Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 119 (2004) 19–44
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

τ-estimators of regression models with structural change of unknown location

Inmaculada Fiteni*

Departamento de Gestión y Predicción de la Energía, Endesa Energía, Paseo de Recoletos no 5, Madrid, Spain

Accepted 19 March 2003

Abstract

This paper concerns with robust estimation of linear regression models with structural change of unknown location under possibly contaminated distributions both for the regressors and for the perturbance term. Existing estimators will be inefficient in this context. Furthermore, they may protect against outlying Y_t, but cannot cope with leverage points, namely outliers in the factor space, which could have large influence on the fit. As a result, these estimators could not discriminate between outlier observations and structural break points, misplacing the shift location. This fact can be of special importance in practice. Therefore, it may be advisable to consider robust estimators under possible leverage points in a structural change context. Thus, we propose the τ-estimator, introduced by Yohai and Zamar (J. Am. Stat. Assoc. 83 (1988) 406) in the standard context of no change. This type of estimator is qualitatively robust, with the best possible breakdown-point and highly efficient under normal errors. The asymptotic distribution of the break location estimator is obtained both for fixed magnitude of shift and for shift with magnitude converging to zero as the sample size increases. The analysis is carried out in the framework of general NED dependence conditions for the data. Monte Carlo experiments illustrates the performance of our estimators in finite samples.

JEL classification: C22

Keywords: Structural change; Robustness; τ-estimators; NED-dependence

1. Introduction

In this paper we consider the problem of estimating a regression model with structural change under possibly contaminated distribution for both the regressors and the

* Corresponding author.
E-mail address: ifiteni@endesa.es (I. Fiteni).

0304-4076/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0304-4076(03)00153-2

perturbance term. Models with structural breaks are often encountered to specify economic time series data. It is a classic result that estimates of regression parameters corresponding to a model with a structural change with neglected breaks, are no longer consistent. There is a number of economic applications considering the existence of structural breaks. Empirical results suggest that, in many areas, allowing for structural breaks has changed conclusions about related inferences. For example, in macroeconomics, it is well known the innovative papers of Perron (1989, 1990) and Rappoport and Reichlin (1989), which independently suggested as a plausible model for a wide variety of economic variables stationarity around a time trend with a break, rejecting the unit root hypothesis previously supported by Nelson and Plosser (1982). Since the work of Friedman and Mieselman (1963), other of the traditional empirical problems in macroeconomics has been whether money has a strong and stable link to aggregate output. Feldstein and Stock (1994) study the stability in money–output regressions for three monetary aggregates, the monetary base, Ml and M2. The structural change in the international trade has been analyzed by Ben-David and Papell (1997), in particular, in the export-GDP and import-GDP ratios. In financial time series, the consideration of structural breaks is also of particular importance. A key part of numerous empirical models is about the magnitude of the impact that a structural change may have on the conclusions, because researchers have failed to take into account a possibly significant time break. As example, Corbae and Ouliaris (1991) and Perron and Vogelsan (1992) introduce this question in the long rung Purchasing Power Parity and they prevent that allowing for structural breaks reverses conclusions about unit roots in this context. This is an important issue recently emerging. Garcia and Perron (1996) analyze this same problem in real interest rates using the Markov switching model. Bekaert and Gray (1998) incorporate the possibility of jumps in their empirical model of exchange rates. Reyes (1999) examines the relationship between firm size and time-varying betas for UK stocks. A survey of empirical applications of the structural change problem in economics is given by Stock (1997) and Maddala and Kim (1998), among others. For further references on parameter instability and breaks, the reader is referred to the reviews and bibliographies by Hackl and Westlund (1989, 1991), Krishnaiah and Miao (1988), Krämer and Sonnberger (1986), Stock (1997) and Csörgo and Horváth (1997), to mention only a few.

While structural change problems are related with consistency and valid inferences, as mentioned earlier, robust estimation methods are mainly concerned with both safety and efficiency under possibly contaminated distributions.[1] The first is an specification issue and the second is basically an estimation problem. The consideration of robust estimators in the usual context (of no structural change) is motivated by the usual poor behavior resulting from the use of classical estimators, even under slight violations

[1] *Contamination* can be outlined as any possible deviation from the assumed parametric model, under which classical procedures lay off being optimal. Belonging to this set, we can consider two standard situations: the occurrence of outliers, the most dangerous type of deviation which could yield a (non-protected) classical estimator to break down, spoiling the estimate completely (the *breakdown point* is a related concept), or a thick-tailed distribution, under which, classical LS estimators could result highly inefficient (corresponding variance is determined by the *influence function*). Indeed, the first situation may be generated by the second one. For definitions see, e.g. Hampel et al. (1986).

of the strict model assumptions. In the structural change context, when estimating a shift point, robust methods become particularly relevant. The fact is that estimators which are not protected against deviations from the model distribution and/or outlying observations can produce disastrous effects on the estimates, in such a way that they will not be able to discriminate between an outlying observation and a structural break point. This aspect will be illustrated by an example, in Remark 1.

Classical estimators of structural change models include the maximum likelihood estimation by Hinkley (1970), Bhattacharya (1987) and Yao (1987) for the *i.i.d.* case and Picard (1985) for a Gaussian autoregressive process, among others. Bai (1994) and Bai and Perron (1998) estimate the unknown change point by least squares (LS), assuming, respectively, a linear process for the error term and strong mixing dependence conditions for the data. Delgado and Hidalgo (2000) estimate the location of a structural change in a nonlinear model. Robust methods have been also considered. Bai (1995, 1998) proposes to use the least absolute deviations (LAD) estimator, which has good properties in terms of robustness but is highly inefficient under normality. Fiteni (2002) proposes an M-estimator, which constitutes a compromise between the efficiency of LS and the robustness of LAD. However, the consideration of regression models with random carriers is crucial when inspecting economic data and above proposals for robust estimation fail to have a high breakdown point under leverage. In this context, we recommend the τ-estimators, which protect against outlying observations both, for the dependent variable (OY axis) and for the random carriers (OX axis). This type of estimator was first developed by Yohai and Zamar (1988) for estimating continuous regression models assuming *i.i.d.* observations. We generalize their results for a structural change model under general NED-dependence conditions for the data.

