Working Paper 94-48
Statistics and Econometrics Series 16
December 1994

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

NONPARAMETRIC ESTIMATION OF STRUCTURAL BREAK POINTS

Miguel A. Delgado and Javier Hidalgo*

Abstract

This paper proposes point and interval estimates of location and size of jumps in multiple regression curves or its derivatives. We are mainly concerned with time series models where structural breaks occur at a given period of time or they are explained by the value taken by some predictor (e.g. threshold models). No previous knowledge of the underlying regression function is required. Left and right limits of the function, with respect to the regressor explaining the break, are estimated at each data point using multivariate multiplicative kernels. The univariate kernel corresponding to the regressor explaining the break is one-sided, with all its mass at the right or left of zero. Since left and right limits are the same, except at the break point, the location of the jump is estimated as the observed regressor value maximizing the difference between left and right limit estimates. This difference, evaluated at the estimated location point, is the estimation of the jump size. A small Monte Carlo study and an empirical application to USA macroecomic data illustrates the performance of the procedure in small samples. The paper also discusses some extensions, in particular the identification of the coordinate explaining the break, the application of the procedure to the estimation of parametric models, and robustification of the method for the influence of outliers.

Key Words

Structural Breaks; Nonparametric Regression; One-sided Kernels; Strong Mixing Processes.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Hidalgo, London School of Economics. Article based on research funded by Spanish Dirección General de Investigación Científica y Técnica (DGICYT), reference number PB92-0247. We thank Pierre Perron for providing the data used in Section 5. Previous versions of the paper have circulated with the title "Inference on Structural Breaks".

1. INTRODUCTION

This paper proposes point and interval estimates of location and size of jumps in multivariate regression curves or its derivatives. We consider dependent observations. Regressors can be strong mixing, possibly nonstationary, or fixed, possibly trends. Thus, the point break can happen at a given period of time or it can be explained by the value taken by some predictor, as is the case in threshold models. The underlying functional form of the regression function is left unspecified.

Structural breaks produce inconsistent estimates, both in a parametric or in a nonparametric context, if they are not taken into consideration. Testing the presence of structural breaks and, more importantly, estimating its location is one of the first steps in model building. There is a large literature on testing structural breaks when the possible timing of the break is unknown. In parametric approaches, a parametric regression function is assumed before and after the jump. Page (1955), Quandt (1960), Hinkley (1969, 1971), Brown, Durbin and Evans (1975), Worsley (1979), and Kim and Siegmund (1989), among others, use independent and identically distributed observations (iid), see also the survey papers by Zacks (1983) and Krishnaiah and Miao (1988). However, there is a lot of empirical evidence on structural breaks in time series dynamic models (see, e.g., Delong and Summers (1988), Perron (1989), and Hendry and Erickson (1991)). Testing structural breaks in linear dynamic models has been considered by Kramer, Ploberger and Alt (1988), and others. Recently, Andrews (1993) has proposed tests for general parametric models under weak dependent data.

Parametric estimation methods for the break point are usually based on the maximum likelihood principle. Hinkley (1971), Hawkins (1977), Bhattacharya (1987), and Yao (1987) consider a shift in location, and Feder (1975) a change in a segmented linear regression with fixed regressors and iid errors. Picard (1985) study the case of a stationary autorregressive process and Bai (1992) more general time series models. These papers deal with segmented, but continuous, linear models.

Parametric estimates can be inconsistent when the functional form of the regression functions before and after the break point are misspecified, and tests based on their estimates can yield very misleading consequences (see Hidalgo 1994). Break point estimates, computed without assuming any parametric specification of the underlying regression function, are interesting by themselves and as a first step in parametric model specification. Based on ideas of Eddy (1980, 1982) for estimating the mode of a density, Müller (1992) has proposed estimates of location and size of jumps in a nonparametric regression (or its derivatives) where the only regressor is time, under iid errors. The size of the jumps are computed as the difference between a left and right-sided kernel estimate. A one-sided kernel has all its mass at the right or left of zero. The method seems to work very well in practice.

In this paper, we employ Müller's (1992) approach in a more general context, likely in econometric applications. We allow the regression function to depend on more than one regressor, and be stochastic, weakly dependent, or fixed, possibly trends. The break point is explained by the value taken by some of the regressors. If this regressor is time, the break point happens at a given period, because there is a change in the trend or because the general structure of the regression model has changed at this period. If the break is explained by some other predictor, the break point is the value of the predictor at which the regression function changes its structure, like in threshold models.

The rest of the paper is organized as follows. In the next Section, we present the estimation method. Section 3 discusses the assumptions required and presents the main theorems of the paper. In Section 4, we report some Monte Carlo simulations. In Section 5, the estimation method is applied to macroeconomic USA data. Finally, in Section 6, we discuss extensions to other conditional location functionals, and suggests ways of further work. Proofs are confined to a mathematical appendix.

2. ESTIMATING THE BREAK POINT AND THE SIZE OF THE BREAK

Let $\{(Y_1,X_1),(Y_2,X_2),\ldots,(Y_n,X_n)\}$ be observations of a multivariate stochastic process, where Y_t is scalar and $X_t= (X_{t1},X_{t2},\ldots,X_{tp})'$ is a p-dimensional vector. Define $E(Y_t | X_t = x) = m_t(x)$. In order to provide asymptotic justification of our estimators, it is convenient to regard $m_t(x)$ as being generated from functions $m(x, t)$ with domain $\mathcal{X} \times (0,1)$, where $\Pr(X_t \in \mathcal{X}) = 1$. That is, $m_t(x) = m(x, \tau_t)$, where $\tau_t = t/n$. Thus, the regressor "time" is defined on the interval (0,1), and it becomes dense as the sample size increases. The regression function depends on the sample size, a device common in the nonparametric literature; see, e.g., Nadaraya (1964), Gasser et al (1985) or Müller (1992). An excellent discussion on the interpretation of this device is in Robinson (1989). So, the regression function can be written compactly as $E(Y_t | X_t) = m(Z_t)$, where $Z_t = (X_t', \tau_t)'$. We consider two different structural break models,

$$m(Z_t) = m_-(Z_t)\, 1(\tau_t \le \tau_o) + m_+(Z_t)\, 1(\tau_t > \tau_o), \tag{2.1}$$

and

$$m(Z_t) = m_-(Z_t)\, 1(X_{tk} \le x_{ok}) + m_+(Z_t)\, 1(X_{tk} > x_{ok}), \text{ for some } k, \tag{2.2}$$

where $1(A)$ is the indicator function of the event A.

Model (2.1) includes changes in the trend of the series as well as changes in the regression model at a given point of time. A special case is the trend model

$$E(Y_t) = m_-(\tau_t)\, 1(\tau_t \le \tau_o) + m_+(\tau_t)\, 1(\tau_t > \tau_o), \tag{2.3}$$

considered by Müller (1992).

In model (2.2), the change is governed by the value taken by some of the regressors. This model is relevant both in time series and cross-sectional applications, being a classical example the threshold model.

In what follows, we refer to models (2.1) and (2.2) compactly as,

$$m(Z_t) = m_-(Z_t)\, 1(Z_{tk} \le \gamma_{ok}) + m_+(Z_t)\, 1(Z_{tk} > \gamma_{ok}), \tag{2.4}$$

where $\gamma_{0,p+1} = \tau_0$, and $\gamma_{ok} = x_{ok}$ for $1 \le k \le p$.

Based on the sample, we first estimate

$$\Delta^{(k)}(z)= m_{+}^{(k)}(z) - m_{-}^{(k)}(z), \tag{2.5}$$

where $m_{+}^{(k)}(z)$ and $m_{-}^{(k)}(z)$ are the left and right hand side limits of $m(z)$ with respect to the k-th regressor , i.e.

$$m_{+}^{(k)}(z)= \lim_{\delta\to 0_+} m(z + e_k\,\delta) \quad \text{and} \quad m_{-}^{(k)}(z)= \lim_{\delta\to 0_-} m(z + e_k\,\delta),$$

where e_k is a vector of zeros with a one in the k-th position.

The break point may be in the first derivative of the regression function with respect to the s-th regressor, and it can be explained by the k-th regressor, that is,

$$m_{(s)}(Z_t)= m_{-(s)}(Z_t)\ 1(Z_{tk}\le z_{k(s)}) + m_{+(s)}(Z_t)\ 1(Z_{tk}> z_{k(s)}). \tag{2.6}$$

where $m_{(s)}(z)= \partial m(z)/\partial z_S$. For instance, consider the trend break model

$$E(Y_t)= \beta_1+ \beta_2\tau_t+ \beta_3(\tau_t-\tau_0)1(\tau_t\le \tau_0).$$

Here, $\Delta^{(1)}(\tau)= 0$, all $\tau\in[0,1]$, but $\Delta^{(1)}_{(1)}(\tau_0)= \beta_3$. The jump may be in the regression function as well as in the derivative. For instance, in the AR(1) model,

$$E(Y_t|Y_{t-1})= (\beta_1 + \beta_2Y_{t-1})\ 1(\tau_t\le \tau_0) + (\beta_1 + \beta_3Y_{t-1})\ 1(\tau_t> \tau_0),$$

with $\beta_2\ne \beta_3$. Here, $\Delta^{(1)}(z)= (\beta_3-\beta_2)Y_{t-1}$ and $\Delta^{(1)}_{(1)}(z)= \beta_3-\beta_2$.

Based on the sample, we first estimate

$$\Delta^{(k)}_{(s)}(z) = m^{(k)}_{+(s)}(z) - m^{(k)}_{-(s)}(z),$$

where $m^{(k)}_{+(s)}(z)$ and $m^{(k)}_{-(s)}(z)$ are the left and right hand side limits of $m_{(s)}(z)$ with respect to the k-th regressor, i.e.

$$m^{(k)}_{+(s)}(z)= \lim_{\delta \to 0_+} m_{(s)}(z + e_k\,\delta) \quad \text{and} \quad m^{(k)}_{-(s)}(z)= \lim_{\delta \to 0_-} m_{(s)}(z + e_k\,\delta).$$

The analysis of higher order derivatives is identical but notationally much more cumbersome, without adding anything from a theoretical point of view; and therefore we will cincunscribe ourselves to the case where the jump is either in the conditional expectation or in its first derivative.

The function $\Delta^{(k)}(.)$ identifies the location of the break point z_{ok}. Notice that $\Delta^{(k)}(z)= 0$ for all $z = (z_1,z_2,\ldots,z_{p+1})'$ such that $z_k \neq z_{ok}$, and $\Delta^{(k)}(z)\neq 0$ only when $z_k = z_{ok}$ for any value of z_j, $j\neq k$. Therefore, whenever we identify a point z in $\mathbb{R}^{p+1}$ such that $\Delta^{(k)}(z)\neq 0$, we immediately identify z_{ok}. Let us introduce the real valued functions k: $\mathbb{R} \longrightarrow \mathbb{R}$ and K:$\mathbb{R}^{p+1} \longrightarrow \mathbb{R}$ such that $K(u)= \Pi^{p+1}_{k=1}\, k(u_k)$, $u=(u_1,\ldots,u_{p+1})'$. The classical kernel regression estimator of $m(z)$ is defined as

$$\hat{m}(z) = \frac{\sum_{t=1}^{n} Y_t\, K[(Z_t - z)/a]}{\sum_{t=1}^{n} K[(Z_t - z)/a]}. \tag{2.7}$$

The kernel estimator depends on the smoothing number $a= a(n)$, which converges to zero as the sample size does to infinity. More general kernel functions $K(.)$ can be used, and different smoothing numbers can be employed for the different regressors; see, e.g. Robinson (1983). We restrict ourselves to the above class of multiplicative kernels with the same smoothing number for each regressor for the sake of presentation and notational convenience. Consistency of the estimator (2.7) requires, despite $a + (a^{p+1}\, n)^{-1} \longrightarrow 0$ as $n \longrightarrow \infty$, that both the conditional expectation $m(z)$ and the probability density function of the regressor X_t are smooth enough.

This requirement for consistency, i.e. the smoothness of the conditional expectation $m(.)$, will motivate the estimator of the break point. The estimator in (2.7) is inconsistent at points of discontinuity of the regression curve, and, therefore, it is useless for the estimation of the jump $\Delta^{(k)}(z)$. On the other hand, if the regression function $m(.)$ is sufficiently smooth before and after the break point, we should be able to estimate consistently $m(.)$ at every other point. Furthermore, when only points before and after the break are used, we can expect that the kernel estimator will be consistent for $m_+^{(k)}(\underset{\sim}{z})$ and $m_-^{(k)}(\underset{\sim}{z})$, e.g. the right and left hand side limits of $m(\underset{\sim}{z})$ with respect to the k-th variable. Let us define $K_-^{(k)}(.)$ and $K_+^{(k)}(.)$ as

$$K_-^{(k)}(u)= k_-(u_k)\ \Pi_{r\neq k}^{p+1}\ k(u_r) \quad \text{and} \quad K_+^{(k)}(u)= k_+(u_k)\ \Pi_{r\neq k}^{p+1}\ k(u_r), \tag{2.8}$$

where $k_-(.)$ and $k_+(.)$ are kernels with domain in $\mathbb{R}_-$ and $\mathbb{R}_+$ respectively. For instance, $k_\pm(u)= k(u)\ 1(u \gtrless 0)/\int k(u)\ 1(u \gtrless 0)\ d\ u$ is a valid one-sided kernel to estimate $m_\pm^{(k)}(\underset{\sim}{z})$, where, at least, $k(.)$ integrates to one. However, more restrictive kernels must be used to estimate the location of the jump, as it will be seen in the next section.

Therefore, the estimation of $\Delta^{(k)}(\underset{\sim}{z})$ will be based on

$$\hat{\Delta}^{(k)}(\underset{\sim}{z})= \hat{m}_+^{(k)}(\underset{\sim}{z}) - \hat{m}_-^{(k)}(\underset{\sim}{z}), \tag{2.9}$$

where

$$\hat{m}_+^{(k)}(\underset{\sim}{z}) = \frac{\sum_{t=1}^n Y_t\ K_+[(Z_t - \underset{\sim}{z})/a]}{\sum_{t=1}^n K_+[(Z_t - \underset{\sim}{z})/a]} \quad \text{and} \quad \hat{m}_-^{(k)}(\underset{\sim}{z}) = \frac{\sum_{t=1}^n Y_t\ K_-[(Z_t - \underset{\sim}{z})/a]}{\sum_{t=1}^n K_-[(Z_t - \underset{\sim}{z})/a]},$$

and $\Delta_{(s)}^{(k)}(\underset{\sim}{z})$ will be estimated by

$$\hat{\Delta}_{(s)}^{(k)}(\underset{\sim}{z})= \partial\hat{\Delta}^{(k)}(\underset{\sim}{z})/\partial \underset{\sim}{z}_s. \tag{2.10}$$

These estimates are similar to those used in kernel regression at the boundary of the regressors domain, e.g. if $Z_t \in [0,1]$ and we try to estimate $E(Y|Z= 0$ or $1)$, which are well known in the statistical literature; see, e.g., Rice (1984), and Gasser, Müller and Mammitzch (1985). Therefore, we can base our estimator of the point (time) of the structural break as the k-th coordinate of $\underset{\sim}{z}$ such that it maximizes the quantities $|\hat{\Delta}^{(k)}(\underset{\sim}{z})|$ or $|\hat{\Delta}_{(s)}^{(k)}(\underset{\sim}{z})|$ respectively. Due to technical problems, which will become apparent in the following section, it is convenient to modify slightly the above criterium, when $p > 0$.

