

$$\begin{aligned} & \frac{T}{h^4} \int_{-\infty}^{\infty} q(z) S^+(z, i v_1) S^-(z, j v_2) dz \\ &= h^{-2} \int_{i v_1/h}^{j v_2/h} q(au + i z_0) \left(K^- \left(u - \frac{v_2}{h} e_0 + \frac{i z_0 - j z_0}{a} (\underline{1} - e_0) \right) \right. \\ & \quad \times \left. \left\{ K^+ \left(u - \frac{i v_1}{h} e_0 \right) - K^+(u) \right\} \right) du + h^{-2} \int_0^{j v_2/h} q(au + i z_0) \\ & \quad \times K^- \left(u - \frac{j v_2}{h} e_0 + \frac{i z_0 - j z_0}{a} (\underline{1} - e_0) \right) K^+(u) du = o(h^{-4}), \end{aligned}$$

where $\underline{1}$ is a vector of ones, proceeding as in the proof of D_1 and D_2 of Lemma C.4 and that if $i z_{0\ell} - j z_{0\ell} \neq 0$

$$\int \prod_{\ell \neq r} k\left(u + \frac{i z_{0\ell} - j z_{0\ell}}{a}\right) q(u) du \rightarrow 0. \quad \square$$

Lemma C.6. Let $q_1(\cdot)$ and $q_2(\cdot)$ be as in Lemma C.4. Then,

$$\sum_{i=1}^T \sum_{j>i} \text{Cov} [q_1(Z_i) S^\pm(Z_i, v_1), q_2(Z_j) S^\pm(Z_j, v_2)] = O\left(\frac{T a^{2(p+1)}}{h^2}\right).$$

Proof. By Lemmas 8.2 and 8.3 of Robinson (1983). \square

Lemma C.7. Let $q_1(\cdot)$ and $q_2(\cdot)$ be as in Lemma C.4. Then,

$$\sum_{i=1}^T \sum_{j>i} \text{Cov} [q_1(Z_i) S^\pm(Z_{i\ell_1} v_1), q_2(Z_j) S^\mp(Z_{j\ell_2} v_2)] = O\left(\frac{Ta^{2(p+1)}}{h^2}\right).$$

Proof. By a routine extension of Lemmas 8.2 and 8.3. of Robinson (1983). \square

Lemma C.8. Let $q_1(\cdot)$ and $q_2(\cdot)$ be as in Lemma C.4. Then,

$$\begin{aligned} \frac{T}{h^4} \text{Cov} [q_1(Z_i) S^\pm(Z_i, v_1), q_2(Z_i) S^\pm(Z_i, v_2)] &= \frac{1}{h^4} f(x_0) q_1(z_0) q_2(z_0) v_1 v_2 \xi_{(1)} \\ &+ o\left(\frac{1}{h^4}\right). \end{aligned}$$

Proof. By Lemma C.2, $E(q_j(Z_i) S^\pm(Z_i, v_j)) = O(a^{p+2}/h)$, $j = 1, 2$. Then, by Lemma C.4. \square

Lemma C.9. Let $q_1(\cdot)$ and $q_2(\cdot)$ be as in Lemma C.4. Then,

$$\frac{T}{h^4} \text{Cov} [q_1(Z_i) S^\pm(Z_{i,i} v_1), q_2(Z_i) S^\mp(Z_{i,j} v_2)] = o\left(\frac{1}{h^4}\right).$$

Proof. By Lemmas C.2 and C.5. \square

References

- Andrews, D.W.K., 1993. Testing for structural instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
- Antoch, J., Huskova, M., 1997. Estimators of changes. Working paper, Charles University, Praha.
- Bai, J., 1994. Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453–472.
- Bai, J., 1995. Least absolute deviation estimation of a shift. *Econometric Theory* 11, 403–436.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.
- Bickel, P.J., Wichura, M., 1971. Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics* 42, 1656–1670.
- Billingsley, P., 1968. *Convergence of Probability Measures*. Wiley, New York.
- Bochner, S., 1955. *Harmonic Analysis and the Theory of Probability*. University of Chicago Press, Chicago, Illinois.
- Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B37*, 149–192.
- Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

- Chu, J.S., Wu, C.K., 1993. Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* 21, 1545–1566.
- Delgado, M.A., Hidalgo, J., 1995. Nonparametric inference on structural breaks. Mimeo, Department of Econometrics, Universidad Carlos III.
- Eddy, W.F., 1980. Optimum kernel estimators of the mode. *Annals of Statistics* 8, 870–882.
- Eubank, R.L., Speckman, P.L., 1994. Nonparametric estimation of functions with jump discontinuities. *Lecture Notes*, Vol. 23, Institute of Mathematical Statistics 130–144.
- Feder, P.I., 1975. On asymptotic distribution theory in segmented regression problem-identified case. *Annals of Statistics* 3, 49–83.
- Fiteni, I., 1998. Robust estimation of structural break points. Working paper, Universidad Carlos III de Madrid.
- Gasser, T., Müller, H.G., Mammitzsch, V., 1985. Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society Series B47*, 238–252.
- Györfi, L., Härdle, W., Sarda, P., Vieu, P., 1989. Nonparametric curve estimation from time series. *Lecture Notes in Statistics*, Vol. 60, Springer, Berlin.
- Hansen, B., 1998. Testing for structural change in conditional models. Working paper, University of Wisconsin.
- Hawkins, D.M., 1977. Testing a sequence of observations for a change in location. *Journal of the American Statistical Association* 72, 180–186.
- Hidalgo, J., 1995. A nonparametric conditional moment test for structural stability. *Econometric Theory* 11, 671–698.
- Hinkley, D., 1969. Inference about the intersection in two phase regression. *Biometrika* 56, 495–504.
- Hinkley, D., 1970. Inference about the change point in a sequence of random variables. *Biometrika* 57, 1–17.
- Kim, H.J., Siegmund, D., 1989. The likelihood ratio test for a change point in simple linear regressions. *Biometrika* 76, 409–423.
- Müller, H.-G., 1992. Change-points in nonparametric regression analysis. *Annals of Statistics* 20, 737–761.
- Müller, H.-G., Song, K.-S., 1994. Maximin estimation of multidimensional boundaries. *Journal of Multivariate Analysis* 50, 265–281.
- Prakasa Rao, B.L.S., 1981. *Nonparametric Functional Estimation*. Academic Press, Orlando.
- Quandt, R.E., 1960. Test of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 55, 324–330.
- Rice, J., 1984. Boundary modification for kernel regression. *Communications in Statistics: Theory and Methods* 13, 893–900.
- Robinson, P.M., 1983. Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
- Robinson, P.M., 1988. Root- N -consistent semiparametric regression. *Econometrica* 56, 931–954.
- Whitt, W., 1970. Weak convergence of probability measures on the function space $C[0, \infty)$. *Annals of Mathematical Statistics* 41, 939–944.
- Worsley, K.L., 1979. On the likelihood ratio test for a shift in locations of normal populations. *Journal of the American Statistical Association* 74, 365–367.
- Yao, Y.-C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Annals of Statistics* 3, 1321–1328.
- Yin, Y.Q., 1988. Detection of the number, locations and magnitudes of jumps. *Communications in Statistics: Stochastic Models* 4, 445–455.