Remark 1. As in Hampel et al. (1986), we illustrate, by an example, the advantage of using a robust estimation procedure in a structural change context. In particular, we contemplate the possible impact of a single outlier observation on the shift estimator when using nonprotected estimators in practice. The considered model is $Y_t = \alpha_t + \beta X_t + U_t,\ t \geqslant 1$. Regressor and error are *i.i.d.* N$(0, 0.1)$ and mutually independent. Then, we generate a first sample of size $n = 30$, with $\beta = 1$ and $\alpha_t = 1 + I(t > 15)$, where $I(\cdot)$ represents the indicator function. The data is shown in Table 1 and Fig. 1, provided in the appendix.

Next, we generate two other samples by replacing the observation corresponding to $t = 22$, i.e., $P = (2.84, 22, 0.08)$, first by $P_1 = (-0.80, 22, 0.08)$ and secondly by $P_2 = (0, 22, 2.0)$, outliers in the OY and OX axis, respectively. For these three samples and considering unknown the break location, we estimate the above model by using four estimation methods: LS, LAD, Huber and τ-estimator, the latter proposed in this paper. Fig. 2, in the appendix, shows the results.

As expected, with the first sample, in absence of outliers, all the estimators locate the break in the true break point position. Moreover, we obtain a similar fitting pattern in all cases. With the second sample, in the presence of an outlying Y_t, the (nonrobust) LS estimator is the only one which spoils the estimation, in such a way that locates the break at $t = 22$, the time corresponding to the outlier observation. LAD and Huber estimators share this same behavior when considering the third sample, in the presence

Table 1
Data corresponding to the original (no contaminated) sample

Time	Y	X	Time	Y	X	Time	Y	X
1	1.03	−0.30	11	0.83	−0.14	21	2.08	−0.27
2	0.67	−0.21	12	1.45	−0.00	**22**	**2.84**	**0.08**
3	0.17	−0.16	13	0.87	0.43	23	3.17	0.54
4	1.27	0.06	14	0.98	0.17	24	1.24	−0.41
5	0.87	−0.06	15	1.47	0.08	25	2.50	0.45
6	1.59	0.33	16	1.57	−0.14	26	2.41	0.30
7	1.43	−0.04	17	1.39	−0.28	27	1.34	−0.43
8	0.98	−0.08	18	2.31	0.26	28	1.07	−0.48
9	0.57	−0.04	19	1.89	0.20	29	2.30	0.32
10	1.06	0.07	20	2.18	0.17	30	2.47	−0.37

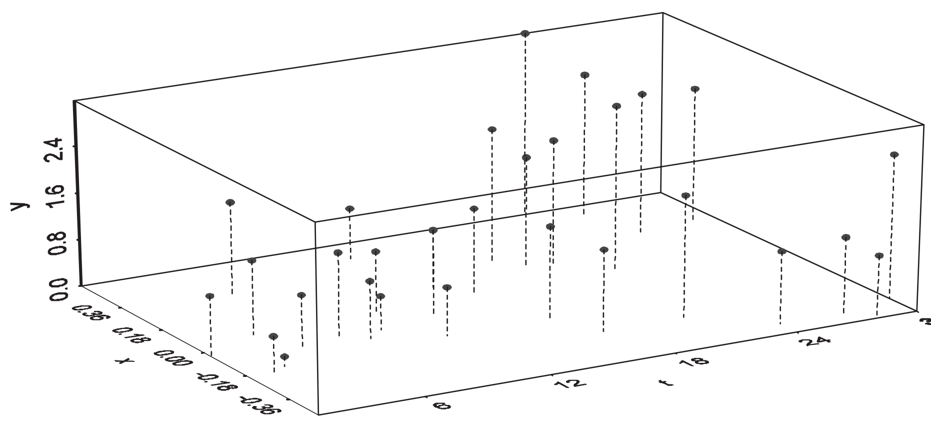

Fig. 1. Data corresponding to the original (no contaminated) sample.

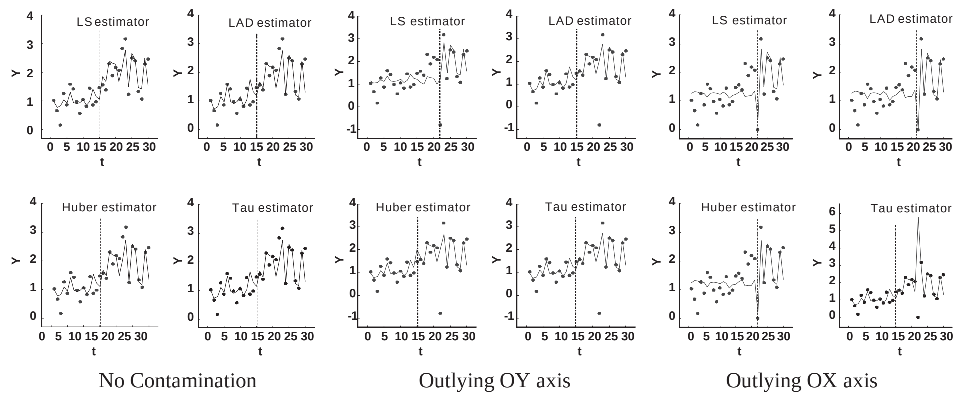

Fig. 2. Scatter plot, fitting curve (solid line) and break location estimate (vertical dotted line) for three samples: no contaminated with outline OY axis and with outlying OX axis.

of a leverage point (P_2). This is the consequence of their lack of robustness in this context. On the contrary, τ-estimator, the only one which prevents against observations of this nature, carries on locating the shift point at the true break position.

In this paper, we provide a suitably robust estimator of a possibly contaminated model with a structural break. Contamination may be present in both the regression carriers and the perturbance term. As in the standard context of no change, existing nonprotected procedures may spoil the estimate completely. As a consequence, in fact, they could mistake an outlier observation for a structural break in this context.

The rest of the paper is organized as follows. Section 2 outlines the procedure for obtaining the τ-estimator in a regression model with structural change and provides the set of sufficient assumptions for justifying the asymptotic properties of the corresponding estimators. The model allows for general forms of serial dependence. In Section 3, rates of convergence and limiting distributions of the estimators will be obtained both for fixed and for shrinking magnitude of shift. The latter is essential for the derivation of feasible confidence intervals for the break point location, provided that only in this case the asymptotic distribution will be pivotal. Section 4 reports a Monte Carlo experiment which illustrates the estimator performance in finite sample situations. Appendixes of proofs and tables are provided in Sections 5 and 6, respectively.