For any positive continuous function $W:\mathbb{R}^{p+1} \longrightarrow \mathbb{R}_+$, $W(\underset{\sim}{z})\ \Delta^{(k)}(\underset{\sim}{z})$ will be different that zero only for those values of $\underset{\sim}{z}$ such that $\underset{\sim}{z}_k = \underset{\sim}{z}_{0k}$. The function

$W(.)$ must satisfy, at least, that $W(z)= 0$ as $\|z\| \longrightarrow \infty$, where $\|.\|$ means Euclidean norm. In fact, $W(z)$ tries to avoid those regions of $\mathbb{R}^{p+1}$ where it is not likely to find data. This suggests to estimate z_{0k} by the k-th coordinate of the vector maximizing $|\hat{\Delta}^{(k)}(z)\ W(z)|$. Different weights will produce different vectors - with the same k-th coordinate - maximizing the objective function. In the next section, we will justify that the function minimizing the variance of the break point estimate corresponds to $W(z) = f_X(x)^r$, r> 0, where $z = (x', z_{p+1})'$, and

$$f_X(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} f_{X_t}(x).$$

where $f_{X_t}(.)$ is the probablilty density function (pdf) of X_t. Recall that z_{p+1} corresponds to the regressor "time", i.e. τ_t, which can be considered independent of X_t and achieves, asymptotically, a uniform pdf in the unit interval. Thus, we propose to estimate z_{0k} by $\hat{z}_{0k}^{(k)}$, where $\hat{z}_0^{(k)} = \left(\hat{z}_{01}^{(k)}, \hat{z}_{02}^{(k)}, \ldots, \hat{z}_{0,p+1}^{(k)}\right)'$ and

$$\hat{z}_0^{(k)} = \underset{z=(x', z_{p+1})'}{\arg\max} |\hat{\Delta}^{(k)}(z)\ f_X(x)^r|,\ r> 0. \tag{2.11}$$

Notice that $\hat{z}_0^{(k)}$ is an estimate of

$$z_0^{(k)} = \underset{z=(x', z_{p+1})'}{\arg\max} |\Delta^{(k)}(z)\ f_X(x)^r|,\ r>0. \tag{2.12}$$

since for any $\delta \neq 0$, $\Delta^{(k)}(z_0^{(k)} + e_k\delta) = 0$. Note that z_{0k} is the k-th element of the vector $z_0^{(k)}$. If we assume that $\Delta^{(k)}(z) = \Delta\ 1(z_k \leq z_{0k})$, where Δ is a constant, the point maximizing $|\Delta^{(k)}(z) f_X(x)^r|$, r> 0, is independent of r. In fact

$$z_0^{(k)} = \underset{\{z=(x', z_{p+1})' : z_k = z_{0k}\}}{\arg\max} |f_X(x)|.$$

That is, we are maximizing the averaged joint density fixing $z_k = z_{0k}$. Intuitively, we are choosing, among all the points z such that the jump is different than zero, the point z "more likely" - i.e. with greater density-.

Similarly, $z_{(s)k}$ is estimated by $\hat{z}^{(k)}_{(s)k}$, where

$\hat{z}^{(k)}_{(s)} = \left(\hat{z}^{(k)}_{(s)1}, \hat{z}^{(k)}_{(s)2}, \ldots, \hat{z}^{(k)}_{(s)p+1}\right)'$ and

$$\hat{z}^{(k)}_{(s)} = \underset{z=(x', z_{p+1})}{\arg\max} \; |\hat{\Delta}^{(k)}_{(s)}(z) f_X(x)^r|, \; r > 0, \tag{2.13}$$

where $\hat{\Delta}^{(k)}_{(s)}(z) = d\, \hat{\Delta}^{(k)}(z)/dz_s$. The break point is estimated by the k-th element of $\hat{z}^{(k)}_{(s)}$, and the size of the jump by $\hat{\Delta}^{(k)}_{(s)}(\hat{z}^{(k)}_{(s)})$.

Since $f_X(x)$ is unknown, it is estimated by

$$\tilde{f}_X(z) = (\tilde{f}^{(k)}_{+}(z)\, \tilde{f}^{(k)}_{-}(z))^{1/2}, \tag{2.14}$$

where

$$\tilde{f}^{(k)}_{+}(z) = (n\, a^{p+1})^{-1} \textstyle\sum_{j=1}^{n} K_{+}((Z_j - z)/a) \text{ and } \tilde{f}^{(k)}_{-}(z) = (n\, a^{p+1})^{-1} \textstyle\sum_{j=1}^{n} K_{-}((Z_j - z)/a),$$

where $z = (x', z_{p+1})'$, for any value of z_{p+1}. Note that the density of the regressor time, $Z_{tp+1} = \tau_t$, is uniform on the interval [0,1], and Z_{tp+1} is independent of the other regressors. This is why $\tilde{f}_X(z)$ is in fact estimating $f_X(x)$ for any value of z_{p+1}. Taking r= 2 in (2.13), (2.14) is very convenient because it avoids the random denominator of the regression estimates in (2.12). Futhermore, any other density estimate choice, like the the usual kernel density estimator, will yield inefficient break point estimates.

Feasible alternatives to $\hat{z}^{(k)}_0$ and $\hat{z}^{(k)}_{(s)}$ are given by

$$\tilde{z}^{(k)}_0 = \underset{z \in \mathbb{Z}_n}{\arg\max} |\hat{\Delta}^{(k)}(z)\, \tilde{f}_X(z)^r| \quad \text{and} \quad \tilde{z}^{(k)}_{(s)} = \underset{z \in \mathbb{Z}_n}{\arg\max} \; |\hat{\Delta}^{(k)}_{(s)}(z)\, f^{\sim}_X(z)^r|, \tag{2.16}$$

respectively, where $\mathbb{Z}_n = \{Z_1, Z_2, \ldots, Z_n\}$. In the next section, it will be provided regularity conditions under which these estimates and those defined in (2.12) and (2.13) have the same limiting distribution.

3. ASYMPTOTIC RESULTS

We analyze here only the case where the only fixed regressor entering in the regression function is a trend. The case for other fixed regressors can be handled in a similar way as with trending regressors. It is only required that the observations become more and more dense in the domain of such a fixed regressor as the sample size increases. The stochastic regressors are serially dependent in the following sense.

Definition: $\{X_t,\ t= 0,\ \pm 1,\ \pm 2,...\}$ is a strong mixing stochastic process with coefficients $\alpha(m)$ if $\lim_{m\to\infty} \alpha(m) = 0$, where

$$\alpha(m)= \sup_t \left\{ \sup_{\{\mathcal{A}\in\mathcal{M}^{t}_{-\infty},\ \mathcal{B}\in\mathcal{M}^{\infty}_{t+m}\}} \left| P(\mathcal{A}\cap\mathcal{B})-P(\mathcal{A})P(\mathcal{B}) \right| \right\},$$

and $\mathcal{M}^b_a$ is the σ-algebra generated by $\{Z_t,\ a\le t\le b\}$.

We need the following assumptions on the data generating process,

A1. The sequence $\{X_t,\ t= 0,\ \pm1,\ \pm2,...\}$ is a strong mixing process, with $\sup_t E\|m(X_t)\|^{\delta+2}< \infty$ for some $\delta> 0$, and mixing coefficients $\alpha(m)$ satisfying $\sum_{m=q}^{\infty} \alpha(m)^{\delta/(2+\delta)}= O(1/q)$.

A2. $\{\varepsilon_t\}$ is an independent sequence of random variables, with $E\{\varepsilon_t|X_s,\ s= 0,\ \pm 1,\ \pm 2,\ ...\}= 0$ and $E\{\varepsilon_t^2|X_s,\ s= 0,\ \pm 1,\ \pm 2,\ ...\}= \sigma^2< \infty$.

A3. The pdf $f_{X_t}(x)$ of X_t has, at least, two continuous derivatives with respect to all its arguments, and the regression function $m(z)$ has, at least, two continuous derivatives with respect to all its arguments, except at the break point where $z_k= z_{ok}$.

A4. $m(z)= m^*(z) + \Delta\ 1(z_k \le z_{ok})$, where $m^*(z)$ is continuous in $\mathbb{R}^{p+1}$.

Assumption A1 generalizes the α-mixing condition of short-range dependence of Rosenblatt (1956) (see also Rosenblatt (1991) for an excelent discussion on mixing conditions). Notice that Assumption A1 allows nonstationary regressors. In particular, this assumption is needed when we have lags of the dependent variable as regressors. The rest of the assumptions are usual for the asymptotic theory of kernel regression estimates under strong mixing; see Robinson (1983). Condition A2 can be relaxed, allowing conditional heteroskedasticity and autocorrelation but paying the price of assuming higher marginal moments for ε_t and X_t, being the only difference the asymptotic variance of the estimator. In the last Section, we will further discuss this point. Condition A3, on smoothness of the regression and pdf of the regressors, is usual in kernel estimation. Condition A4 assumes that the regression function changes only in mean but the slopes are not altered. This assumption can be relaxed, but then we have to assume a different convergence rate for the bandwidth number in B4 below.

Define $k_{+(\ell)}(u)= d^{\ell}\, k_+(u)/d\, u^{\ell}$, $k_{-(\ell)}(u)= d^{\ell}\, k_-(u)/d\, u^{\ell}$, $K^{(k)}_{+(\ell)}(u)= k_{+(\ell)}(u_k)\, \Pi^{p+1}_{r\neq k}\, k(u_r)$ and $K^{(k)}_{-(\ell)}(u)= k_{-(\ell)}(u_k)\, \Pi^{p+1}_{r\neq k}\, k(u_k)$. We make the following assumptions on the kernel function and the smoothing number $a= a(n)$, when the jump or break is in the regression function $m(.)$. Required conditions will be given latter on, when the jump occurs in the first derivative of the regression function.

B0. The kernel function $k(.)$ is such that $\int_{\mathbb{R}} k(u)\, d\, u= 1$ and $\int_{\mathbb{R}} |uk(u)|\, d\, u < \infty$.

B1. The kernels $k_+(.)$ and $k_-(.)$ are such that $k_+(0)= k_-(0)= 0$, and $k_-(-\infty)= k_+(\infty)= 0$, and are continuously differentiable, such that $k_{\pm(1)}(0)\neq 0$.

B2. $k_{+(i)}(u)= (-1)^i\, k_{-(i)}(-u)$, for i= 0, 1.

B3. $\int u^r\, k_{\pm}(u)\, d\, u= \delta_{ro}$, r= 0, 1, where δ_{ro} is the Kronecker's delta and $\int u^2\, k_{\pm}(u)\, d\, u < \infty$, $\int |k_{\pm(i)}(u)|du < \infty$, i=1,2.

B4. The sequence $a= a(n)$ satisfies $\lim_{n\to\infty} (n\, a^{p+1})^{-1}= 0$ and $\lim_{n\to\infty} n\, a^{p+5} < \infty$.

Conditions on the one-sided kernels are similar to those imposed by Müller (1992). They rule out truncated kernels like truncated normal, Epanechnikov or uniform. However from any kernel function we can easily construct kernels with these properties. Let $k(.)$ be a density function with domain in $\mathbb{R}_+$, then

$$k_+(u) = u\,(c_1 + c_2\, u)\, k(u)$$

where c_1 and c_2 are chosen such that

$$\int_0^{\infty} u\ (c_1 + c_2\ u)\ k(u)\ du = 1 \quad \text{and} \quad \int_0^{\infty} u^2\ (c_1 + c_2\ u)\ k(u)\ du = 0,$$

will satisfy B1 and B3, being $k_-(-u)= k_+(u)$. For instance, if $k(u)= \exp(-u)$ $1(u\geq 0)$, then $k_+(u)= u\ (3 - u)\ \exp(-u)\ 1(u\geq 0)$. Condition B4 is satisfied by the bandwidth choice $a = C\ n^{-1/\upsilon}$ for $p+1< \upsilon\leq p+5$, where C is a constant independent of n.

Remark 3.1: If we do not assume A4, we need to change B4 by

$$\lim_{n\to\infty} \{n\ a^{p+3} + (n\ a^{p+1})^{-1}\} = 0.$$

That is, the bandwidth sequence must converge to zero faster than the bandwidth sequence in B4. ■

Let us introduce the following notation,

$$\hat{\delta}(\upsilon)= \hat{\Delta}^{(k)}(z_0^{(k)}+ a\ \upsilon\ e_k\ (na^{p+1})^{-1/2}))\ \tilde{f}(z_0^{(k)}+ a\ \upsilon\ e_k\ (na^{p+1})^{-1/2})^{r}, \tag{3.1}$$

$$\xi_n(\upsilon)= na^{p+1}\ (\hat{\delta}(\upsilon)- \hat{\delta}(0)) \quad \text{and} \quad \upsilon_n^{*}= \arg\max_{\upsilon\in\mathbb{R}} |\xi_n(\upsilon)|.$$

Since, by construction,

$$(na^{p+1})^{1/2}\left((\hat{z}_{0k}^{(k)} - z_{0k})/a\right) = \upsilon_n^{*}, \tag{3.2}$$

a Central Limit Theorem (CLT) for the left hand side of (3.2) is obtained from a CLT for $\xi_n(\upsilon)$, which is provided by the following theorem. Let $\alpha_0^{(k)}$ and $\alpha_{(s)}^{(k)}$ be p-dimensional vectors with components equal to the first p elements of $z_0^{(k)}$ and $z_{(s)}^{(k)}$ respectively, i.e. $z_0^{(k)}=(\alpha_0^{(k)\prime}, z_{0p+1}^{(k)})'$ and $z_{(s)}^{(k)}=(\alpha_{(s)}^{(k)\prime}, z_{(s)p+1}^{(k)})'$.

Theorem 3.1. Assume that A1–A4 and B0–B4 hold, with $\upsilon\in[-M, M]$, M finite, and $\Delta^{(k)}(z_0^{(k)})> 0$. Then $\xi_n(\upsilon)$ converges weakly to $\xi(\upsilon)$ on $\mathcal{C}([-M, M])$, where $\xi(.)$ is a continuous Gaussian process with moment structure
$E[\xi(\upsilon)]= \Delta^{(k)}(z_0^{(k)})\upsilon^2 K_{-(1)}^{(k)}(0)\ f_X(\alpha_0^{(k)})^r/2$ and
$\mathrm{Cov}[\xi(\upsilon_1), \xi(\upsilon_2)]= 2\upsilon_1\upsilon_2\sigma^2\int K_{-(1)}^{(k)}(u)^2 du\ f_X(\alpha_0^{(k)})^{2r-1}$. If $\Delta^{(k)}(z_0^{(k)})< 0$, replace

$K^{(k)}_{-(1)}(.)$ by $K^{(k)}_{+(1)}(.)$ everywhere. ∎

The proof of this theorem and any other result are in the Appendix.

The above theorem implies that

$$\xi(\upsilon)= \Delta^{(k)}(\chi_0^{(k)})\upsilon^2 K^{(k)}_{-(1)}(0)\; f_X(x_0^{(k)})^r/2+ V\upsilon f_X(x_0^{(k)})^r,$$

where $V\sim N(0,\ 2\sigma^2\int K^{(k)}_{-(1)}(u)\ du/f_X(x_0^{(k)}))$. Moreover, using Whitt (1970) results, the above weak convergence can be extended to $\mathscr{C}(-\infty,\ \infty)$. Observing the limiting process $\xi(\upsilon)$, it is easy to notice that it has a maximum at $\upsilon^{*}= -\ V/(\Delta^{(k)}(\chi_0^{(k)})\ K^{(k)}_{-(1)}(0))$, and, since the supremum is a continuous function, by the continuous mapping theorem, $\upsilon^{*}_n \xrightarrow{d} \upsilon^{*}$. Hence, we obtain a CLT for the left hand side of (3.2) as a Corollary.

Corollary 3.1. Under the same conditions of Theorem 3.1.

$$(na^{p+1})^{1/2}\left((\hat{\chi}^{(k)}_{ok} - \chi^{(k)}_{ok})/a\right) \xrightarrow{d} N\left(0,\ \frac{2\,\sigma^2\int (K^{(k)}_{-(1)}(u))^2\, du}{f_X(x_0^{(k)})(\Delta^{(k)}(\chi_0^{(k)})\ K^{(k)}_{-(1)}(0))^2}\right). \;∎$$

Remark 3.2: In view of the asymptotic variance of the break point and A4, the weight function $W(\chi)\ =\ f_X(x)^r$, for any r >0, produce the point estimate with smaller asymptotic variance. Intuitively, using this weight function we are avoiding those points χ whose k-th component is $\chi^{(k)}_{ok}$, but the joint density function of the regressors is very small (i.e. we are avoiding those points where observations are not very likely). In particular, among all the points $\chi\in \mathbb{R}^{p+1}$ with k-th component $\chi^{(k)}_{ok}$, we are choosing these points maximazing the density $f_X(.)$∎

Remark 3.3: Under A4, $\Delta^{(k)}(\chi_0^{(k)})= \Delta$ a constant. However, this assumption can be relaxed by assuming the rate of convergence of the bandwidth sequence given in Remark 3.1, which is faster than the rate of convergence implied by B4, and in view of Corollary 3.1, the rate of convergence of the point estimate will be slower for p≥ 0. ∎

Remark 3.4: If we allow conditional heteroskedasticity (or/and autocorrelation) of ε_t, and we assume that higher marginal moments of $m(Z_t)$ and ε_t exist, it is straightforward to prove that Corollary 3.1. also follows but now σ^2 must be replaced by $E(\varepsilon_t|X_t=x_{ok}^{(k)})=\sigma^2(x_{ok}^{(k)})$ in the heteroskedastic case (see Lemma 6 in the Appendix for further details). In the last Section, we will discuss how to deal with such problems. ∎

When the regression curve is continuous, but the first derivative with respect to some regressor is not, i.e. model (2.6), we need the following assumptions.

A3'.- The pdf $f_{X_t}(x)$ of X_t has, at least, three continuous derivatives with respect to all its arguments. The regression function $m(.)$ has, at least, three continuous derivatives with respect to all its arguments, except at the break point $x_{(s)k}$.

A4'. $m(x)/\partial x_s = m^*_{(s)}(x) + \Delta\, 1(x_s \le x_{(s)k})$, where $m^*_{(s)}(x)$ is continuous in $\mathbb{R}^{p+1}$.

B1'.- The kernel functions $k_+(.)$ and $k_-(.)$ are such that $k_+^{(\ell)}(0) = k_-^{(\ell)}(0)= 0$, and $k_+^{(\ell)}(\infty)= k_+^{(\ell)}(-\infty)= 0$, $\ell= 0, 1,2$, and $k_+^{(3)}(0)\neq 0$ and $k_-^{(3)}(0)\neq 0$.

B2'.- $k_+^{(1)}(u) = (-1)^i\, k_-^{(1)}(-u)$, for i= 0, 1, 2, 3.

B4'.- The sequence $a= a(n)$ satisfies $\lim_{n\to\infty} (na^{p+3})^{-1} = 0$ and $\lim_{n\to\infty} na^{p+7} < \infty$.

Remark 3.5: Assumption A4' implies that second derivatives of the regression function are constant with respect to all its arguments. We can avoid this condition by assuming that

$$\lim_{n\to\infty} \{n\, a^{p+5} + (n\, a^{p+3})^{-1}\} = 0. \blacksquare$$

As in the previous case, we can construct kernels with the above properties very easily. For instance, let $k(.)$ be a density function in $\mathbb{R}_+$. Then

$$k_+(u) = u^3\,(c_1 + c_2\, u)\, k(u),$$

where c_1 and c_2 are chosen such that

$$\int_0^\infty u^3(c_1+c_2u)k(u)\, du= 1 \quad \text{and} \quad \int_0^\infty u^4(c_1+c_2u)k(u)\, du = 0.$$

For instance, if $k(u)= \exp(-u)1(u\geq 0)$, then

$$k_+(u)= u^3(3-u)\exp(-u)/6 \ 1(u\geq 0) \tag{3.3}$$

will satisfy B1' and B3.

Now we need the following notation,

$\hat{\delta}'(\upsilon)= \hat{\Delta}^{(k)}_{(s)}(\gamma^{(k)}_{(s)}) + a\upsilon\, e_k\, (na^{p+3})^{-1/6})\, \tilde{f}(\gamma^{(k)}_{(s)}) + a\upsilon\, e_k\, (na^{p+3})^{-1/6})^r$,

$\xi'_n(\upsilon)= (na^{p+3})^{2/3}\, (\hat{\delta}'(\upsilon)- \hat{\delta}'(0))$ and $\upsilon'^*_n = \underset{\upsilon\, \in\, [-M,M]}{\arg\max} |\xi'_n(\upsilon)|$.

Theorem 3.2. Assume that A1-A2, A3' and B1',B2', B3, B4', with $\upsilon \in [-M, M]$, M finite, hold, and $\Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)})> 0$. Then $\xi'_n(\upsilon)$ converges weakly to $\xi'(\upsilon)$ on $\mathcal{C}([-M, M])$, where $\xi'(.)$ is a continuous Gaussian process such that

$E[\xi'(\upsilon)]= \Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)})\, \upsilon^4\, K^{(k)}_{-(3)}(0)\, f_X(x^{(k)}_0)^r/24$ and

$\mathrm{Cov}[\xi'(\upsilon_1),\ \xi'(\upsilon_2)] = 2\upsilon_1\upsilon_2\, \sigma^2\!\int (K^{(k)}_{-(2)}(u))^2 du\ f_X(x^{(k)}_0)^{2r-1}$.

If $\Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)}) < 0$, replace $K^{(k)}_{-(3)}(.)$ by $K^{(k)}_{+(3)}(.)$ everywhere.■

Employing same arguments as before, $\upsilon'^*_n= (na^{p+3})^{1/6}\, ((\hat{\gamma}^{(k)}_{(s)k} - \gamma_{(s)k})/a)$. Since, $\xi'_n(\upsilon) = \Delta^{(k)}(\gamma^{(k)}_0)\, \upsilon^4\, K^{(k)}_{-(3)}(0)\, f_X(x^{(k)}_0)^r/24 + \upsilon\, V'$, where $V'\sim N(0,\ 2\sigma^2 \int (K^{(k)}_{-(2)}(u))^2 d\,u/f_X(x^{(k)}_0))$, has a minimum at $\upsilon'^*= - \left(6\ V'/(\Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)})\ K^{(k)}_{-(3)}(0))^{1/3}\right)$, by the continuous mapping theorem, $(\upsilon'^*_n)^3 \xrightarrow{d} (\upsilon'^*)^3$. This justifies the following Corollary.

Corollary 3.2. Under the same conditions as in Theorem 3.2.

$$(na^{p+3})^{1/2} \left((\hat{\gamma}^{(k)}_{(s)k} - \gamma_{(s)k})/a\right)^3 \xrightarrow{d} N\left(0,\ \frac{36\text{x}2\,\sigma^2 \int (K^{(k)}_{-(2)}(u))^2\, d\,u}{f_X(x^{(k)}_{(s)})(\Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)})\ K^{(k)}_{-(3)}(0))^2}\right).■$$

From Corollaries 3.1. and 3.2., it is immediate to obtain the limiting distribution of the feasible estimate defined in (2.13).

Remark 3.6: Remarks 3.2., 3.3. and 3.4. after Corollary 3.1. applies also here. ■

Corollary 3.3. (a) Under the same conditions of Theorem 3.1., $\hat{\gamma}^{(k)}_{ok} - \tilde{\gamma}^{(k)}_{ok} = o_p\left((n\, a^{p+1})^{-1/2}\right)$. (b) Under the same conditions of Theorem 3.2., $\hat{\gamma}^{(k)}_{(s)k} - \tilde{\gamma}^{(k)}_{(s)k} = o_p\left((n\, a^{p+3})^{-1/2}\right)$.

This Corollary forms a basis to construct confidence intervals on the location of the break point. The variance σ^2 is estimated by

$$\tilde{\sigma}^2 = n^{-1} \sum_{t=1}^{n} \left\{ Y_t - \tilde{m}_-(Z_t)\, 1(Z_{tk} < \tilde{\gamma}^{(k)}_{ok}) - \tilde{m}_+(Z_t)\, 1(Z_{tk} \geq \tilde{\gamma}^{(k)}_{ok}) \right\}^2 ,$$

where

$$\tilde{m}_+(\gamma) = \frac{\sum_{t=1}^{n} Y_t K[(Z_t - \gamma)/a] 1(Z_{tk} \geq \tilde{\gamma}^{(k)}_{ok})}{\sum_{t=1}^{n} K[(Z_t - \gamma)/a]\, 1(Z_{tk} \geq \tilde{\gamma}^{(k)}_{ok})}$$

$$\tilde{m}_-(\gamma) = \frac{\sum_{t=1}^{n} Y_t K[(Z_t - \gamma)/a] 1(Z_{tk} < \tilde{\gamma}^{(k)}_{ok})}{\sum_{t=1}^{T} K[(Z_t - \gamma)/a]\, 1(Z_{tk} < \tilde{\gamma}^{(k)}_{ok})},$$

and $K(.)$ is not necessarily a one sided kernel.

Next Corollary justifies interval estimates for $\gamma^{(k)}_0$ and $\Delta^{(k)}(\gamma^{(k)}_0)$. Let $\tilde{\alpha}^{(k)}_0$ and $\tilde{\alpha}^{(k)}_{(s)}$ be the first p components of $\tilde{\gamma}^{(k)}_0$ and $\tilde{\gamma}^{(k)}_{(s)}$ respectively.

Corollary 3.4. (a) Under the same conditions as in Theorem 3.1.

$$\left\{ \frac{2\, \tilde{\sigma}^2 \int (K^{(k)}_{-(1)}(u))^2 d\, u}{\tilde{f}_X(\tilde{\alpha}^{(k)}_0)(\tilde{\Delta}^{(k)}(\tilde{\gamma}^{(k)}_0)\; K^{(k)}_{-(1)}(0))^2} \right\}^{-1/2} (na^{p+1})^{1/2} \left((\tilde{\gamma}^{(k)}_{ok} - \gamma_{ok})/a \right) \xrightarrow{d} N(0,\, 1),$$

$$\left\{ \frac{2\, \tilde{\sigma}^2 \int (K^{(k)}_{-}(u))^2 d\, u}{\tilde{f}_X(\tilde{\alpha}^{(k)}_0)} \right\}^{-1/2} (na^{p+1})^{1/2}\, (\hat{\Delta}^{(k)}(\tilde{\gamma}^{(k)}) - \Delta^{(k)}(\gamma^{(k)}_0)) \xrightarrow{d} N(0,1).$$

(b) Under the same conditions as in Theorem 3.2.

$$\left\{\frac{36\text{x}2\ \tilde{\sigma}^2 \int (K^{(k)}_{-(2)}(u))^2 du}{\tilde{f}_X(\tilde{\alpha}^{(k)}_{(s)})(\tilde{\Delta}^{(k)}_{(s)}(\tilde{\gamma}^{(k)}_{(s)})\ K^{(k)}_{-(3)}(0)))^2}\right\}^{-1/2} (na^{p+3})^{1/2}\left[(\tilde{\gamma}^{(k)}_{(s)k}-\gamma^{(k)}_{(s)k})/a\right]^3 \xrightarrow{d} N(0,1),$$

$$\left\{\frac{36\text{x}2\ \tilde{\sigma}^2 \int (K^{(k)}_{-(2)}(u))^2 du}{\tilde{f}_X(\tilde{\alpha}^{(k)}_0)}\right\}^{-1/2} (na^{p+3})^{1/2}\ (\hat{\Delta}^{(k)}_{(s)}(\tilde{\gamma}^{(k)}_{(s)})-\Delta^{(k)}_{(s)}(\gamma^{(k)}_{(s)})) \xrightarrow{d} N(0,\ 1).$$

∎

4. MONTE CARLO EXPERIMENTS

We will consider two models. The first one is the single threshold model

$$Y_t = X_t + \alpha\ 1(X_t \geq 0) + \varepsilon_t,\ t= 1,\ldots,n, \tag{4.1}$$

where X_t ~ iid $N(0,\ 1)$ and ε_t ~ iid $N(0,\ 1)$, X_t and ε_t are independent, for sample sizes n= 50, 100 and 200. The second one is a trend model, where

$$Y_t = \tau_t + \alpha\ 1(\tau_t \geq 0.5) + \varepsilon_t,\ t= 1,\ldots,n, \tag{4.2}$$

where ε_t ~ iid $N(0,1)$, for sample sizes n= 50, 100 and 200.

Monte Carlo biases and root mean squared errors (RMSE) based on 5,000 replications for models (4.1) and (4.2) are reported in Tables I and II respectively, for α= 0.5, 1 and 2. In all experiments we choose $a= Cn^{-1/5}$ for C= 1, 2, 3 and 4. The simulations were run in FORTRAN 77 Double Precision on an Apollo Workstation HP 715, using the IMS library for generating the random numbers.

In both models, the estimator behaves well and does not seem to be very sensitive to the bandwidth choice. As expected, the biases increase with C, but the variances decreases as C increases. Even for the smallest sample size, the estimator biases and RMSEs are quite small, decreasing rapidly as the sample size increases.

As it should be expected, results for the trend model (4.2) are better than for model (4.1) where a stochastic regressor determines the break.

TABLE 1 AND 2 ABOUT HERE

5. AN EMPIRICAL APPLICATION

We have applied the estimator to some USA macroeconomic time series used in the influential paper by Nelson and Plosser (1982) which have been used afterwards for illustration of other statistical procedures related to structural break point analysis (see, for instance Zivot and Andrews (1992) or Perron (1989)). The series considered are: annual nominal wages (1900-1979), annual common stock prices (1871-1970) and quarterly real gross national product (GNP) (1947:I-1986:III). All these series are in logs and we also consider the rate of growth of GNP (i.e. the first differences of the series in logs).

We consider a model where the regressor explaining the break is "time" (τ_t), and the other stochastic regressors are lags of the dependent variable. That is

$$m(Z_t) = \Delta\ 1(\tau_t \leq \tau_0) + m^{*}(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}), \tag{5.1}$$

with p= 1, 2 and 3, where $m^{*}(.)$ is an unknown function.

In Table 3 to 7, we report point estimates and 95% confidence intervals of the location of the break point in the different series. Figures 1 to 4 report plots of $\hat{\Delta}(Z_t)\tilde{f}_X(X_t)$, t=1,..,n, where $\tilde{f}_X(X_t)$ is computed according to (2.15).

Results for Nominal Wages are in Table 3 and Figure 1. The point estimate is around the 1929 crash and the interval estimates contain around the 6% of the sample. The year 1929 has been suggested as possible break point in Nominal Wages by several authors (e.g. Perron 1989). Figure 1 shows a clear peak of the scaled jump function in 1929. This function does not change a lot with respect to the bandwidth choice. For the choice $C= 1$, the function is too erratic , suggesting that this choice may be too small. As the number of lags increases, the confidence intervals become narrower. Probably, the errors are correlated and as we introduce more lags, the autocorrelation decreases. The autocorrelation is expected to be positive and, therefore, the confidence interval will become smaller as lags values of Y_t are introduced into the model.

TABLE 3 AND FIGURE 1 ABOUT HERE

The point estimates for the common stock prices are around the First World War (1914-1917). Other authors (e.g. Perron 1989) have suggested the great crash of 1929 as the possible break point. Looking at the original series, there is a clear peak in 1929, which could be interpreted as an additive outlier. Figure 2 shows that the jump function suddenly decreases to zero at 1929. Our method is not robust to outlying observations and such outliers can spoil our estimates. Table 4 shows that the confidence intervals are too wide, and they become wider as the sample size increases. It may be an indication of the bad performance of our estimator when outliers are present in the sample. It is expected that robust versions proposed in Section 6.3. will work better in these situations.

TABLE 4 AND FIGURE 2 ABOUT HERE

Table 5 and Figure 3 report results for the Quarterly Real GNP series. The point estimate is around 1967 (the Vietnam War). Authors differ about the location of the break point for this series. While Perron (1989) suggests the oil crisis of 1973 as a possible location and no unit roots, Zitov and Andrews (1992) defend that there is a unit root, and the possible structural break, if it exists, occurs at the second quarter of 1972. The confidence intervals are quite narrow containing typically less that the 6% of the sample. Table 6 and Figure 4 report results for the Growth Rate of Quarterly GNP. The break point estimate is still around 1967 but the confidence intervals are much wider.

There are several arguments for explaining this fact. On one hand, after taking first differences the series becomes much more volatile, the residual variance increases, and then the variance of the point estimates. There is also observed negative correlation in the series and, since this autocorrelation is not taken into account, the confidence intervals become wider. It may be also the case that model (5.1) is not valid here and the derivatives of $m(.)$ change before and after the break. Then, after taking differences of the original series, the jump in the transformed series could be smaller. That is, the model for the GNP series may be

$$m(Z_t) = m_-(\tau_t, Y_{t-1}, Y_{t-2},\dots,Y_{t-p})\, 1(\tau_t \le \tau_0) + m_+(\tau_t, Y_{t-1}, Y_{t-2},\dots,Y_{t-p})\, 1(\tau_t > \tau_0),$$

instead of (5.1). So, the jump size can be smaller after taking differences, and the variance of the break point estimate increases (note that the size of the jump is in the denominator of the asymptotic variance). Moreover, our results may indicate together with Zitov and Andrews (1989) results, that the series may have a unit root but the rate of growth have change around 1967.