2. Model and assumptions

Let $\{Z_t=(Y_t,X_t)\}_{t=1}^n$ be a sample of Z, a $\mathbb{R}\times\mathbb{R}^p$-valued stochastic process defined on the probability space $(\text{-},\mathscr{F},P)$, such that

$$Y_t = X_t'\beta_{10}I(t\leqslant[n\pi_0]) + X_t'\beta_{20}I(t>[n\pi_0]) + U_t, \qquad t=1,\dots,n \tag{1}$$

a linear model with a simple shift, where $[\cdot]$ represents the nearest integer function and $\{U_t\}_{t=1}^n$ is the sequence of perturbances. Define the parameters vector $\xi_0=(\beta_{10}',\beta_{20}',\pi_0)'$, where $\beta_{j0}\in\Theta\subset\mathbb{R}^p$, for $j=1,2$, with $\beta_{10}\neq\beta_{20}$, and $\pi_0\in\Pi\subset(0,1)$ is the shift point location, which is also unknown.

The focus of this paper is to estimate, using the proposal of the τ-estimators, a linear regression model under the maintained hypothesis that there exists a shift with unknown location and size $\lambda=\beta_{10}-\beta_{20}\neq 0$. Eq. (1) allows all the regression parameters to switch between regimes, but the results generalize to the case where only a subset of parameters change at a given time point. The difference is only an issue of efficiency, because a partial structural change model consists on incorporating additional null restrictions about some components of λ.

To outline the estimator, we make use of the following two definitions:

Definition 1 (Huber, 1981). Given $\{v_t\}_{t=1}^n$, a sample of a random variable v, the M-estimator of scale (or M-scale), denoted by $s_n(v)$, is such that $n^{-1}\sum_{t=1}^n \rho(v_t/s_n(v))=b$, for a given function $\rho:\mathbb{R}\to\mathbb{R}$ and $b=\mathrm{E}_\Phi[\rho(v)]$, where Φ represents the standard normal distribution.

S-estimators, provided by Rousseauw and Yohai (1984), are defined by the minimization of the residuals M-scale. They showed that these estimators may have a high breakdown point, but their loss of efficiency under normality is sizable. To solve this, Yohai and Zamar (1988) propose the τ-estimators of the regression coefficients as the minimizers of the residuals τ-scale, defined below. They guarantee simultaneously the best possible breakdown point and efficiency at the normal model.

Definition 2 (Yohai and Zamar, 1986). Consider two real functions ρ_1 and ρ_2 and let s_n be the M-estimator of scale based on ρ_1. Then, given a sample $\{v_t\}_{t=1}^n$ of a random variable v, the τ-estimator of scale (or τ-scale), denoted by τ_n, is such that $\tau_n^2(v) = s_n^2(v)n^{-1}\sum_{t=1}^n \rho_2(v_t/s_n(v))$.

Particular functions ρ_1 and ρ_2 will determine the type of estimator. For $\rho_2(u) = u^2$, we obtain the least-squares estimator and if we consider $\rho_1 = \rho_2$, then $\tau_n = \sqrt{b}s_n$, corresponding to the S-estimator. As mentioned before, the τ-estimator is more robust than the first under possible contaminated distributions and more efficient than the second at a Gaussian model.

Next, given model (1), we define the τ-estimator $\hat{\xi}_n$ of the parameters vector ξ_0 as follows:

$$\hat{\xi}_n = \underset{\xi \in \Theta^2 \times \Pi}{\arg\min}\, \mathscr{V}_n(\xi), \tag{2}$$

such that,

$$\mathscr{V}_n(\xi) = \mathscr{V}_{1n}(\beta_1, \pi) + \mathscr{V}_{2n}(\beta_2, \pi), \tag{3}$$

$$\mathscr{V}_{1n}(\beta_1, \pi) = s_n^2(\xi)\frac{1}{n}\sum_{t=1}^{[n\pi]} \rho_2\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right) \quad \text{and}$$

$$\mathscr{V}_{2n}(\beta_2, \pi) = s_n^2(\xi)\frac{1}{n}\sum_{t=[n\pi]+1}^{n} \rho_2\left(\frac{r_t(\beta_2)}{s_n(\xi)}\right), \tag{4}$$

where $r_t(\beta) = Y_t - X_t'\beta$ is the residual function and, for $j = 1, 2$, $\mathscr{V}_{jn}(\beta_j, \pi)$ represents the residuals τ-scale, pre-$[n\pi]$ and post-$[n\pi]$, respectively. Finally, $s_n(\xi)$ denotes the corresponding M-scale, such that

$$\frac{1}{n}\sum_{t=1}^{[n\pi]} \rho_1\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right) + \frac{1}{n}\sum_{t=[n\pi]+1}^{n} \rho_1\left(\frac{r_t(\beta_2)}{s_n(\xi)}\right) = b \tag{5}$$

with $b = \mathrm{E}_\Phi[\rho_1(v)]$.

As a matter of notation, we let $\{\mathscr{I}_t^j; j = 1, 2\}$ denote a particular time partition, such that $\mathscr{I}_t^1 = I\ (t \leqslant [n\pi])$ and $\mathscr{I}_t^2 = I\ (t > [n\pi])$. For any integer j, the function $\rho_j : \mathbb{R} \to \mathbb{R}$ is such that $\psi_j(z) = \partial\rho_j(z)/\partial z$, $\tilde{\psi}_j(z) = z\psi_j(z)$, $\tilde{\dot{\psi}}_j(z) = z\dot{\psi}_j(z)$, $\tilde{\tilde{\dot{\psi}}} = z^2\dot{\psi}_j(z)$ and $\tilde{\rho}_j(z) = 2\rho_j(z) - \tilde{\psi}_j(z)$. Let $\|\cdot\|$ denote the Euclidean norm of a vector or a matrix and $\|\cdot\|_r$ the L_r norm of a random q-vector (i.e., $\|X\|_r = (\sum_{i=1}^q \mathrm{E}|X_i|^r)^{1/r}$). The symbol "$\to^{\mathrm{p}}$" represents convergence in probability, "$\to^{\mathrm{d}}$" convergence in distribution and "⇒" weak convergence in the space $D[0,1]$ under the Skorokhod metric (see, e.g. Pollard, 1984).