TABLES 4 AND 5 AND FIGURES 3 AND 4 ABOUT HERE

6. SUGGESTIONS FOR FURTHER WORK

We indicate some extensions and applications of the above procedure that are of possible econometric interest.

6.1. Estimating the structural break under error autocorrelation.

For the sake of simplicity, we have assumed in the main theorems that the errors are uncorrelated. This assumption seems to be too strong in a time series context but it can be easily removed. Let us define $\gamma_j = E[\varepsilon_i \varepsilon_{j+i}]$, $j=1,2,\dots$ as the autocovariance and $f_\varepsilon(0) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j$ as the spectral density function evaluated at zero. Then, it is reasonably straightforward to prove that, under suitable conditions,

$$(n\ a^{p+1})^{1/2}\left((\hat{\gamma}_{0k}^{(k)} - \gamma_{0k}^{(k)})/a\right) \xrightarrow{d} N\left(0,\ \frac{2\pi f_{\varepsilon}(0)\ 2\ \sigma^2 \int (K_{-(1)}^{(k)}(u))^2\, d\, u}{f_X(x_0^{(k)})(\Delta^{(k)}(\gamma_0^{(k)})\ K_{-(1)}^{(k)}(0)))^2}\right).$$

Moreover, a consistent estimator of $f_{\varepsilon}(0)$ is given by

$$\tilde{f}_{\varepsilon}(0) = (2\pi)^{-1} \sum_{j=-m}^{m} w(j/m)\ \tilde{\gamma}_j,$$

where $w(.)$ is a weighting function with $w(0)$=1 and $w(x)$=0 when $|x|>1$, and $\tilde{\gamma}_j = \frac{1}{n-j} \sum_{k=1}^{n-j} \tilde{\varepsilon}_k \tilde{\varepsilon}_{k+j}$ estimates γ_j, where

$$\tilde{\varepsilon}_t = Y_t - \tilde{m}_-(Z_t)\ 1(Z_{tk} < \tilde{\gamma}_{0k}^{(k)}) - \tilde{m}_+(Z_t)\ 1(Z_{tk} \geq \tilde{\gamma}_{0k}^{(k)}). \tag{6.1}$$

6.2. Estimating the structural break point under heteroskedasticity of unnknown form.

In cross-sectional applications, it is restrictive to assume conditional homoskedasticity. If $E(\varepsilon_t^2 | X_t = x) = \sigma^2(x)$, under suitable regularity conditions, it is easy to prove that

$$(n\ a^{p+1})^{1/2}\left((\hat{\gamma}_{0k}^{(k)} - \gamma_{0k}^{(k)})/a\right) \xrightarrow{d} N\left(0,\ \frac{2\ \sigma^2(x_0^{(k)}) \int (K_{-(1)}^{(k)}(u))^2\, d\, u}{f_X(x_0^{(k)})(\Delta^{(k)}(\gamma_0^{(k)})\ K_{-(1)}^{(k)}(0)))^2}\right).$$

The conditional variances can be estimated from (6.1) by

$$\hat{\sigma}^2(x) = \sum_{j=1}^{n} \tilde{\varepsilon}_j^2\ W_j(x),$$

where $W_j(x)$ are nonparametric probabilistic weights.

6.3. Finding the regressor explaining the break.

It may be the case that the break is explained by more than one regressor. We have in mind the model

$$m(Z_t) = m_-(Z_t)\, 1\left(g(Z_t) \le g_0\right) + m_+(Z_t)\, 1\left(g(Z_t) > g_0\right), \quad i,j = 1,2,$$

where $g(.)$ is an unknown function. It is interesting to estimate the functional form of $g(.)$. However, we are unable to identify this function using the methodology presented above. Suppose, for the sake of simplicity, that $g(.)$ depends only on the first two regressors. We could estimate

$$m_{ij}(Z_t) = \lim_{\substack{\delta_1 \to 0 \\ \delta_2 \to 0}} m\left(\gamma + (-1)^i \delta_1 + (-1)^j \delta_2\right)$$

and

$$\Delta_{ijkl}(\gamma) = \left| m_{ij}(\gamma) - m_{kl}(\gamma)\right|, \quad i \ne k, \ j \ne l.$$

Applying the method developed in Section 2, we could estimate the value maximizing the above function, but it is not very useful.

We may be interested in finding the regressor explaining the break. We could conclude that the regressor k^* explains the break, where

$$k^* = \arg\max_{k \in \{1,\ldots,p\}} \left|\hat{\Delta}^{(k)}(\hat{\gamma}_0^{(k)})\right|.$$

6.4. Two step estimation of parametric models

Suppose that the regression function is parameterized, but there is a structural break with unknown location. Suppose that the regression function before and after the break is linear. That is,

$$E(Y_t|X_t) = Z_t'\beta_1\, 1(Z_{kt} \le \gamma_0) + Z_t'\beta_2\, 1(Z_{kt} > \gamma_0).$$

The break can affect to the intercept ($\beta_{11} \ne \beta_{21}$ and $\beta_{1j} = \beta_{2j}$, all $j > 1$) or the slopes ($\beta_{11} = \beta_{21}$ and $\beta_{1j} \ne \beta_{2j}$, all $j > 1$). Feder (1975) proposed an estimation method for a segmented (but continuous) regression model with a fixed regressor and iid errors. Bai (1992) also consider the same model but with autocorrelated errors where the break occurs in the trend. A two step procedure permits to estimate β_1 and β_2 even when the regression is not continuous. In a first step γ_0 is estimated by $\hat{\gamma}_0$ using the method developed in Section 2. Then β_1 and β_2 are estimated by

$$\hat{\beta}_1 = \left[\sum_t Z_t Z_t' \, 1(Z_{kt} \le \hat{\gamma}_0)\right]^{-1} \sum_t Y_t \, Z_t \, 1(Z_{kt} \le \hat{\gamma}_0),$$

and

$$\hat{\beta}_2 = \left[\sum_t Z_t Z_t' \, 1(Z_{kt} > \hat{\gamma}_0)\right]^{-1} \sum_t Y_t \, Z_t \, 1(Z_{kt} > \hat{\gamma}_0),$$

respectively. It is expected that these estimates will be, under suitable regularity conditions, asymptotically equivalent to estimates computed when γ_0 is perfectly known.

6.5. Robustness

Outliers can produce very misleading consequences in our estimation method, as it has been illustrated in Section 5. One sided kernels are more sensitive to outliers than symmetric kernels because more weight is put on the outlying observations. The influence of outliers can be bounded by using a M-type version of the usual kernel method as proposed by Tsybakov (1993), Robinson (1984) and Härdle (1984), to mention only a few. Instead of estimating $m(x)$ as the functional of conditional location, we can estimate a robust conditional location functional $r(x)$ defined as the solution to

$$E\left[\phi\left(Y - r(x)\right) \big| X = x\right] = 0,$$

where $\phi(.)$ is a bounded real function. So, we can estimate

$$r_+(\gamma) = \lim_{\delta \to 0_+} r(\gamma + e_k\delta) \quad \text{and} \quad r_-(\gamma) = \lim_{\delta \to 0_-} r(\gamma + e_k\delta),$$

by $\hat{r}_+(\gamma)$ and $\hat{r}_-(\gamma)$, defined as the solution to

$$\frac{\sum_{t=1}^n \phi(Y_t - \hat{r}_+(\gamma))\, K_+[(Z_t - \gamma)/a]}{\sum_{t=1}^n K_+[(Z_t - \gamma)/a]} = 0 \quad \text{and} \quad \frac{\sum_{t=1}^n \phi(Y_t - \hat{r}_-(\gamma))\, K_-[(Z_t - \gamma)/a]}{\sum_{t=1}^n K_-[(Z_t - \gamma)/a]} = 0,$$

respectively. Then, we estimate the jump as

$$\hat{\Delta}^{(k)}(\gamma) = \left|\hat{r}_+(\gamma) - \hat{r}_-(\gamma)\right| \, \tilde{f}_X(\gamma)^r \;, \; r > 0.$$

Remark: The proofs below are for the case where regressors are stationary for notational simplicity. In the general case, notice that instead of $\int h(x) f_X(x)\, dx$ we will have $\lim_{n\to\infty} n^{-1}\sum_t \int h(x) f_{X_t}(x)\, dx$.

Proof of Theorem 3.1.

Let us introduce some notation. Hereforth, $\tau = \gamma_0^{(k)}$, $g_n = (na^{p+1})^{1/2}$, $e = e_k$, $\Delta(\upsilon) = \Delta^{(k)}(\upsilon)$, $f(\upsilon) = f_X(\upsilon)$, $K_+(\upsilon) = K_+^{(k)}(\upsilon)$, $K_+^s(\upsilon) = K_{+(s)}^{(k)}(\upsilon)$, $K_-(\upsilon) = K_-^{(k)}(\upsilon)$, $K_-^s(\upsilon) = K_{-(s)}^{(k)}(\upsilon)$, $\gamma = \int_0^\infty K_+^1(\omega)^2\, d\,\omega$, $\mathcal{K}_t^-(\upsilon) = K_-((X_t - (\tau + ave/g_n))/a)$, $\mathcal{K}_t^+(\upsilon) = K_+((X_t - (\tau + ave/g_n))/a)$, $\hat{P}_+(\upsilon) = g_n^{-2}\sum_t Y_t\, \mathcal{K}_t^+(\upsilon)$, $\hat{P}_-(\upsilon) = g_n^{-2}\sum_t Y_t\, \mathcal{K}_t^-(\upsilon)$, $\hat{f}_+(\upsilon) = g_n^{-2}\sum_t \mathcal{K}_t^+(\upsilon)$, $\hat{f}_-(\upsilon) = g_n^{-2}\sum_t \mathcal{K}_t^-(\upsilon)$, $\mathcal{P}_+(\upsilon) = \hat{P}_+(\upsilon) - \hat{P}_+(0)$, $\mathcal{P}_-(\upsilon) = \hat{P}_-(\upsilon) - \hat{P}_-(0)$, $\ell_+(\upsilon) = \hat{f}_+(\upsilon) - \hat{f}_+(0)$, $\ell_-(\upsilon) = \hat{f}_-(\upsilon) - \hat{f}_-(0)$.

It is easily seen that,

$$\hat{\delta}(\upsilon) - \hat{\delta}(0) = \sum_{i=1}^4 A_i(\upsilon) + a(\upsilon),$$

where

$A_1 = [\mathcal{P}_+(\upsilon) - \mathcal{P}_-(\upsilon)][\hat{f}_-(\upsilon) - f(\tau)]$, $A_2(\upsilon) = \mathcal{P}_-(\upsilon)[\hat{f}_-(\upsilon) - \hat{f}_+(\upsilon)]$,
$A_3(\upsilon) = -\,\ell_+(\upsilon)\,[\hat{P}_+(0) - m_+(\tau)f(\tau)]$, $A_4(\upsilon) = -\,\ell_-(\upsilon)\,[\hat{P}_-(0) - m_-(\tau)f(\tau)]$,
$a(\upsilon) = [\mathcal{P}_+(\upsilon) - \mathcal{P}_-(\upsilon)]f(\tau) - \ell_+(\upsilon)m_+(\tau)f(\tau) + m_-(\tau)\ell_-(\upsilon)f(\tau)$.

Thus, the theorem follows by Propositions 1 to 5, which are based on results proved in the Lemmata.

PROPOSITION 1: $A_1(\upsilon) = O_p(g_n^{-3})$.
Proof.- By lemmas 5 and 7, $\mathcal{P}_+(\upsilon) - \mathcal{P}_-(\upsilon) = O_p(g_n^{-2})$, while by Lemma 2 $[\hat{f}_-(\upsilon) - f(\tau)] = O_p(g_n^{-1} + a^2)$, noting that, by B4, $O_p(a^2 g_n^{-2}) = O_p(g_n^{-3})$.

PROPOSITION 2: $A_2(\upsilon) = O_p(ag_n^{-2})$.
Proof.- Notice that

$$\begin{aligned} A_2(\upsilon) &= [\mathcal{P}_-(\upsilon) - \mathcal{P}_+(\upsilon)][\ell_-(\upsilon) - \ell_+(\upsilon)] + [\mathcal{P}_-(\upsilon) - \mathcal{P}_+(\upsilon)][\hat{f}_-(0) - \hat{f}_+(0)] \\ &\quad + \mathcal{P}_+(\upsilon)[\ell_-(\upsilon) - \ell_+(\upsilon)] + \mathcal{P}_+(\upsilon)[\hat{f}_-(0) - \hat{f}_+(0)] \\ &= O_p(ag_n^{-2}), \end{aligned}$$

by lemmas 2, 4, 5 and 7, noting that, by B4, $O_p(a^3 g_n^{-1}) = O_p(a g_n^{-2})$.

PROPOSITION 3: $A_3(\upsilon) = O_p(a g_n^{-2})$

Proof.- $\ell_+(\upsilon) = O_p(a g_n^{-1})$ by lemmas 6 and 9 and $[\hat{\mathcal{P}}_+(0) - m_+(\tau) f(\tau)] = O_p(g_n^{-1} + a^2)$ by Lemma 1.

PROPOSITION 4: $A_4(\upsilon) = O_p(a g_n^{-2})$

Proof.- Applying Lemma 1, 6 and 9 as in Proposition 3.

PROPOSITION 5: $g_n^2 a(\upsilon) \Rightarrow \xi(\upsilon)$

Proof.- By lemmas 4 to 12, $\mathrm{E}[g_n^2\, a(\upsilon)] = \Delta(\tau)\, f(\tau)^2\, K_-^1(0)\, \upsilon^2/2 + o(1)$, and $\mathrm{Cov}(g_n^2\, a(\upsilon_1),\, g_n^2\, a(\upsilon_2)) = 2\,\sigma^2\, \upsilon_1 \upsilon_2 \gamma f(\tau)^3 + o(1)$. Define $\bar{a}(\upsilon_k) = g_n^2\, \{a(\upsilon_k) - \mathrm{E}(a(\upsilon_k))\}$. Then, for fixed $\upsilon_1, \upsilon_2, \ldots, \upsilon_\ell \in [-M, M]$,

$$\{\bar{a}(\upsilon_1), \bar{a}(\upsilon_2), \ldots, \bar{a}(\upsilon_\ell)\} \xrightarrow{d} N(0, \Sigma),$$

by Robinson (1983) Lemma 7.1, where $\Sigma = (\sigma_{ij})$, and $\sigma_{ij} = 2\upsilon_i \upsilon_j\, f(\tau)^3\, m_+(\tau)^2\, \gamma$. Thus, the proposition follows by showing that the sequence $\bar{a}(\upsilon)$ is tight. According to Billingsley (1968), it suffices to prove that

$$\mathrm{E}\{(\bar{a}(\upsilon_1) - \bar{a}(\upsilon_2))^2\} \le C(\upsilon_1 - \upsilon_2)^2,$$

where C is a generic constant, what follows by Lemmas 13 and 14. But, by Whitt (1970), we can extend the result to $(-\infty, \infty)$. Therefore,

$$g_n^2 \left(\hat{\delta}(\upsilon) - \hat{\delta}(0)\right) \xrightarrow[\text{weakly}]{d} \Delta(\tau)\, K_-^1(0)\, f(\tau)^2\, \upsilon^2/2 + \upsilon\, Z$$
$$\equiv -\,\Delta(\tau)\, K_+^1(0)\, f(\tau)^2\, \upsilon^2/2 + \upsilon\, Z\, f(\tau)^2,$$

where $Z \sim N\left(0,\ 2\,\sigma^2\, \gamma/f(\tau)\right)$. ■.

Proof of Theorem 3.2.

We need to introduce some extra notation. Hereforth, $h_n = (n\, a^{p+3})^{1/6}$, $\Delta_{(s)}(\upsilon) = \Delta_{(s)}^{(k)}(\upsilon)$, $\mathcal{K}_t^{\pm}(\upsilon) = K_{\pm}((X_t - (\tau + a\upsilon e/h_n))/a)$, $\mathcal{K}_{t\pm}^1(\upsilon) = K_{\pm}^1((X_t - \tau + ea\upsilon/h_n)/a)$, $\hat{P}_{\pm}^1(\upsilon) = a h_n^{-6} \sum_t Y_t K_{t\pm}^1(\upsilon)$, $\hat{f}_{\pm}^1(\upsilon) = a h_n^{-6} \sum_t \mathcal{K}_{t\pm}^1(\upsilon)$, $\mathcal{P}_{\pm}^1(\upsilon) = \hat{P}_{\pm}^1(\upsilon) - \hat{P}_{\pm}^1(0)$, $\ell_{\pm}^1(\upsilon) = \hat{f}_{\pm}^1(\upsilon) - \hat{f}_{\pm}^1(0)$, where $\hat{P}_{\pm}(\upsilon) = (n a^{p+1})^{-1} \sum_t Y_t \mathcal{K}_t^{\pm}(\upsilon)$, $\hat{f}_{\pm}(\upsilon) = (n a^{p+1})^{-1} \sum_t \mathcal{K}_t^{\pm}(\upsilon)$.