Assuming that the objective function is twice differentiable, we can define, for each π, the partial regression parameter estimators, pre-$[n\pi]$ and post-$[n\pi]$, from the first-order conditions (Proposition A.1 in the appendix will be used). Thus, the estimator $\hat{\xi}_n$, outlined by Eq. (2), can be derived in two steps. First, for each possible partition $\pi \in \Pi, \hat{\beta}_n(\pi) = (\hat{\beta}'_{1n}(\pi), \hat{\beta}'_{2n}(\pi))'$ is obtained as

$$\sum_{t=1}^{n} \left(W_n(\hat{\beta}_n(\pi), \pi)\psi_1 \left(\frac{r_t(\hat{\beta}_{jn}(\pi))}{s_n(\hat{\beta}_n(\pi), \pi)} \right) + \psi_2 \left(\frac{r_t(\hat{\beta}_{jn}(\pi))}{s_n(\hat{\beta}_n(\pi), \pi)} \right) \right) X_t \mathscr{I}_t^j = 0, \quad j = 1,2 \tag{6}$$

with $W_n(\xi)$ defined by (A.3). In a second step, the estimator $\hat{\pi}_n$ will be derived as a global minimizer of the objective function, such that $\hat{\pi}_n = \arg\min_{\pi \in \Pi} \{\mathscr{V}_{1n}(\hat{\beta}_{1n}(\pi), \pi) + \mathscr{V}_{2n}(\hat{\beta}_{2n}(\pi), \pi)\}$. The τ-estimator of the regression parameters will be $\hat{\beta}_n = \hat{\beta}_n(\hat{\pi}_n)$ and the size of the jump is estimated by $\hat{\lambda}_n = \hat{\beta}_{1n} - \hat{\beta}_{2n}$.

Next, we outline the temporal dependence structure we assume in our model, the '*Near Epoch Dependence*' (NED), which constitutes a plausible alternative to the widespread mixing process. It is well known that the strong restrictions needed to ensure processes are mixing threaten to limit the usefulness of the mixing concept (see, e.g., Davidson 1994, Chapter 14). Only specific aspects of mixing, encapsulated in the concept of a mixingale, are required for main limit results to hold. And the key factor of a NED process consists on making use of this fact: although it may not be mixing, it will be 'approximately' mixing in the sense of being well approximated by the near epoch of a mixing process, permitting the application of limit theorems, of which the mixingale property is the most important. It includes linear processes, strong mixing processes and many other dependent structures as special cases. This concept was introduced by Ibramigov (1962), and has been formalized in different ways by Billingsley (1968), McLeish (1975a, b), Bierens (1981), Andrews (1988), Wooldridge and White (1988), Hansen (1991) and Pötscher and Prucha (1991), among others. This is defined below:

Definition 3. Let $\{V_t\}_{-\infty}^{\infty}$ be a strong mixing sequence, possibly vector-valued, on a probability space $(-, \mathscr{G}, P)$ and, define $\mathscr{G}_{t-m}^{t+m} = \sigma(V_{t-m}, \ldots, V_{t+m})$, such that $\{\mathscr{G}_{t-m}^{t+m}\}_{m=0}^{\infty}$ is an increasing sequence of σ-subfields of $\mathscr{G}$. For $r \geqslant 0$, a sequence of integrable random vectors $\{W_t\}_{-\infty}^{\infty}$ is said to be L_r-NED of size $-q_0$ on the strong mixing base $\{V_t\}$ of size $-q_1$ if there exists a sequence of nonnegative constants $\{d_t\}_1^{\infty}$ and a nonnegative sequence $\{v_m\}_0^{\infty}$, such that $v_m \to 0$ as $m \to 0$, and,

(i) for $r = 0$, $\Pr(\|W_t - \mathrm{E}[W_t \mid \mathscr{G}_{t-m}^{t+m}]\| > \varepsilon) \leqslant d_t v_m \ \forall \varepsilon > 0$,
(ii) for $r > 0$, $\|W_t - \mathrm{E}[W_t \mid \mathscr{G}_{t-m}^{t+m}]\|_r \leqslant d_t v_m$,

hold for all $t \geqslant 1$ and $m \geqslant 0$. Besides, $v_m = \mathrm{O}(m^{-q})$ for all $q > q_0$ and $\{\alpha_m\}_{m \geqslant 0}$, the sequence of the strong mixing number of $\{V_t\}$, is such that $\alpha_m = \mathrm{O}(m^{-q})$ for all $q > q_1$.

Finally, the following set of assumptions outlines the setup under which the asymptotic properties of estimators will be derived.

A.1 Assumptions on ρ_1 and ρ_2.

A.1.1. For $j=1,2$, let $\rho_j:\mathbb{R}\to\mathbb{R}$ be a function that satisfies the following properties: (i) $\rho_j(0)=0$. (ii) $\rho_j(-u)=\rho_j(u)$. (iii) $0\leqslant u\leqslant v$ implies that $\rho_j(u)\leqslant\rho_j(v)$. (iv) ρ_j is even and twice continuously differentiable. (v) Given $a_j=\sup\rho_j(u)$, then $0<a_j<\infty$. (vi) There exists a constant m such that $\rho_j(u)$ is constant for $|u|>m$. (vii) If $\rho_j(u)<a_j$ and $0\leqslant u<v$, then $\rho_j(u)<\rho_j(v)$.

A.1.2. Let $b=\mathrm{E}_\Phi[\rho_1(u)]$, then $(b/a_1)=0.5$ holds for a_1 defined by A1.1-(v).

A1.3. ρ_2 satisfies that $2\rho_2(u)-\psi_2(u)u\geqslant 0$.

A2. Model assumptions.
Given $(\theta,s)\in\Theta\times[h_1,h_2]$, $0<h_1,h_2<\infty$, define the sequences $\{\eta_{jt}(\theta;s)\}_{t\leqslant n}$, for $j=1,2$, where $\eta_{jt}(\theta;s)=\psi_j(s^{-1}(U_t+\theta'X_t))X_t$, for each t. Let $\eta_{jt}=\eta_{jt}(0_p;\sigma_0)$, $\forall t\leqslant n$, where 0_p is a p-vector of zeroes and σ_0 is such that $\lim_{n\to\infty}n^{-1}\sum_{t=1}^{n}\mathrm{E}[\rho_1(U_t/\sigma_0)]=b$, with b defined by A1.2. The subscript t of these sequences indicates the dependency on the data $\{Z_t\}$ and θ could be dependent on n, in which case it will be denoted by θ_n. The index $j=1,2$ will be used throughout.

A2.1. $\Theta\subset\mathbb{R}^p$ is a compact and convex set.

A2.2. $\pi_0\in\Pi$, provided that Π has closure in $(0,1)$.

A2.3. (i) $\{Z_t=(Y_t,X_t')'\}_{t\leqslant n}$ is a random vector with domain in Z, L_0-*NED* on a strong mixing base $\{w_t$: $t=\ldots,0,1,\ldots\}$ with constant $d_t=1$, where Z is a Borel subset of $\mathbb{R}^{p+1}$ defined on the probability space $(\text{-},\mathscr{F},P)$. Let $\mathscr{F}_n=n^{-1}\sum_1^n\mathscr{F}(Z_t)$, such that $\{\mathscr{F}_n\}_{n\geqslant1}$ is tight on Z. (ii) X_t is L_2-bounded and such that $\lim_{n\to\infty}n^{-1}\sum_{t=1}^{n-1}\sum_{m=1}^{n-t}\|X_tX_{t+m}'\|<\infty$.

A2.4. For some $r>2,\{\eta_{jt}\}_{t\leqslant n}$ is a random vector sequence of mean zero, L_2-*NED* of size $-\frac{1}{2}$ on a strong mixing base $\{w_{jt}$: $t=\ldots,0,1,\ldots\}$ of size $-r/(r-2)$, with constants $d_{jt}=1$ and $\sup_{t\leqslant n}\mathrm{E}\|\eta_{jt}\|^r<\infty$.

A2.5. (i) $\sup_{t\leqslant n}|\mathrm{E}[\sup_{s>0}\psi_j(s^{-1}U_t)|\mathscr{F}_t^x]|=0$, where $\mathscr{F}_t^x=\sigma(-\infty,\ldots,X_t)$. (ii) $\forall(\theta,s)\in\Theta\times[h_1,h_2],\eta_{jt}(\theta;s)$ is Borel measurable in Z_t, such that $\dot\eta_{jt}(\theta;s)=\partial\eta_{jt}(\theta;s)/\partial\theta'$, continuous in $(Z_t,\theta,s)\in Z\times\Theta\times[h_1,h_2]$ by A1.1-(iv), satisfies that $\sup_{t\leqslant n}\mathrm{E}[\sup_{\theta\in\Theta}\|\dot\eta_{jt}(\theta;\sigma_0)\|^{1+\varepsilon}]<\infty$, for some $\varepsilon>0$.

A2.6. There exists a b_0 such that, for all $b>b_0$, the smallest eigenvalues of the positive matrices $b^{-1}\sum_{t=[n\pi_0]+1}^{[n\pi 0]+b}\dot\eta_{jt}(\theta;\sigma_0)$ and $b^{-1}\sum_{t=[n\pi_0]-b}^{[n\pi_0]}\dot\eta_{jt}(\theta;\sigma_0)$ are bounded away from zero uniformly in $\theta\in\Theta$.

A2.7. Define $M(\theta)=\lim_{n\to\infty}n^{-1}\sum_{t=1}^{n}\mathrm{E}[W_0\dot\eta_{1t}(\theta;\sigma_0)+\dot\eta_{2t}(\theta;\sigma_0)]=W_0M_1(\theta)+M_2(\theta)$, where, for $j=1,2$, $M_j(\theta)=\lim_{n\to\infty}n^{-1}\sum_{t=1}^{n}\mathrm{E}[\dot\eta_{jt}(\theta;\sigma_0)]$ is a positive definite matrix $\forall\theta\in\Theta$ and $W_0=\lim_{n\to\infty}(\sum_{t=1}^{n}\mathrm{E}[\tilde\rho_2(U_t/\sigma_0)]/\sum_{t=1}^{n}\mathrm{E}[\tilde\psi_1(U_t/\sigma_0)])$. The $\lim_{n\to\infty}n^{-1}\sum_{t=1}^{[n\pi]}\mathrm{E}[\dot\eta_{jt}(\theta;\sigma_0)]$ exists uniformly in $(\theta,\pi)\in\Theta\times\Pi$ and equals $\pi M_j(\theta)$, $\forall(\theta,\pi)\in\Theta\times\Pi$. Finally, let $M=M(0_p)$.

A2.8. Let $\{q_t(\theta,s)=q(s^{-1}(U_t+\theta'X_t))\}_{t\leqslant n}$ be a sequence defined for a general function $q(\cdot,\cdot)$, $\forall(\theta,s)\in\Theta\times[h_1,h_2]$, $0<h_1,h_2<\infty$. Then, for the particular cases of $q(\cdot)=\rho_j(\cdot);\tilde\psi_j(\cdot);\tilde{\tilde\psi}_j(\cdot)$, it holds that $\lim_{n\to\infty}n^{-1}\sum_{t=1}^{[n\pi]}\mathrm{E}[q_t(\theta,s)]$

exists uniformly in $(\theta, s, \pi) \in \Theta \times [h_1, h_2] \times \Pi$ and equals $\pi H_q(\theta, s)$, such that $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathrm{E}[q_t(\theta, s)] = H_q(\theta, s) < \infty$. Similarly, for the functions $q(\cdot) = \psi_j(\cdot); \tilde{\dot{\psi}}_j(\cdot)$, we assume that $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{[n\pi]} \mathrm{E}[q_t(\theta, \sigma_0) X_t] = \pi H_q^X(\theta)$, uniformly in $(\theta, \pi) \in \Theta \times \Pi$, such that $H_q^X(\theta) = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathrm{E}[q_t(\theta, \sigma_0) X_t]$, finite $\forall \theta \in \Theta$. Finally, $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \mathrm{E}[\sup_{\theta\in\Theta; s\in[h_1;h_2]} |q_t(\theta; s)|^{1+\varepsilon}] < \infty$ holds for all above mentioned function.

A2.9. Given $S_{12} = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{E}[\eta_{1t}\eta_{2s}']$, define $S = \lim_{n\to\infty} var[n^{-1/2} \sum_{t=1}^{n} W_0\eta_{1t} + \eta_{2t}] = W_0^2 S_1 + S_2 + 2W_0 S_{12}$, where, for $j = 1, 2$, $S_j = \lim_{n\to\infty} var[n^{-1/2} \sum_{t=1}^{n} \eta_{jt}]$ is a finite and positive $p \times p$ matrix. $\forall \pi \in (0, 1)$ it holds that $\lim_{n\to\infty} var[n^{-1/2} \sum_{t=1}^{[n\pi]} \eta_{jt}] = \pi S_j$.