It is easily seen that,

$$\hat{\delta}'(\upsilon) - \hat{\delta}'(0) = \sum_{i=1}^{8} B_i(\upsilon) + b(\upsilon),$$

where

$$B_1(\upsilon) = \hat{f}^1_-(0)\, [\mathcal{P}_+(\upsilon)-\mathcal{P}_-(\upsilon)],\ B_2(\upsilon) = \mathcal{P}_-(\upsilon)\, [\hat{f}^1_-(\upsilon)-\hat{f}^1_+(\upsilon)],$$
$$B_3(\upsilon) = \ell_-(\upsilon)\, [\hat{P}^1_+(0)-\hat{P}^1_-(0)],\ B_4(\upsilon) = -\hat{P}^1_-(0)\, [\ell_+(\upsilon)-\ell_-(\upsilon)],$$
$$B_5(\upsilon) = -\mathcal{P}^1_-(\upsilon)\, [\hat{f}_-(\upsilon)-\hat{f}_+(\upsilon)],\ B_6 = [\mathcal{P}^1_+(\upsilon)- \mathcal{P}^1_-(\upsilon)][\hat{f}_-(\upsilon)-\hat{f}_+(\upsilon)],$$
$$B_7(\upsilon)= \ell^1_-(\upsilon)[\hat{P}_+(0)- m(\tau)f(\tau)],\ B_8(\upsilon)= -\ell^1_+(\upsilon)[\hat{P}_-(0)- m(\tau)f(\tau)],$$
$$b(\upsilon)= \left([\mathcal{P}^1_+(\upsilon)- \mathcal{P}^1_-(\upsilon)]-m(\tau)[\ell^1_+(\upsilon)-\ell^1_-(\upsilon)]\right) f(\tau).$$

Then the Theorem follows from the following propositions.

PROPOSITION 1.6.: $B_1(\upsilon)= o_p(h_n^{-4})$.
Proof.- By consistency of $\hat{f}^1_-(0)$ to $f^1(\tau)$ and lemmas 5 and 7, $B_1(\upsilon)=O(g_n^{-2}+ a^2/g_n)$. But $g_n= (na^{p+1})^{-1/2}$. So by B4' $B_1(\upsilon)=o(h_n^{-4})$.

PROPOSITION 1.7.: $B_2(\upsilon)= o_p(h_n^{-4})$.
Proof.- By Lemma 4, $\mathcal{P}_-(\upsilon)= o(g_n^{-2}+ a^2/g_n)$, and by Lemma 18, $(\hat{f}^1_-(\upsilon)-\hat{f}^1_+(\upsilon))= O(a^2+ (na^{p+2})^{-1/2})$. So by B4' $B_1(\upsilon)=o(h_n^{-4})$.

PROPOSITION 1.8.: $B_3(\upsilon)= o_p(h_n^{-4})$.
Proof.- Use lemmas 6 and 9 for $\ell_-(\upsilon)$ and that $[\hat{P}^1_+(0)-\hat{P}^1_-(0)]= O_p(1)$ by similar arguments to Lemma 1. Then apply B4'.

PROPOSITION 1.9.: $B_4(\upsilon)= o_p(h_n^{-4})$.
Proof.- $\hat{P}^1_-(0)= O_p(1)$ by similar arguments to Lemma 1 and $[\ell_+(\upsilon)-\ell_-(\upsilon)]= o_p(h_n^{-4})$ by lemmas 6 and 9 and B4'.

PROPOSITION 1.10: $B_5(\upsilon)= o_p(h_n^{-4})$.
Proof.- Apply Lemma 19 for $\mathcal{P}^1_-(\upsilon)$, Lemma 2 for $[\hat{f}_-(\upsilon)-\hat{f}_+(\upsilon)]$ and then B4'.

PROPOSITION 1.11: $B_6(\upsilon)= o_p(h_n^{-4})$.
Proof.- Apply lemmas 20 and 21 for $[\mathcal{P}^1_+(\upsilon)- \mathcal{P}^1_-(\upsilon)]$, and $\hat{f}_\pm(0)= f(\tau) + O_p(g_n^{-1})$ by Lemma 2, then apply B4'.

PROPOSITION 1.12: $B_7(\upsilon)= o_p(h_n^{-4})$.
Proof.- Apply lemmas 20 and 21 for $\ell^1_-(\upsilon)$ and $[\hat{P}_+(0)- m(\tau)f(\tau)]= \bar{m}_+(\tau)f(\tau) + O_p(g_n^{-1})$ by Lemma 1.

PROPOSITION 1.13: $B_8(\upsilon) = o_p(h_n^{-4})$.
Proof.- Same as Proposition 1.12.

PROPOSITION 1.14: $h_n^2\, b(\upsilon) \Rightarrow \xi(\upsilon)$.
Proof: Using similar arguments as in Proposition 1.5.

Proof of Corollary 3.3.

We only prove (a) since (b) is identical. By definition of υ^* and $\tilde{\gamma}_{ok}^{(k)}$, we should show that

$$\lim_{n\longrightarrow\infty} \Pr\{|\upsilon^* - \tilde{\gamma}_{ok}^{(k)}| > \delta\, (n\, a^{p-1})^{-1/2}\} = 0.$$

Take υ^* fixed, the above probability is bounded by $[1- F(x+\upsilon^*)-F(-x+\upsilon^*)]^n$, where $x = \delta\, (n\, a^{p-1})^{-1/2}$ and $F(.)$ is the distribution function of $\tilde{\gamma}_{ok}^{(k)}$. Then, the probability is of an order of magnitude

$$\lim_{n\longrightarrow\infty} (1-2x)^n = \lim_{n\longrightarrow\infty} \exp\{-2(n\, a^{1-p})^{1/2}\}.\blacksquare$$

LEMMATA

This Lemma is used for the proof of Theorem 3.1.

Notation: From now on, $\int_a^b \ell(\omega)\, d\omega \equiv \int_a^b \left[\int_{\mathbb{R}^p} \ell(\omega)\, \Pi_{j\neq k}\, d\omega_j\right] d\omega_k$.

Define $A_{\pm}(x,\, \upsilon) = K_{\pm}\left((x - (\tau - a\upsilon/g_n))/a\right) - K_{\pm}\left((x-\tau)/a\right)$, $\bar{m}(x) = m(x)f(x)$,

$\bar{m}_{\pm}(x) = m_{\pm}(x)f(x)$, $\bar{m}^{s}_{\pm}(x) = \partial^{s}\bar{m}_{\pm}(x)/\partial x^{s}$, $S^{\pm}_{t}(\upsilon) = [\mathcal{K}^{\pm}_{t}(\upsilon) - \mathcal{K}^{\pm}_{t}(0)]$.

Lemma 1.- $\hat{P}_{\pm}(0) = \bar{m}_{\pm}(\tau)f(\tau) + O_p(g_n^{-1} + a^2)$.

Proof.- Applying Robinson's (1983) Theorem 5.1. □

Lemma 2.- $\hat{f}_{\pm}(0) = f(\tau) + O_p(g_n^{-1} + a^2)$.

Proof.- Applying Robinson's (1983) Theorem 4.1. □

Lemma 3.- $E\left\{ |A_{\pm}(X_t,\, \upsilon_1)|^{\alpha} \right\} = O\left(a^{p+1}/g_n^{\alpha}\right)$.

Proof: The l.h.s. of the above expression is bounded by

$$C \int_{\mathbb{R}^{p+1}} |A_{\pm}(\omega,\, \upsilon_1)|^{\alpha}\, d\omega,$$

where, hereforth, C is a generic constant, (in this case C bounds the density function $f_X(.)$). By a change of variable and taking into account that $K_-(\omega) = 0$, for all $\omega \geq 0$,

$$\int_{\mathbb{R}^{p+1}} |A_-(\omega,\, \upsilon)|^{\alpha}\, d\omega = a^{p+1}\int_{-\infty}^{-\upsilon/g_n} |K_-(\upsilon e/g_n + \omega) - K_-(\omega)|^{\alpha}\, d\,\omega$$

$$+\, a^{p+1}\int_{-\upsilon/g_n}^{0} |K_-(\omega)|^{\alpha}\, d\omega$$

$$= O\left(a^{p+1}/g_n^{1+\alpha} + a^{p+1}/g_n^{\alpha}\right),$$

by applying a mean value theorem (mvt) argument. The other cases are symmetrical.□

For the next two lemmas, we suppose that $\upsilon > 0$. The results for $\upsilon < 0$ are identical, and therefore ommited.

Lemma 4.- $E[\mathcal{P}_{\pm}(\upsilon)] = O(a\ g_n^{-1})$.

Proof: By a change of variable and a Taylor expansion w.r.t the k-th coordinate,

$$E[\mathcal{P}_{+}(\upsilon)] = \int_0^{\infty} \{\bar{m}(a\omega + ae\upsilon/g_n + \tau) - \bar{m}(a\omega + \tau)\}\ K_{+}(\omega)\ d\omega$$

$$= \bar{m}_{+}^{1}(\alpha)\ a\upsilon/g_n + O(a^2/g_n^2).$$

Then, apply condition B4. $E[\mathcal{P}_{-}(\upsilon)] = O(a\ g_n^{-1})$ using the same arguments. □

Lemma 5.- $E[\mathcal{P}_{+}(\upsilon) - \mathcal{P}_{-}(\upsilon)] = -\Delta(\tau)\ f(\tau)\ K_{-}^{1}(0)\upsilon^2/2g_n^2 + o(a^2/g_n)$.

Proof: In this Lemma, without loss of generality, we will assume that $\tau=0$. Moreover, we will define $\bar{m}(\alpha) = f(\alpha)m(\alpha)$ and $\nu = \upsilon/g_n$.

$$E[\mathcal{P}_{+}(\upsilon) - \mathcal{P}_{-}(\upsilon)] = (a^{p+1})^{-1}\int m(\alpha)\left\{K_{+}\Big((\alpha+ae\nu)/a\Big) - K_{+}(\alpha/a)\right\} f(\alpha)\ d\alpha$$

$$- (a^{p+1})^{-1}\int m(\alpha)\left\{K_{-}\Big((\alpha+ae\nu)/a\Big) - K_{-}(\alpha/a)\right\} f(\alpha)\ d\alpha$$

$$= \alpha(\nu) - \alpha(0).$$

where

$$\alpha(\nu) = (a^{p+1})^{-1}\int \bar{m}(\alpha)\left\{K_{+}\Big((\alpha+ae\nu)/a\Big) - K_{-}\Big((\alpha+ae\nu)/a\Big)\right\} d\alpha.$$

By the change of variable $u=(\alpha+ae\nu)/a$,

$$\alpha(\nu) = \Delta(0)f(0) + \int_{\nu}^{\infty}\left\{\bar{m}_{+}\Big(a(u-e\nu)\Big) - \bar{m}_{+}(0)\right\} K_{+}(u)\ du + \int_0^{\nu} \bar{m}_{-}\Big(a(u-e\nu)\Big)K_{+}(u)\ du$$

$$- \int_{-\infty}^{0}\left\{\bar{m}_{-}\Big(a(u-e\nu)\Big) - \bar{m}_{-}(0)\right\} K_{-}(u)\ du - \bar{m}_{+}(0)\int_0^{\nu} K_{+}(u)\ du.$$

After applying a Taylor expansion w.r.t. the k-th coordinate, and using the facts that, by A4, $\bar{m}_{-}^{1}(0) = \bar{m}_{+}^{1}(0)$ and $\int u\ K_{\pm}(u)\ du = 0$, we can write

$$\alpha(\nu) = \Delta(0)f(0) - \Delta(0)f(0)\int_0^{\nu} K_{+}(u)\ du + \int_{\nu}^{\infty}\bar{m}_{+}^{2}\Big(\xi a(u-e\nu)\Big)a^2(u-e\nu)^2K_{+}(u)\ du$$

$$+ \int_{-\infty}^{0} \bar{m}_{-}^{2}\Big(\xi a(u-e\nu)\Big)\ a^2(u-e\nu)^2 K_{-}(u)\ du + \int_{0}^{\nu} \bar{m}_{-}^{2}\Big(\xi a(u-e\nu)\Big)\ a^2(u-e\nu)^2 K_{+}(u)\ du,$$

where ξ, hereforth, is a generic constant such that $\xi\in(0,1)$. Since $a^2\int_0^{\nu}(u-e\nu)^2\ K_{+}(u)\ du = O(a^2\ \nu^3)$, we have

$$\begin{aligned}\alpha(\nu)= \bar{m}_{+}^{2}(0)\int_{0}^{\infty} a^2(u-\nu)^2K_{+}(u)du + \bar{m}_{-}^{2}(0)\int_{-\infty}^{0} a^2(u-\nu)^2K_{-}(u)du + \Delta(0)f(0)\\ -\Delta(0)f(0)K_{+}^{1}(0)\nu^2 + \int_{\nu}^{\infty}\Big\{\bar{m}_{+}^{2}\Big(\xi a(u-e\nu)\Big) - \bar{m}_{+}^{2}(0)\Big\}\ a^2(u-e\nu)^2K_{+}(u)\ du\\ + \int_{\nu}^{\infty}\Big\{\bar{m}_{-}^{2}\Big(\xi a(u-e\nu)\Big) - \bar{m}_{-}^{2}(0)\Big\}\ a^2(u-e\nu)^2K_{-}(u)\ du + + O(a^2\nu^3).\end{aligned}$$

Thus,

$$\alpha(\nu)-\alpha(0)= \bar{m}_{+}^{2}(0)\ a^2\nu^2/2 - \Delta(0)f(0)K_{+}^{1}(0)\nu^2/2 + T_1 + T_2 + O(a^2\nu^3),$$

where

$$T_1 = \int_{\nu}^{\infty}\Big\{\bar{m}_{+}^{2}\Big(\xi a(u-e\nu)\Big)-\bar{m}_{+}^{2}(0)\Big\}\ a^2(u-e\nu)^2K_{+}(u)\ du - \int_{0}^{\infty}\Big\{\bar{m}_{+}^{2}(\xi au)-\bar{m}_{+}^{2}(0)\Big\}\ a^2u^2K_{+}(u)\ du$$

$$T_2 = \int_{-\infty}^{0}\Big\{\bar{m}_{-}^{2}\Big(\xi a(u-e\nu)\Big)-\bar{m}_{-}^{2}(0)\Big\}\ a^2(u-e\nu)^2K_{-}(u)\ du - \int_{-\infty}^{0}\Big\{\bar{m}_{-}^{2}(\xi au)-\bar{m}_{-}^{2}(0)\Big\}\ a^2u^2K_{-}(u)\ du$$

Thus, it suffices to prove that $T_i= o(a^2\nu)$, i=1,2. After the change of variable $v= u-\nu$,

$$T_1 = a^2\int_0^{\infty}\Big\{\bar{m}_{+}^{2}(\xi av)-\bar{m}_{+}^{2}(0)\Big\}\ v^2\Big\{K_{+}(v+\nu)-K_{+}(v)\Big\}\ dv= T_{11} + T_{12}$$

where

$$T_{11} = a^2\int_0^{M}\Big\{\bar{m}_{+}^{2}(\xi av)-\bar{m}_{+}^{2}(0)\Big\}\ v^2\Big\{\nu\ K_{+}^{1}(v)+ \nu^2K_{+}^{2}(\xi v)\Big\}\ dv,$$

$$T_{12}= a^2\int_M^{\infty}\Big\{\bar{m}_{+}^{2}(\xi av)-\bar{m}_{+}^{2}(0)\Big\}\ v^2\Big\{\nu\ K_{+}^{1}(v)+ \nu^2K_{+}^{2}(\xi v)\Big\}\ dv.$$

Because $m(.)$ and $f(.)$ are continuous - and thus uniformly continuous in a compact interval-, for an arbitrary $\varepsilon > 0$,

$$T_{11}= \varepsilon\ a^2\int_0^{M}|\nu\ K_{+}^{1}(v)v^2|dv + \varepsilon\ a^2\ \nu^2\int_0^{M}|v^2\ K_{+}^{2}(\xi v)|\ dv = o(a^2\nu),$$

For M large enough, $\int_M^{\infty}\bar{m}_{+}^{2}(v)\ v^2\ K_{+}^{2}(v)dv < \infty$, then $T_{12}= o(a^2\nu)$. Using similar arguments, it is also proved that $T_2= O(a^2\nu)$. □

Lemma 6.- $E[\ell_{\pm}(\upsilon)] = O(a\ g_n^{-1})$.