A2.10. Let $\lambda_n \to 0$ with $n\|\lambda_n\|^2 \to \infty$.

Assumption A1 is standard in the τ-estimator context and guarantees robustness and efficiency under Gaussianity. Thus, we obtain a estimation procedure which assures simultaneously stability under any possible type of contamination and high relative efficiency with respect to LS when the latter is optimal. In particular, condition A1.l determines the functional form of ρ_j, for $j = 1, 2$, differentiable, symmetric, bounded and monotone increasing. A1.2. is a restrictive assumption on ρ_1, under which the highest breakdown point is guaranteed (Yohai and Zamar 1988, Theorem 3.1). As it will be explained in Remark 2 (Section 3), efficiency under normality is obtained by A1.3, a requirement imposed on ρ_2. Moreover, this assumption will imply that $W_n(\xi)$, defined by (A.3), is nonnegative. As a consequence, the τ-estimator of regression parameters, in (6), can be viewed as an M-estimator with an adaptive ψ-function, let $\psi_n(u) = W_n(\xi)\psi_1(u) + \psi_2(u)$, a weighted average of ψ_1 and ψ_2 with weights depending on the data. A similar result is obtained by Yohai and Zamar (1988) in the standard context of no change.

Conditions in A2 will permit the development of an asymptotic theory for the model estimators. In particular, the assumption of a bounded parameter set in A2.1 and A2.2 is restrictive, although it may not be of any practical significance. A2.3 and A2.4 establish the weak temporal dependence structure in this specific robust regression context. It is noteworthy that a suitable transformation of a NED process will also inherit a NED structure dependence (see, e.g., Davidson 1994, Chapter 17). Therefore, under smoothness conditions on ρ_j, as imposed by A1.1, we may base (and state) the NED property on the temporal behavior of regressors and/or the perturbance term allowing to establish the dependence conditions in more primitive terms. For example, a model either with independent regressors and a linear process for the error term or with both regressors and perturbance term following an asymptotic dependence structure of a strong mixing process, pertain to the class of processes considered here. Note that condition A2.5-(i) is quite restrictive, although it is also required in the standard context of no change. For *i.i.d.* observations, it would imply a symmetric distribution for the error term. Assumption A2.5-(ii) is imposed given that the break point is estimated by a global searching of a minimizer for the objective function. On the other hand, its identification will be guarantee under A2.6, assuming that there exist enough observations near it. Assumptions A2.7–A2.9 are standard requirements for obtaining asymptotic covariance stationarity for the estimated parameters. However, they rule out models with lagged-dependent variables as regressors when associated coefficients are subject to change, which can be especially interesting in dynamic models. But in this case, the conditions could be weaken and the results of theorems surely hold (for interested readers, a detailed discussion is given by Fiteni 2002, Remark 4). Finally, the reason for A2.10 is that with a fixed break, its location estimator will have an asymptotic distribution which depends upon nuisance parameters, and thus not useful for related inferences. By letting λ_n tend towards zero, a simpler asymptotic pivotal distribution will be found, as it can be seen in the next section.

3. Asymptotic properties of the estimators

In this section, we provide the asymptotic behavior of τ-estimators in a structural change regression model, defined by (2)–(4). First, we consider the corresponding rates of convergence, in Theorem 1, which will allow us to derive the asymptotic distribution of the estimators, our main result provided by Theorem 2.

Theorem 1. *Under* Al *and* A2.1–A2.8, *it holds that* $(\hat{\beta}_{jn}-\beta_{j0})=\mathrm{Op}(n^{-1/2})$, *for* $j=1,2$ *and* $(\hat{\pi}_n-\pi_0)=\mathrm{Op}(n^{-1}\|\lambda\|^{-2})$.

It is convenient to remark that the rate of convergence corresponding to the regression coefficients estimators are, as usual, $n^{-1/2}$, as if the true break location were known. The break point estimator converges to the true break point, at a rate depending on λ, the size of the jump. This dependence allows us to incorporate two standard settings about λ: fixed and asymptotically decreasing (at a rate given by A2.10). For the latter, it will be denoted by λ_n. Therefore, the rate of convergence corresponding to $\hat{\pi}_n$ will be $\mathrm{Op}(n^{-1})$ when the break is constant and $\mathrm{Op}(n^{-1}\|\lambda_n\|^{-2})$ when a local change. This result is well known for any other estimation procedure. The τ-estimator of the break fraction remains consistent, in any case, even when the shift is asymptotically decreasing with the sample size, by A2.10. Given the above rates of convergence, we are in the position to prove the limiting distribution of the model estimators, in the next theorem.

Theorem 2. *Under* Al *and* A2.1–A2.9, *it holds that*,

(i)

$$\begin{bmatrix} \sqrt{n}(\hat{\beta}_{1n}-\beta_{10}) \\ \sqrt{n}(\hat{\beta}_{2n}-\beta_{20}) \end{bmatrix} \xrightarrow{\mathrm{d}} M^{-1}S^{1/2}\begin{bmatrix} \tau_0^{-1/2}Z_p & 0_{p\times p} \\ 0_{p\times p} & (1-\tau_0)^{-1}Z_p \end{bmatrix}, \tag{7}$$

where Z_p *represents the* p*-dimensional standard Gaussian vector and* $0_{p\times p}$ is a $(p\times p)$*-matrix of zeroes.*

(ii) *Assuming* A2.10,

$$\frac{(\lambda_n' M\lambda_n)^2}{\lambda_n' S\lambda_n}\, n(\hat{\pi}_n-\pi_0) \;\Rightarrow\; \arg\max_w\{W(w)-\tfrac{1}{2}|w|\}, \tag{8}$$

where $W(\cdot)$ represents an independent two-sided standard Brownian motion defined in $\mathbb{R}$.

(iii) *Assuming λ constant,*

$$n(\hat{\tau}_n - \tau_0) \;\Rightarrow\; \arg\max_w \{\lambda' W^*(w) - \tfrac{1}{2}\lambda' M(\lambda)\lambda |w|\}, \tag{9}$$

where $W^*(\cdot)$ represents a process defined in $\mathbb{Z}$, the integer set, such that,

$$W^*(w) = \left\{ \begin{array}{ll} 0, & w = 0 \\ s_n^0 \sum_{t=w}^{-1} \eta_t\,(0_p, s_n^0), & w = -1, -2, \ldots \\ s_n^0 \sum_{t=1}^{w} \eta_t\,(0_p, s_n^0), & w = 1, 2, \ldots \end{array} \right\} \tag{10}$$

with s_n^0 defined by $n^{-1}\sum_{t=1}^n \rho_1(U_t/s_n^0) = b$.