Proof: Identical to the proofs of lemmas 4.□

Lemma 7.- (a) $\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \mathcal{P}_+(\upsilon_2)\} = g_n^{-4}\ f(\tau)\ \upsilon_1\upsilon_2\ [m_+^2(\tau) + \sigma^2]\ \gamma + o(g_n^{-4})$.

(b) $\mathrm{Cov}\{\mathcal{P}_-(\upsilon_1),\ \mathcal{P}_-(\upsilon_2)\} = g_n^{-4}\ f(\tau)\ \upsilon_1\upsilon_2\ [m_-^2(\tau) + \sigma^2]\ \gamma + o(g_n^{-4})$.

Proof: **We only prove (a), the other is identical. Without loss of generality, we only consider the case** $0 < \upsilon_1 < \upsilon_2$. **Because of independence of** ε_t, **the above** covariance is equal to

$$\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \mathcal{P}_+(\upsilon_2)\} = \sigma^2 n\ g_n^{-4}\ \mathrm{Cov}\ \{S_1^+(\upsilon_1),\ S_1^+(\upsilon_2)\} +$$

$$n\ g_n^{-4}\ \mathrm{Cov}\{m(X_1)S_1^+(\upsilon_1),\ m(X_1)S_1^+(\upsilon_2)\} + g_n^{-4} \sum_{t\neq s}\sum \mathrm{Cov}\{m(X_t)S_t^+(\upsilon_1),\ m(X_s)S_s^+(\upsilon_2)\}.$$

By Lemma 4, $E[S_1^+(\upsilon_j)] = o(g_n^{-2})$, and $E[m(X_1)S_1^+(\upsilon_j)] = o(g_n^{-2})$. By Lemma 3,

$$\sum_{t\neq s}\sum \left\{\mathrm{Cov}[m(X_t)S_t^+(\upsilon_1),\ m(X_s)S_s^+(\upsilon_2)]\right\} = O(n\ a^{2(p+1)}/g_n^2) = o(1),$$

by using similar arguments as Robinson (1983).

Therefore,

$$\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \mathcal{P}_+(\upsilon_2)\} = \sigma^2 T_1 + T_2 + o_p(g_n^{-4}),$$

where

$$T_1 = n\ g_n^{-4}\ E[S_t^+(\upsilon_1)\ S_t^+(\upsilon_2)].$$

$$T_2 = n\ g_n^{-4}\ E[m(X_t)^2 S_t^+(\upsilon_1)\ S_t^+(\upsilon_2)],$$

It suffices to prove the convergence of T_2. Note that, after the change of variable $x - \tau = u$,

$$T_1 = n\, g_n^{-4} \left\{ \int_{a\upsilon_2/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+\!\left(\frac{u - a\upsilon_2/g_n}{a}\right) K_+\!\left(\frac{u - a\upsilon_1/g_n}{a}\right) d\,u \right.$$

$$- \int_{a\upsilon_1/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a)\, K_+\!\left(\frac{u - a\upsilon_1/g_n}{a}\right) d\,u$$

$$- \int_{a\upsilon_2/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a)\, K_+\!\left(\frac{u - a\upsilon_2/g_n}{a}\right) d\,u$$

$$\left. + \int_0^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a)^2\, d\,u \right\}$$

$$= T_{21} + T_{22},$$

where

$$T_{21} = n\, g_n^{-4} \left\{ \int_{a\upsilon_2/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+\!\left(\frac{u - a\upsilon_2/g_n}{a}\right) \times \left\{ K_+\!\left(\frac{u - a\upsilon_1/g_n}{a}\right) - K_+(u/a) \right\} du, \right.$$

$$\left. - \int_{a\upsilon_1/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a) \left\{ K_+\!\left(\frac{u - a\upsilon_2/g_n}{a}\right) - K_+(u/a) \right\} du \right\}$$

$$T_{22} = n\, g_n^{-4} \int_0^{a\upsilon_1/g_n} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a)^2\, du.$$

Using a Taylor expansion for $K_+(u/a)$ around zero, as we did in Lemma 5, $T_{22} = O(g_n^{-5}) = o(g_n^{-4})$. On the other hand, $T_{21} = B_1 + B_2$, where

$$B_1 = n\, g_n^{-4} \int_{a\upsilon_2/g_n}^{\infty} m(u+\tau)^2 f_X(u+\tau) \left\{ K_+\!\left(\frac{u - a\upsilon_1/g_n}{a}\right) - K_+(u/a) \right\} \left\{ K_+\!\left(\frac{u - a\upsilon_2/g_n}{a}\right) - K_+(u/a) \right\} du$$

$$B_2 = n\, g_n^{-4} \int_{a\upsilon_1/g_n}^{a\upsilon_2/g_n} m(u+\tau)^2 f_X(u+\tau)\, K_+(u/a) \left\{ K_+\left(\frac{u - a\upsilon_1/g_n}{a} \right) - K_+(u/a) \right\} du$$

By a mvt argument and the change of variable $x = u/a$, for some $\zeta \in (0, 1)$,

$$B_2 = g_n^{-4} \int_{\upsilon_1/g_n}^{\upsilon_2/g_n} m(ax+\tau)^2 f_X(ax+\tau)\, K_+(x)\, K_+^1\left(x + \zeta\upsilon_1/g_n\right) \frac{x\,\upsilon_1}{g_n}\, dx$$

$$= O(g_n^{-5}) = o(g_n^{-4}).$$

$$B_1 = g_n^{-2} \int_{\upsilon_2/g_n}^{\infty} m(ax+\tau)^2 f_X(ax+\tau) \left\{ K_+\left(x - \upsilon_{1/g_n}\right) - K_+(x) \right\}$$

$$\left\{ K_+\left(x - \upsilon_{2/g_n}\right) - K_+(x) \right\} dx$$

$$= g_n^{-2} \int_{\upsilon_2/g_n}^{\infty} m(ax+\tau)^2 f_X(ax+\tau) \left\{ \upsilon_1 K_+^1(x)/g_n + \upsilon_1^2 K_+^2\left(x + \zeta\upsilon_1/g_n\right)/g_n^2 \right\}$$

$$\left\{ \upsilon_2 K_+^1(x)/g_n + \upsilon_2^2 K_+^2\left(x + \zeta\upsilon_2/g_n\right)/g_n^2 \right\} dx$$

$$= g_n^{-4}\, \upsilon_1\upsilon_2 \int_{\upsilon_2/g_n}^{\infty} m(ax+\tau)^2 f_X(ax+\tau)\, K_+^1(x)^2\, d\,x + O(g_n^{-5})$$

$$= g_n^{-4}\, \upsilon_1\upsilon_2\, m_+(\tau)^2 f_X(\tau) \int_{\upsilon_2/g_n}^{\infty} K_+^1(x)^2\, d\,x + O\left(g_n^{-5} + a\, g_n^{-4}\right)$$

$$= g_n^{-4}\, \upsilon_1\upsilon_2\, m_+(\tau)^2 f_X(\tau)\, \gamma + o(g_n^{-4}). \ \square$$

Lemma 8.- $\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1), \mathcal{P}_-(\upsilon_2)\} = o(g_n^{-4})$.

Proof: As in Lemma 7, by A2 and A3,

$\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1), \mathcal{P}_-(\upsilon_2)\} = \sigma^2\, n\, g_n^{-2}\, \mathrm{Cov}\, \{S_1^+(\upsilon_1), S_1^-(\upsilon_2)\}$ +

$$n\, g_n^{-4}\, \mathrm{Cov}\{m(X_t)S_t^+(\upsilon_1),\ m(X_t)S_t^-(\upsilon_2)\} + g_n^{-4} \sum_{t\neq s} \mathrm{Cov}\{m(X_t)S_t^+(\upsilon_1),\ m(X_s)S_s^-(\upsilon_2)\}.$$

Using the same arguments as in Lemma 7,

$$\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \mathcal{P}_-(\upsilon_2)\} = \sigma^2 T_1 + T_2 + o(g_n^{-4}),$$

where

$$T_2 = n\, g_n^{-4}\, \mathrm{E}[m(X_t)^2\ S_t^+(\upsilon_1)\ S_t^-(\upsilon_2)],$$

$$T_1 = n\, g_n^{-4}\, \mathrm{E}[S_t^+(\upsilon_1)\ S_t^-(\upsilon_2)].$$

We consider, without loss of generality the case $0 < \upsilon_1 < \upsilon_2$. Using the fact that $K_+(\omega) = 0$ if $\omega \le 0$ and $K_-(\omega) = 0$ if $\omega \ge 0$, $T_2 = T_{21} + T_{22}$, where

$$T_{21} = n\, g_n^{-4} \int_{a\upsilon_1/g_n}^{a\upsilon_2/g_n} m(u+\tau)^2\, f_X(u+\tau)\, K_-\left(\frac{u - a\upsilon_2/g_n}{a}\right) \times \left\{K_+\left(\frac{u - a\upsilon_1/g_n}{a}\right) - K_+(u/a)\right\} du,$$

$$T_{22} = n\, g_n^{-4} \int_{0}^{a\upsilon_2/g_n} m(u+\tau)^2\, f_X(u+\tau)\, K_-\left(\frac{u - a\upsilon_2/g_n}{a}\right) K_+(u/a)\, du.$$

Using same arguments as in Lemma 7, $T_{21} = o(g_n^{-4})$ and $T_{22} = o(g_n^{-4})$.

Lemma 9.-

$$\mathrm{Cov}\{\ell_\pm(\upsilon_1),\ \ell_\pm(\upsilon_2)\} = g_n^{-4}\, f_X(\tau)\, \gamma\, \upsilon_1\upsilon_2 + o(g_n^{-4}).$$

Proof.- Identical to Lemma 7. □

Lemma 10.-

$$\mathrm{Cov}\{\ell_+(\upsilon_1),\ \ell_-(\upsilon_2)\} = o(g_n^{-2}),$$

Proof.- Identical to Lemma 8. □

Remark: Lemmas 9 and 10 imply that

$$\mathrm{Cov}\{m_-(\tau)\ell_-(\upsilon_1) - m_+(\tau)\ell_+(\upsilon_1);\ m_-(\tau)\ell_-(\upsilon_2) - m_+(\tau)\ell_+(\upsilon_2)\} =$$

$$g_n^{-4}\ [m_-(\tau)^2 + m_+(\tau)^2]\ \upsilon_1\upsilon_2\ \gamma + o(g_n^{-4}).$$

Lemma 11.- (a) $\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \ell_-(\upsilon_2)\} = o(g_n^{-4})$,

(b) $\mathrm{Cov}\{\mathcal{P}_-(\upsilon_1),\ \ell_+(\upsilon_2)\} = o(g_n^{-4})$.

Proof: Identical to Lemma 8. □

Lemma 12.- $\mathrm{Cov}\{\mathcal{P}_+(\upsilon_1),\ \ell_+(\upsilon_2)\} = g_n^{-4}\ f_X(\tau)\ m_+(\tau)\ \upsilon_1\upsilon_2 + o(g_n^{-4})$.

Proof: Identical to Lemma 7. □

Remark: By lemmas 7 to 12,

$$\mathrm{Cov}(a(\upsilon_1),\ a(\upsilon_2)) = 2\ \sigma^2\ g_n^{-4}\ \upsilon_1\upsilon_2\ \gamma\ f(\tau) + o(g_n^{-4}).$$

Lemma 13.- $g_n^4\ \mathrm{E}\left[\left(\mathcal{P}_+(\upsilon_1) - \mathcal{P}_-(\upsilon_1) - (\mathcal{P}_+(\upsilon_2) - \mathcal{P}_-(\upsilon_2))\right)^2\right] \le C(\upsilon_1 - \upsilon_2)^2 + o(1)$.

Proof: The left hand side of the above inequality is bounded by

$$2\ g_n^4\ \left\{\mathrm{Var}\left[\mathcal{P}_+(\upsilon_1) - \mathcal{P}_+(\upsilon_2)\right] + \mathrm{Var}\left[\mathcal{P}_-(\upsilon_1) - \mathcal{P}_-(\upsilon_2)\right]\right\}.$$

Now, applying same arguments as in the proof of Lemma 7,

$$\mathrm{Var}\left[\mathcal{P}_+(\upsilon_1) - \mathcal{P}_+(\upsilon_2)\right] = g_n^{-4}\ f_X(\tau)\ (\upsilon_1 - \upsilon_2)^2\ \{m_+(\tau)^2 + \sigma^2\}\ \gamma + o(g_n^{-4}).\ \square$$

Lemma 14.- $g_n^4\mathrm{E}\left[\left(\ell_+(\upsilon_1) - \ell_+(\upsilon_2)\right)^2\right] \le C(\upsilon_1 - \upsilon_2)^2 + o(1)$.

Proof: Identical to Lemma 14. □

The lemmas below are applied in propositions 1.6. to 1.13, which are used for proving Theorem 3.2. Therefore, we are assuming than conditions for Theorem 3.2. hold. Without loss of generality, we will assume that τ=0, also $\nu = \upsilon/h_n$.

Lemma 15.-

$$\mathrm{E}[\mathcal{P}_+(\upsilon) - \mathcal{P}_-(\upsilon)] = O(h_n^{-4}).$$

Proof.- Since $f(.)$ and $m(.)$ are continuous everywhere,

$$\mathrm{E}[\mathcal{P}_+(\upsilon) - \mathcal{P}_-(\upsilon)] = \mathrm{E}[\mathcal{P}_+(\upsilon) - m(-\xi ae)f(-\xi ae) + m(0)f(0)]$$
$$- \mathrm{E}[\mathcal{P}_-(\upsilon) - m(-\tau ae)f(-\tau ae) + m(0)f(0)].$$

We only discuss the convergence of the first term, the second term converges using the same arguments. After standard kernel estimation algebra, the first term is asymptotically equivalent to

$$a^2\bar{m}^2(-a\upsilon)\int_0^\infty u^2K_+(u)du - a^2\bar{m}^2(0)\int_0^\infty u^2K_+(u)du$$
$$= a^2\left(\bar{m}^2(-a\upsilon) - \bar{m}^2(0)\right)\int_0^\infty u^2K_+(u)du$$
$$= O(a^3h_n^{-1})$$
$$= O(h_n^{-4}),$$

by B4'.

Lemma 16.-

$$\text{(a) } \mathrm{Cov}[\mathcal{P}_\pm(\upsilon_1),\ \mathcal{P}_\pm(\upsilon_2)] = o(h_n^{-8}).$$
$$\text{(b) } \mathrm{Cov}[\mathcal{P}_+(\upsilon_1),\ \mathcal{P}_-(\upsilon_2)] = o(h_n^{-8}).$$

Proof.- We only prove (a). The other cases are identical. Without loss of generality, we only consider the case $0<\upsilon_1<\upsilon_2$. By identical arguments as in Lemma 7, we need only to consider the order of magnitude of

$$\sigma^2(na^{p+1})^{-2}\sum_t E\left\{\left[K_+((X_t+ae\upsilon_1)/a)-K_+(X_t/a)\right]\left[K_+((X_t+ae\upsilon_2)/a)-K_+(X_t/a)\right]\right\}$$
$$+ (na^{p+1})^{-2}\sum_t E\left\{m(X_t)^2\left[K_+((X_t+ae\upsilon_1)/a)-K_+(X_t/a)\right]\left[K_+((X_t+ae\upsilon_2)/a)-K_+(X_t/a)\right]\right\}.$$

It suffices to prove the convergence of the first term (note that $m(.)$ is differentiable everywhere. After a change of variable, the first term in the above expression is equal to

$$\sigma^2(na^{p+1})^{-1}\left\{\int_0^\infty f(au)\left[K_+(u-e\upsilon_1)-K_+(u)\right]\left[K_+(u-e\upsilon_2)-K_+(u)\right]du\right.$$
$$\left. + \int_{-\upsilon_1}^0 f(au)\ K_+(u-e\upsilon_1)K_+(u-e\upsilon_2)du + \int_{-\upsilon_2}^{-\upsilon_1} f(au)\ K_+(u-e\upsilon_2)du\right\}.$$

Applying the mean value theorem, around u for the first term and around 0 for the last two terms (recall again that $K_+(0)=0$), we have that the above expression is $O(h_n^{-2}/(na^{p+1})) = O(h_n^{-8}a^2) = o(h_n^{-8})$.■

These last two lemmas show that $\mathcal{P}_+(\upsilon)-\mathcal{P}_-(\upsilon) = o_p(h_n^{-4})$.