(iv) *The distribution of $\sqrt{n}((\hat{\beta}_{1n} - \beta_{10})', (\hat{\beta}_{2n} - \beta_{20})')'$ and that of $n(\hat{\pi}_n - \pi_0)$ are asymptotically independent for the two cases of λ.*

Again, the estimated regression parameters have a standard limiting distribution, as if the shift point location were known. As usual, the limiting distribution of the break date estimator depends upon nuisance parameters when considering constant the jump size. The difficulty is due to the Op(n^{-1}) rate of convergence. Only when setting an asymptotic framework where the magnitude of the shift converges to zero, we reduce the rate of convergence and a pivotal limiting distribution is found. As pointed by Bai (1994), when the sample size increases, because λ_n converges to zero, more observations in a neighborhood of the true shift point are needed to discern the shift point so that the Central Limit Theorem eventually applies (though the size of the neighborhood will increase at a rate $\|\lambda_n\|^{-2}$, much slower than the sample size). This is why a Brownian motion is embedded in the limiting process. In particular, it is characterized by $W(w)$, a two-sided Brownian motion, given by $W(w) = W_1(-w)I(w<0) + W_2(w)I(w \geqslant 0)$, with $W_1(w)$ and $W_2(w)$ representing two independent standard Brownian processes. The specific distribution function of $\arg\max_w\{W(w) - |w|/2\}$ is given by

$$F(t) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\sqrt{t}\mathrm{e}^{-1/st} + \frac{3}{2}\,\mathrm{e}^t \Phi\left(-\frac{3}{2}\sqrt{t}\right) - \left(\frac{1}{2}\,t + \frac{5}{2}\right)\Phi\left(-\frac{1}{2}\sqrt{t}\right) \tag{11}$$

for $t > 0$, we can see, e.g., in Bai (1994) or Fiteni (2002).

Finally, considering part (i) of Theorem 2, we derive, in the next corollary, the asymptotic distribution of τ-estimator corresponding to the jump size.

Corollary 1. *Under* A1 *and* A2,

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{\mathrm{d}} (\tau_0(1-\tau_0))^{-1/2} M^{-1} S^{1/2} Z_p.$$

In the next section, we focus on the implementation in practice of this type of estimator by means of simulation experiments. For this study, a particular τ-estimator based on the bisquare family of ψ-functions, will be used. This is described below.

Remark 2. As we can see in Theorem 3.1 by Yohai and Zamar (1988), the high breakdown-point property depends only on ρ_1. Therefore, ρ_2 will be chosen so that we obtain high efficiency under Gaussianity. We consider a particular τ-estimator corresponding to the bisquare family of ψ-functions, which satisfies the required conditions A1. This is defined as follows:

$$\psi_{B,c}(u) = \begin{cases} u(1-\frac{u^2}{c^2})^2 & \text{for } |u| < c, \\ 0 & \text{for } |u| \geqslant c \end{cases}$$

for a given constant c. The corresponding ρ-function will be

$$\rho_{B,c}(u) = \begin{cases} \frac{u^2}{2}(1-\frac{u^2}{c^2}+\frac{u^4}{3c^4}) & \text{for } |u| < c, \\ \frac{c^2}{6} & \text{for } |u| \geqslant c. \end{cases}$$

Condition A1.1 holds for any value of c. If we take $\rho_1 = \rho_{B,c_1}$, such that $c_1 = 1.56$ and $b = \mathrm{E}_{\Phi}[\rho_{B,c_1}(u)] = 0.203$, A1.2 will be satisfied and the corresponding τ-estimator will have a breakdown point equals 0.5. Requirement A1.3 also holds, given that

$$2\rho_{B,c}(u) - u\psi_{B,c}(u) = \begin{cases} \frac{u^4}{c^2}(1-\frac{2}{3}\,\frac{u^2}{3c^2}) & \text{for } |u| < c, \\ \frac{c^2}{3} & \text{for } |u| \geqslant c \end{cases}$$

and $\rho_2 = \rho_{B,c_2}$, with $c_2 = 6.08$, such that $\mathrm{E}_{\Phi}[\dot{\psi}_0(u)]^2/\mathrm{E}_{\Phi}[\psi_0^2(u)] = 0.95$. The resulting estimator will have simultaneously maximum breakdown point and a relative efficiency of 0.95 at the normal model.

4. Monte Carlo experiment

This simulation study focuses on the relative finite sample performance of robust estimators of structural change models under different distributional scenarios for the regressors. As illustrated in Remark 1, the behavior of existing estimators of a structural break may be altered by the presence of outliers in the regressors space. Now, we contemplate the possibility of regressors distributions with thick tails (see footnote 1). Under outlying observations, nonprotected estimators could spoil the estimates, misplacing the break location. Under thick-tailed distributions, they will be also inefficient, as will be illustrated below.

Again, four estimators will be considered: (a) the LS estimator is the most efficient under Gaussian distributions, but it gets the lowest breakdown point under any type of contamination. (b) the LAD estimator is the most robust under contaminated distributions for the perturbance term, but it is highly inefficient under Gaussianity and its breakdown point equals zero under leverage. (c) the Huber estimator is a particular

M-estimator with a related ψ-function given by $\psi(u)=cu\min\{|u|/c,1\}/|u|$, for a suitable constant c. It sets up an intermediate solution between LAD and LS estimators, more efficient than the first and more robust than the second, but only against symmetric heavy-tailed error distributions, because its breakdown point equals zero when the carriers are also contaminated. Lastly, (d) the τ-estimator is qualitatively robust, with maximum breakdown point under contaminated distributions for both the regressors and the error term, and highly efficient under Gaussianity.

These four estimators will be compared in the following simulation study. Data are generated according to the model:

$$Y_t = 1 + X_t + I(t/n > 0.5) + U_t, \quad t = 1,\dots,n \tag{12}$$

such that $\xi_0=(\beta'_{10},\beta'_{20},\pi_0)'=(1,1,2,1,0.5)'$, $U_t \sim \mathrm{N}(0,1)$ and $X_t \sim i.i.d.F(X)$. Our goal is to illustrate the potential gain of using τ-estimators in the presence of leverage. Then, we generate model (12) under different distributional scenarios F: standard normal, double exponential, t_3, t_5 and two mixed normal distributions, such that $F(X) = (1-\varepsilon)\Phi(X) + \varepsilon\Phi(\sigma X)$, with $\sigma = 3$ and $\varepsilon = 0.1$ and 0.25, which will be denoted by $N90$ and $N75$, respectively. The regressor is standardized for comparative purposes, in order to get a variance equals one for all the cases.