Lemma 17.-

$$\text{(a) } \hat{f}_-(\upsilon)-\hat{f}_+(\upsilon) = O_p(a^3 + (na^{p+1})^{-1/2}),$$
$$\text{(b) } \hat{f}^1_-(\upsilon)-\hat{f}^1_+(\upsilon) = O_p(a^2 + (na^{p+2})^{-1/2}).$$

Proof.- It suffices to prove (a), being (b) completely identical.

$$\hat{f}_-(\upsilon)-\hat{f}_+(\upsilon)= [\ell_-(\upsilon) - \ell_+(\upsilon)] + [\hat{f}_-(0) - \hat{f}_+(0)].$$

Now, $[\ell_-(\upsilon) - \ell_+(\upsilon)]= o_p(h_n^{-4})$ by the previous two lemmas, and $\hat{f}_\pm(0)-f(0)= O_p(a^3+ (na^{p+1})^{-1/2})$ by standard kernel manipulations.

Lemma 19.-

$$\mathcal{P}^1_-(\upsilon)= O_p(a^2h_n^{-1}+ a^{-1}h_n^{-2}).$$

Proof.-

$$E[\mathcal{P}^1_-(\upsilon)]= (a^{p+2})^{-1} \int_{-\infty}^{-a\upsilon} \bar{m}(x)\left[K^1_-(x+ae\upsilon_1)/a)-K^1_-(x/a)\right]dx- (a^{p+2})^{-1}\int_{-a\upsilon}^{0} K^1_-(x/a)\ dx.$$

By B1', the second term is $O(h_n^{-2}a^{-1})$. With regard to the first term, it is equal to

$$a^{-1}\int_{-\infty}^{-\upsilon} m_-(au)f(au)\left[K^1_-(u+e\upsilon)-K^1_-(u)\right]du$$
$$= a^{-1}\int_{-\infty}^{-\upsilon}\left[m_-(au)f(au)-m_-(0)f(0)\right]\left[K^1_-(u+e\upsilon)-K^1_-(u)\right]du$$
$$+ a^{-1}m_-(0)f(0) \int_{-\infty}^{-\upsilon}\left[K^1_-(u+e\upsilon)-K^1_-(u)\right]du.$$

Using the fact that both m_- and f are three times continuously differentiable, $\int K^1_-(u)du=0$ and B1', B2' and B3, the last expression is $O(a^2h_n^{-1}+ h_n^{-2})$. With regard to its variance, by similar arguments as lemma 6 or 16, taking $\upsilon_1=\upsilon_2$, it is of an order of magnitude

$$(na^{p+2})^{-1}(a^{p+2})^{-1}\ E\left\{K^1_-(X_t+ae\upsilon_1)/a)-K^1_-(X_t/a)\right\}^2 = O(h_n^{-4}). \ \blacksquare$$

Lemma 20.-

$$E[\mathcal{P}^1_+(\upsilon)-\mathcal{P}^1_-(\upsilon)]= \Delta'(0)f(0)\ K^3_+(0)h_n^4/24 + o(h_n^{-4}).$$

Proof.- $E[\mathcal{P}^1_+(\upsilon)-\mathcal{P}^1_-(\upsilon)]= \alpha(\upsilon) - \alpha(0)$, where

$$\alpha(\upsilon)= (a^{2+p})^{-1} \int m(x)f(x) \left\{K^1_+((x+ae\upsilon)/a) - K^1_-((x+ae\upsilon)/a)\right\} dx.$$

By similar algebra as in lemma 5, $\alpha(\upsilon) = T_1+ T_2+ T_3+ o_p(h_n^{-4})$, where

$$T_1= a^{-1}\int_{\upsilon}^{\infty}\left(\bar{m}_-(0) + \bar{m}^1_+(0)a(u+e\upsilon) + \bar{m}^2_+(0)\ a^2(u+e\upsilon)^2/2 + \bar{m}^3_+(0)a^3(u+e\upsilon)^3/6\right)K^1_+(u)du,$$
$$T_2= a^{-1}\int_{0}^{\upsilon}\left(\bar{m}_-(0)+ \bar{m}^1_-(0)a(u+e\upsilon)+ \bar{m}^2_-(0)a^2(u+e\upsilon)^2/2+ \bar{m}^3_-(0)a^3(u+e\upsilon)^3/6\right)K^1_-(u)\ du,$$
$$T_3= a^{-1}\int_{-\infty}^{0}\left(\bar{m}_-(0)+ \bar{m}^1_-(0)a(u+e\upsilon)+ \bar{m}^2_-(0)a^2(u+e\upsilon)^2/2+ \bar{m}^3_-(0)a^3(u+e\upsilon)^3/6\right)K^1_-(u)\ du.$$

Using A4, B1' and B3 and after some algebra, very similar to Lemma 5,

$$\alpha(\upsilon)-\alpha(0)=\bar{m}^1_-(0)\int_0^{\upsilon}(u+e\upsilon)\,K^1_+(u)-\bar{m}^1_+(0)\int_0^{\upsilon}(u+e\upsilon)\,K^1_+(u)\,du+o(h_n^{-4}).$$
$$=-\Delta^1(0)\,f(0)\int_0^{\upsilon}(u+e\upsilon)\,K^1_+(u)+o(h_n^{-4})$$
$$=\Delta^1(0)/24\;K^3_+(0)\;f(0)h_n^{-4}+o(h_n^{-4}).\ \blacksquare$$

Lemma 21.-

(a) $\mathrm{Cov}[\mathcal{P}^1_+(\upsilon_1),\ \mathcal{P}^1_+(\upsilon_2)]=h_n^{-8}(m(0)^2+\sigma^2)\,f(0)\,\upsilon_1\upsilon_2\int(K^3_+(u))^2\,du+o(h_n^{-8})$,

(b) $\mathrm{Cov}[\mathcal{P}^1_-(\upsilon_1),\ \mathcal{P}^1_-(\upsilon_2)]=h_n^{-8}(m(0)^2+\sigma^2)\,f(0)\,\upsilon_1\upsilon_2\int_{-\infty}^{0}(K^3_-(u))^2\,du+o(h_n^{-8})$,

(c) $\mathrm{Cov}[\mathcal{P}^1_+(\upsilon_1),\ \mathcal{P}^1_-(\upsilon_2)]=o(h_n^{-8})$.

Proof.-We only prove (a). The other cases are identical. Without loss of generality, we only consider the case $0<\upsilon_1<\upsilon_2$. By identical arguments as in Lemma 6, we need only to consider the order of magnitude of

$$(na^{p+3})^{-1}(a^{p+1})^{-1}\Big\{\sigma^2\int\Big[K^1_+((x+ae\upsilon_1)/a)-K^1_+(x/a)\Big]\Big[K^1_+((x+ae\upsilon_2)/a)-K^1_+(x/a)\Big]f(x)dx$$
$$+\int m(x)^2\Big[K^1_+((x+ae\upsilon_1)/a)-K^1_+(x/a)\Big]\Big[K^1_+((x+ae\upsilon_2)/a)-K^1_+(x/a)\Big]f(x)dx\Big\}.$$

We only consider the second term, which is asymptotically equivalent to

$$(na^{p+3})^{-1}(a^{p+1})^{-1}\{A_1+A_2+A_3\},$$

where

$$A_1=\int_{a\upsilon_2}^{\infty}m(x)^2\Big[K^1_+((x+ae\upsilon_1)/a)-K^1_+(x/a)\Big]\Big[K^1_+((x+ae\upsilon_2)/a)-K^1_+(x/a)\Big]f(x)dx,$$
$$A_2=-\int_{a\upsilon_1}^{a\upsilon_2}m(x)^2\Big[K^1_+((x+ae\upsilon_1)/a)-K^1_+(x/a)\Big]K^1_+(x/a)f(x)dx,$$
$$A_3=\int_0^{a\upsilon_1}m(x)^2\;K^1_+(x/a)\;K^1_+(x/a)\;f(x)dx.$$

By repeating the same steps as in the proof of Lemma 7, we get the desired result. Observe that $m(.)$ is differentiable.

REFERENCES

Andrews, D.W.K. (1993), "Testing for structural instability and structural change with unknown change point", *Econometrica*, 61, 821-856.

Bai, J. (1992), "Estimation of Structural change in econometric models", Manuscript, Dep. of Economics, Univ. de California, Berkeley.

Bhattacharya, P.K. (1987), "Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case", *Journal of Multivariate Analysis*, 23, 183-208.

Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for testing the constancy of regression relationships over time", *Journal of the Royal Statistical Society, Ser. B*, 37, 149-192.

Delong, J.B. and L.H. Summers (1988), "How does macroeconomic policy affect output?", *Brooking Papers of Economic Activity*, no. 2, 433-480.

Eddy, W.F. (1980), "Optimum kernel estimators of the mode", *Annals of Statistics*, 8, 870-882.

Eddy, W.F. (1982), "The asymptotic distribution of kernel estimators of the mode", *Zeitchrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58, 7-20.

Feder, P.I. (1975), "On asymptotic distribution theory in segmented regression problem-identified case", *Annals of Statistics*, 3, 49-83.

Gasser, T., H.G. Müller and V. Mammitzsch (1985), "Kernels for nonparametric curve estiamtion", *Journal of the Royal Statistical Society, Ser. B*, 47, 238-252.

Hawkins, D.M. (1977), "Testing a sequence of observations for a change in location", *Journal of the American Statistical Association*, 72, 180-186.

Hendry, D.F. and N.R. Ericsson (1991), "An econometric analysis of UK money demand in monetary trends in the United states and the United Kingdom by Milton Friedman and Ana Schwartz", *American Economic Review* 81, 8-38.

Hidalgo, J. (1994), "A conditional nonparametric moment test for structural stability". Forthcoming in *Econometric Theory.*

Hinkley, D. (1969), "Inference about the intersection in two phase regression", *Biometrika*, 56, 495-504.

Hinkley, D. (1971), "Inference about the change point in a sequence of random variables", *Biometrika*, 57, 1-17.

Kramer, W., W. Ploberger and R. Alt (1988), "Testing for structural changes in dynamic models", *Econometrica*, 56, 1355-1370.

Krishnaiah, P.R. and B.Q. Miao (1988), "Review about staimation of change points", in *Handbook of Statistics*, Vol 7, ed. by P.R. Krishnaiah and C.R. Rao, New York, Elsevier.

Müller, H-G. (1992), "Change-points in nonparametric regression analysis", *Annals of Statistics*, 20, 737-761.

Nadaraya, E.A. (1964), "On estimating regresion", *Theory of Probability and its Applications* 9, 141-142.

Nelson, C.R. and C.I. Plosser (1982), "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics*, 10, 139-162.

Page, E.S. (1955), "A test for a change in a parameter ocurring at an unknown point", *Biometrika*, 42, 523-527.

Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

Picard, D. (1985), "Testing and estimating change points in time series", *Journal of Applied Probability*, 14, 411-415.

Quandt, R.E. (1960), "Test of the hypothesis that a linear regresion 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. (1989), "Nonparametric estimation of time-varying parameters" in *Statistical Analysis and Forecasting of Economic Structural Change*, (P.Hackl, ed.), Springer-Verlag, Berlin, pp. 253-264.

Rosenblatt, M. (1956), "A central limit theorem and a strong mixing condition", *Proceedings of the National Academy of Sciences, USA*, 42, 43-47.

Rosenblatt, M. *Stochastic Curve Estimation*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 3.

Worsley, K.L. (1979), "On the likelihood ratio test for a shift in locations of normal populations", *Journal of the American Statistical Association*, 72, 180-186.

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.

Zacks, S. (1983), "Survey of classical and Bayesian approaches to the change point problem: fixed sample and sequential procedures of testing and estimation", in *Recent Advances in Statistics*, ed. by M.H. Rizvi et al, Academic Press, New York.

Zitov, E. and D.W.K. Andrews (1992), "Further evidence on the great crash, the oil price shock, and the unit-root hypothesis", *Journal of Business and Economic Statistics*, 10, 251-270.

TABLE 1

Biases and RMSE of change point estimates based on model (4.1) and 5000 replications for different sample and jump sizes.

	n = 50		n = 100		n = 200	
	Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
α= 0.5						
C= 1	-0.098	0.337	-0.087	0.270	-0.071	0.212
C= 2	-0.146	0.294	-0.135	0.235	-0.114	0.189
C= 3	-0.160	0.287	-0.152	0.233	-0.136	0.190
C= 4	-0.166	0.287	-0.159	0.232	-0.150	0.195
α= 1						
C= 1	-0.057	0.244	-0.037	0.162	-0.019	0.101
C= 2	-0.102	0.231	-0.084	0.169	-0.062	0.120
C= 3	-0.123	0.238	-0.109	0.182	-0.090	0.137
C= 4	-0.135	0.245	-0.122	0.190	-0.108	0.150
α= 2						
C= 1	-0.011	0.143	-0.003	0.081	-0.001	0.046
C= 2	-0.054	0.163	-0.037	0.103	-0.023	0.067
C= 3	-0.078	0.180	-0.061	0.125	-0.047	0.087
C= 4	-0.092	0.190	-0.077	0.138	-0.063	0.101

TABLE 2

Biases and RMSE of change point estimates based on model (4.2) and 5000 replications for different sample and jump sizes.

	n = 50		n = 100		n = 200	
	Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
α= 0.5						
C= 1	-0.005	0.117	-0.005	0.089	0.001	0.053
C= 2	-0.019	0.115	-0.021	0.090	-0.016	0.057
C= 3	-0.024	0.116	-0.026	0.091	-0.022	0.058
C= 4	-0.026	0.115	-0.029	0.092	-0.026	0.059
α= 1						
C= 1	0.020	0.071	0.012	0.041	0.011	0.023
C= 2	0.006	0.077	0.001	0.045	-0.000	0.025
C= 3	0.001	0.080	-0.003	0.046	-0.005	0.026
C= 4	-0.002	0.081	-0.006	0.047	-0.008	0.027
α= 2						
C= 1	0.033	0.032	0.020	0.018	0.009	0.011
C= 2	0.025	0.034	0.012	0.020	0.003	0.012
C= 3	0.022	0.035	0.009	0.021	-0.001	0.013
C= 4	0.020	0.036	0.007	0.021	-0.003	0.014

TABLE 3

Point estimates and confidence intervals for the location of the break point of series of Nominal Wages (1900-1970). The bandwidth choice is $h = C\, n^{-1/(p+5)}$ for C= 1,2,3, and 4. The τ_0 estimates are in parenthesis.

	Point estimate	95% Confidence Interval
One Lag (p= 1)		
C= 1	1930 (0.443)	1929.57 to 1930.43 (0.436 to 0.489)
C= 2	1930 (0.443)	1928.89 to 1931.10 (0.427 to 0.459)
C= 3	1930 (0.443)	1928.49 to 1931.51 (0.421 to 0.464)
C= 4	1931 (0.457)	1929.24 to 1932.75 (0.432 to 0.482)
Two Lags (p= 2)		
C= 1	1930 (0.493)	1929.65 to 1930.34 (0.444 to 0.454)
C= 2	1929 (0.434)	1928.01 to 1929.98 (0.420 to 0.449)
C= 3	1930 (0.449)	1928.56 to 1931.43 (0.428 to 0.470)
C= 4	1930 (0.449)	1928.25 to 1931.74 (0.424 to 0.474)
Three Lags (p= 3)		
C= 1	1929 (0.441)	1928.76 to 1929.24 (0.437 to 0.445)
C= 2	1929 (0.441)	1928.14 to 1929.85 (0.428 to 0.454)
C= 3	1929 (0.441)	1927.36 to 1930.66 (0.417 to 0.466)
C= 4	1929 (0.441)	1926.56 to 1931.43 (0.405 to 0.477)

TABLE 4

Point estimates and confidence intervals for the location of the break point of series of Common Stock Prices (1871-1970). The bandwidth choice is $h = C\, n^{-1/(p+5)}$ for C= 1,2,3, and 4. The τ_0 estimates are in parenthesis.

	Point estimate	95% Confidence Interval
One Lag (p= 1)		
C= 1	1915 (0.455)	1911.87 to 1918.13 (0.422 to 0.486)
C= 2	1915 (0.455)	1908.40 to 1921.60 (0.388 to 0.521)
C= 3	1916 (0.465)	1904.51 to 1927.49 (0.348 to 0.580)
C= 4	1916 (0.465)	1900.38 to 1931.62 (0.306 to 0.622)
Two Lags (p= 2)		
C= 1	1914 (0.454)	1910.82 to 1917.18 (0.421 to 0.486)
C= 2	1914 (0.454)	1908.15 to 1919.85 (0.393 to 0.514)
C= 3	1915 (0.464)	1905.08 to 1924.92 (0.361 to 0.566)
C= 4	1916 (0.474)	1902.12 to 1929.88 (0.331 to 0.617)
Three Lags (p= 3)		
C= 1	1922 (0.536)	1919.78 to 1924.22 (0.513 to 0.558)
C= 2	1913 (0.443)	1906.83 to 1919.16 (0.380 to 0.507)
C= 3	1915 (0.464)	1902.82 to 1927.18 (0.338 to 0.589)
C= 4	1915 (0.464)	1894.55 to 1935.45 (0.253 to 0.675)

TABLE 5

Point estimates and confidence intervals for the location of the break point of series of Quarterly Real Gross National Product (1947:I–1986:III). The bandwidth choice is $h = C\, n^{-1/(p+5)}$ for C= 1,2,3, and 4. The τ_0 estimates are in parenthesis.

	Point estimate	95% Confidence Interval
One Lag (p= 1)		
C= 1	1965:IV (0.475)	1965.90 to 1966.10 (0.472 to 0.477)
C= 2	1966:III (0.490)	1966.61 to 1966.88 (0.490 to 0.497)
C= 3	1966:IV (0.500)	1966.84 to 1967.16 (0.496 to 0.503)
C= 4	1967:I (0.502)	1967.07 to 1967.42 (0.501 to 0.512)
Two Lags (p= 2)		
C= 1	1967:II (0.516)	1967.40 to 1967.60 (0.513 to 0.518)
C= 2	1966:III (0.497)	1966.61 to 1966.89 (0.493 to 0.500)
C= 3	1966:III (0.497)	1966.59 to 1966.91 (0.493 to 0.501)
C= 4	1966:IV (0.503)	1966.81 to 1967.18 (0.498 to 0.508)
Three Lags (p= 3)		
C= 1	1970:II (0.603)	1970.43 to 1970.57 (0.600 to 0.604)
C= 2	1966:II (0.500)	1966.36 to 1966.64 (0.496 to 0.504)
C= 3	1966:II (0.500)	1966.29 to 1966.71 (0.495 to 0.505)
C= 4	1966:II (0.500)	1966.22 to 1966.77 (0.493 to 0.507)

TABLE 6

Point estimates and confidence intervals for the location of the break point of series of Rate of Growth of Quarterly Real Gross National Product (1947:I–1986:III). The bandwidth choice is $h = C\, n^{-1/(p+5)}$ for $C=$ 1,2,3, and 4. The τ estimates are in parenthesis.

	Point estimate	95% Confidence Interval
One Lag (p= 1)		
$C=1$	1967:IV (0.522)	1963.05 to 1972.44 (0.403 to 0.642)
$C=2$	1967:III (0.516)	1962.12 to 1972.87 (0.379 to 0.653)
$C=3$	1967:II (0.509)	1961.09 to 1973.40 (0.353 to 0.666)
$C=4$	1967:II (0.509)	1960.37 to 1974.12 (0.334 to 0.684)
Two Lags (p= 2)		
$C=1$	1967:III (0.519)	1962.63 to 1972.36 (0.394 to 0.644)
$C=2$	1967:II (0.513)	1961.50 to 1972.99 (0.365 to 0.660)
$C=3$	1967:II (0.513)	1960.61 to 1973.88 (0.342 to 0.683)
$C=4$	1967:II (0.513)	1959.80 to 1974.69 (0.322 to 0.704)
Three Lags (p= 3)		
$C=1$	1967:I (0.516)	1963.47 to 1970.52 (0.425 to 0.607)
$C=2$	1966:IV (0.509)	1960.78 to 1972.71 (0.356 to 0.663)
$C=3$	1966:IV (0.509)	1958.27 to 1975.22 (0.291 to 0.728)
$C=4$	1966:IV (0.509)	1955.76 to 1977.74 (0.226 to 0.793)

Figure 1

Jump Function for Nominal Wages (1900–1970).

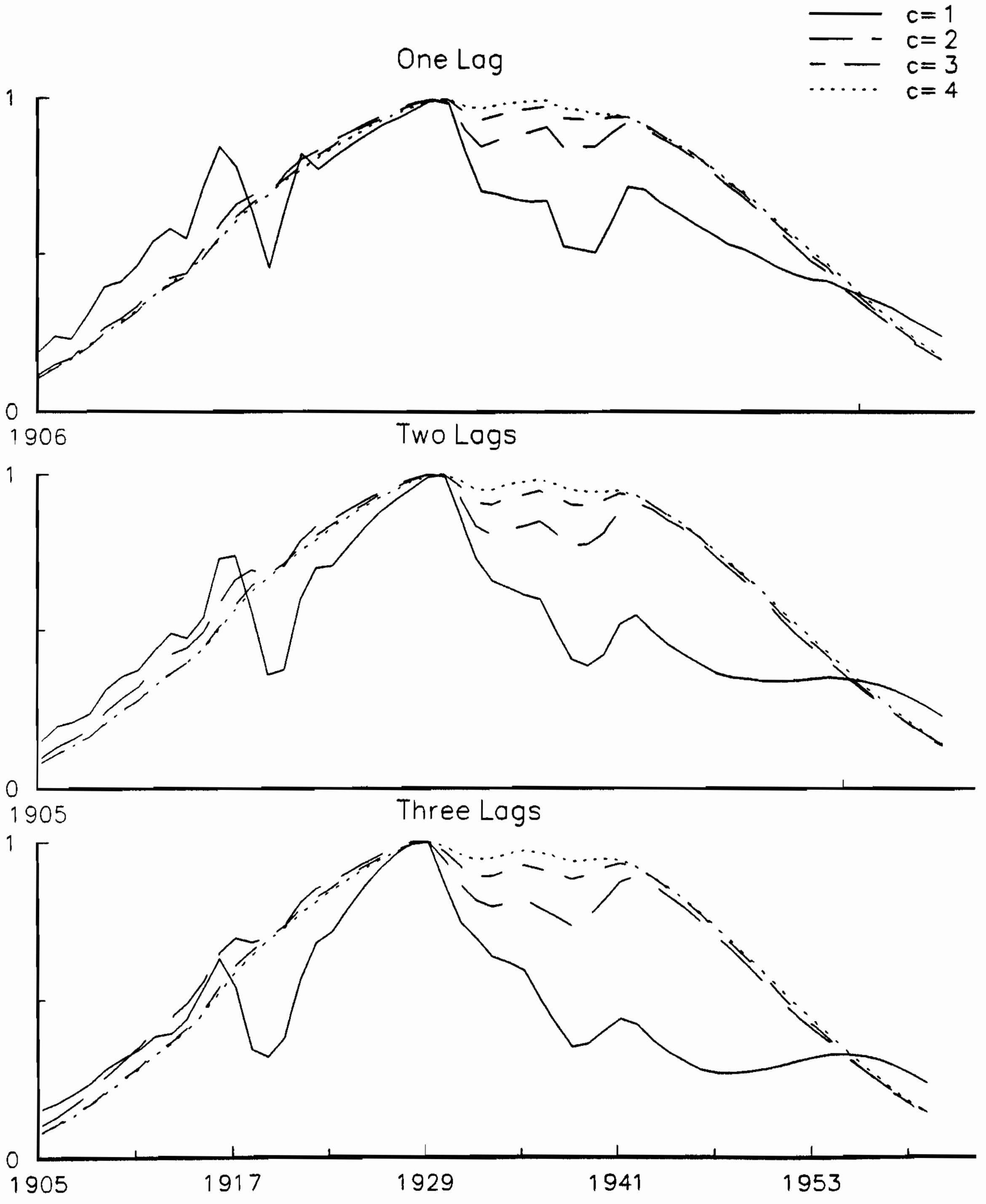

Figure 2

Jump Function for Common Stock Prices (1871–1970).

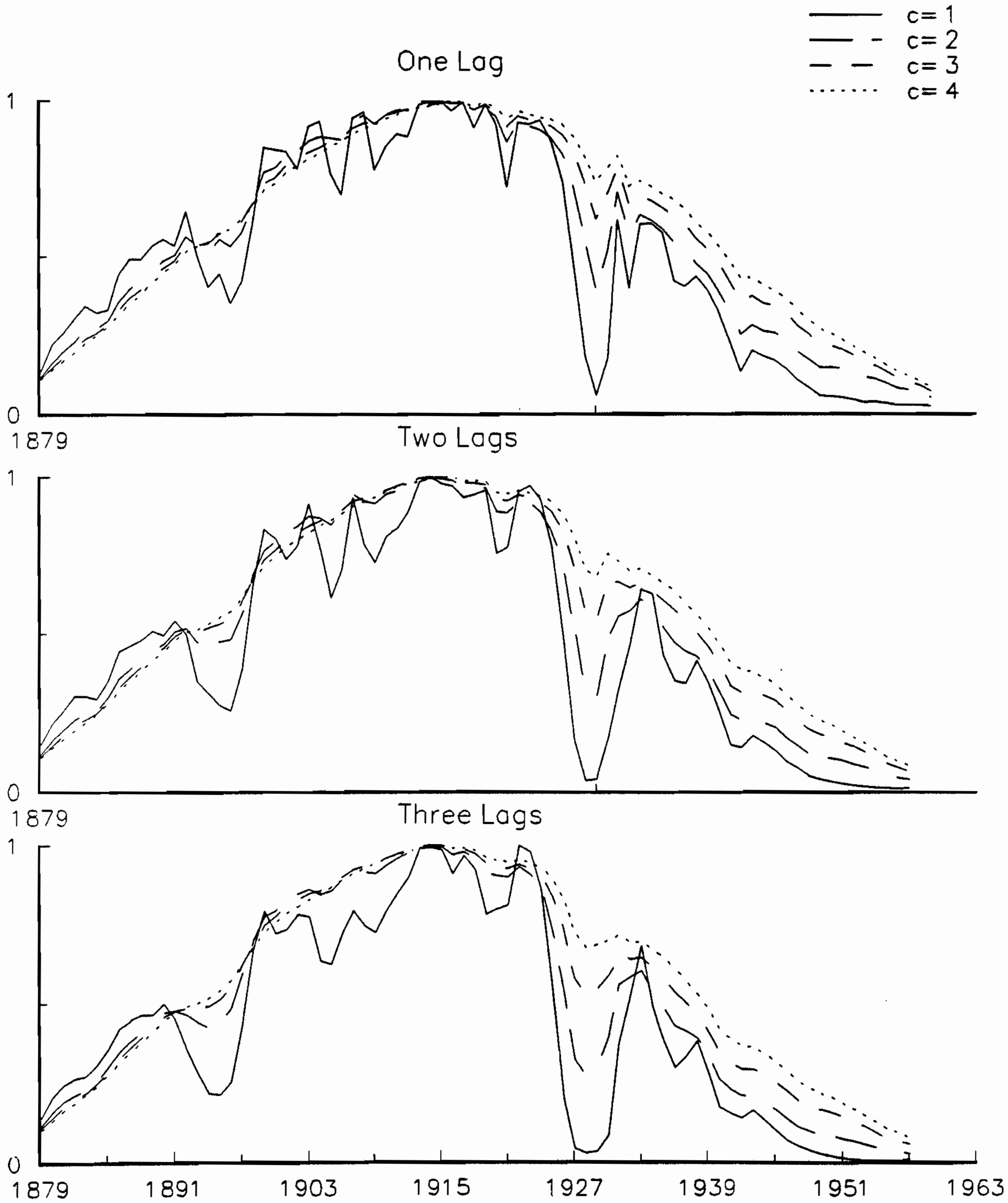

Figure 3

Jump Function for Quarterly GNP (1947:I,1986:III)

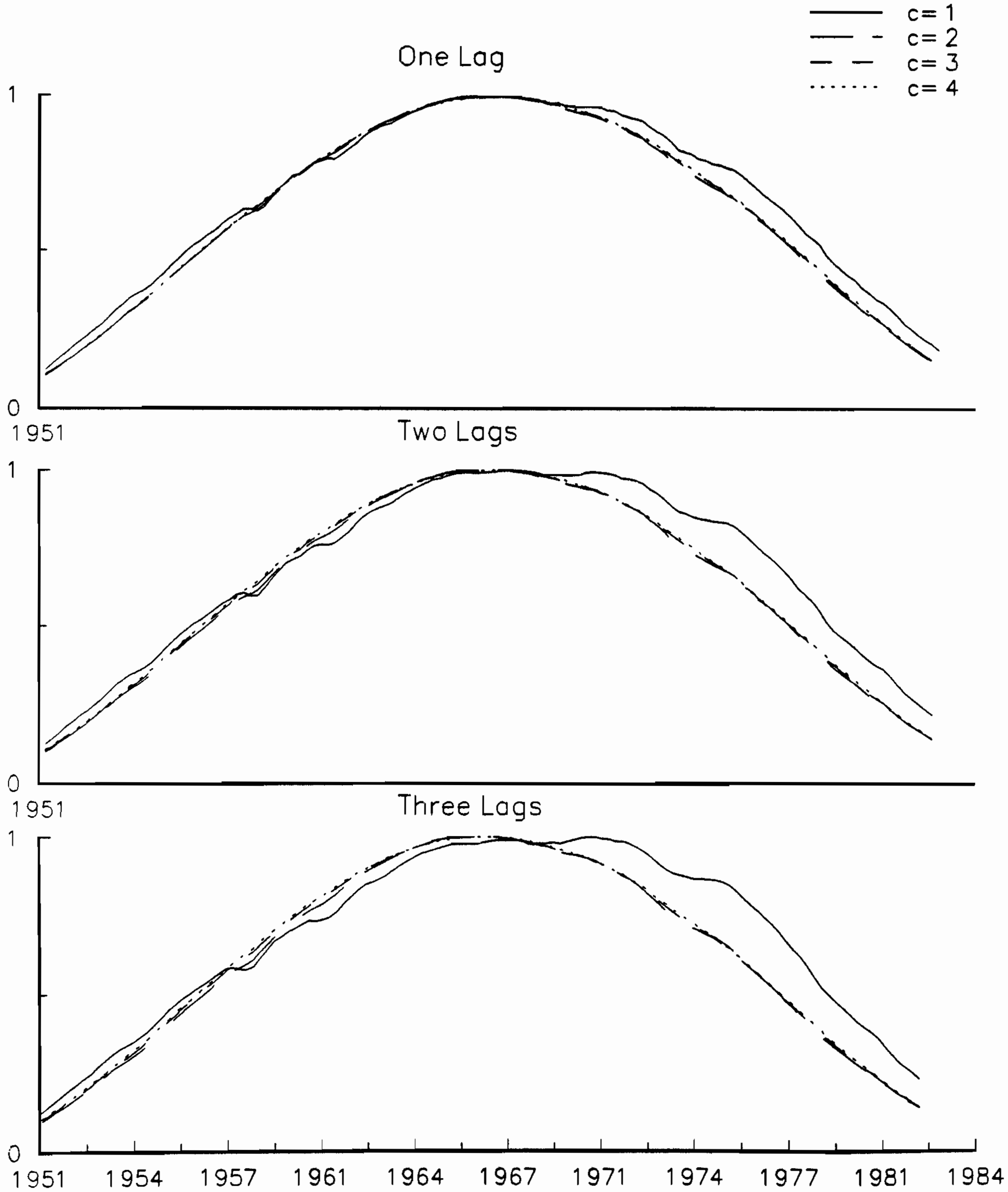

Figure 4

Jump Function for Growth Rate in Quarterly GNP (1947:I,1986:III)