Under each of above distributions F and from 2000 repetitions, we estimate bias and mean square error (MSE) corresponding to LS, LAD, Huber and τ-estimators of ξ_0. The computed Huber estimator is scale-invariant, considering the median absolute deviation (MAD) as the scale estimator and the constant $c = 1.345$, according to the minimax version (see Huber, 1981). The computed τ-estimator belongs to the family of bisquare functions, with the constant c_1 and c_2 defined as in Remark 2.

Programs are written in FORTRAN90, Double Precision and IMSL routines were used for generating random numbers. We have applied the algorithm designs proposed by Koenker and D'Orey (1987) for the LAD estimation, Huber and Dutter (1974) and Huber (1977) for the Huber estimator and Yohai and Zamar (1988) for the τ-estimator. All of them have been adapted for this structural change estimation context. For interested readers, programs will be provided under request.

Tables 2 and 3, in the last section, show the results. In Table 2, we present bias and MSE for the break location estimator, which are as expected. Under all (but the Gaussian) distributions F, the τ-estimator appears to be the most efficient in terms of MSE and for all n. At the normal model, however, the LS estimator obtains estimates with the lowest spread. In fact, it provides the maximum likelihood estimation in this case. Simulation evidence also confirms that inefficient results will be provided by the LAD estimator in this context. It yields the largest spread estimates under all the distributional scenarios and for any sample size. Lastly, the Huber estimator produces an intermediate solution, in terms of efficiency, between LS and LAD, for all n and any of the distributional scenario we have considered for the regressor term.

Estimated bias and MSE corresponding to the regression coefficient estimators pre- and post-break are reported in Table 3 (only t_5 and double exponential distributional scenarios are shown in order to save space; remaining cases are similar and available under request). As expected, their behavior is similar to that of the break location estimator for each regressor distribution. The τ-estimator arises as the most efficient for

Table 2
Break estimator

Point estimation		$n=50$		$n=100$	
Model	Estimators	Bias	MSE	Bias	MSE
	LS	−0.369	**1.301**	0.159	**1.230**
	MAD	−0.030	1.941	0.113	2.255
N(0, 1)	Huber	−0.328	1.672	0.269	1.515
	τ-est.	−1.350	2.152	0.030	1.366
	LS	0.169	2.337	−0.235	1.589
	MAD	0.068	3.005	−0.220	2.255
$\frac{1}{2}\exp(-\lvert x\rvert)$	Huber	0.069	2.466	−0.175	1.515
	τ-est.	−0.183	**2.275**	−0.090	**1.366**
	LS	0.385	2.416	−0.621	1.686
	MAD	−0.540	2.999	−0.513	2.327
t_3	Huber	0.163	2.525	−0.538	1.834
	τ-est.	−0.640	**2.277**	−0.574	**1.544**
	LS	0.257	2.312	−0.196	1.567
t_5	MAD	−0.248	2.957	−0.012	2.269
	Huber	−0.286	2.355	−0.192	1.629
	τ-est.	−0.366	**2.125**	−0.072	**1.447**
	LS	−0.112	2.363	−0.501	1.556
	MAD	−0.390	2.995	−0.165	2.176
$N90$	Huber	0.188	2.409	−0.113	1.613
	τ-est.	−0.128	**2.189**	−0.401	**1.463**
	LS	−0.086	2.371	−0.617	1.519
	MAD	−0.040	2.981	−0.302	2.159
$N75$	Huber	0.125	2.446	−0.175	1.575
	τ-est.	−0.128	**2.188**	−0.361	**1.468**

Bias and mean squared error (MSE) for LAD, Huber and τ-estimators (2000 replications). Model: $Y_t = 1 + X_t + I(t/n > \tau_0) + U_t$, with $t = 1,\dots,n$, where $U_t \sim \mathrm{N}(0,1)$ and $X_t \sim F(X)$, such that $F = \mathrm{N}(0,1), \frac{1}{2}\exp(-|x|), t_3, t_5, N90$ and $N75$. Bold indicates the lowest MSE. Values must be divided by 10^3.

every defined thick-tailed distribution and the LAD estimates performs comparatively rather badly. For each estimation procedure, considered distributional scenarios give estimates with similar spread, which becomes narrower when increasing the sample size.

We conclude noting that the consideration of robust estimators for structural change models can be of special usefulness in practice if we suspect the presence of outliers in the data or, more generally, possibly thick-tailed distributions for either the regressors or the perturbance term. Using protected estimators in this context, we could gain in terms of efficiency and furthermore we can prevent the risk of confusing between an outlier or a structural break point.

Table 3
Coefficient regression estimators

Point estimation		$n = 50$		$n = 100$	
Model	Estimator	Bias	MSE	Bias	MSE
$e^{-\|x\|}/2$					
	LS Intercept pre-break	−5.699	8.781	−4.341	3.518
	Intercept post-break	7.015	9.255	4.196	3.365
	Slope pre-break	0.424	11.33	0.057	4.027
	Slope post-break	−1.050	12.72	0.615	4.371
	LAD Intercept pre-break	−5.947	11.90	−4.793	5.551
	Intercept post-break	6.984	12.53	4.952	5.303
	Slope pre-break	−1.164	15.67	0.260	5.866
	Slope post-break	−0.622	16.99	0.589	6.258
	Huber Intercept pre-break	−6.152	9.685	−4.497	3.552
	Intercept post-break	6.457	9.912	4.452	3.676
	Slope pre-break	0.009	**11.82**	0.191	3.858
	Slope post-break	−0.904	14.78	0.464	4.622
	τ-est. Intercept pre-break	−5.273	**8.794**	−4.244	**3.349**
	Intercept post-break	5.097	**8.241**	3.767	**3.195**
	Slope pre-break	0.586	12.02	0.361	**3.517**
	Slope post-break	−0.540	**11.24**	0.691	**3.820**
t_3					
	LS Intercept pre-break	−6.310	9.391	−4.864	3.928
	Intercept post-break	7.292	9.443	2.994	3.409
	Slope pre-break	0.138	17.90	0.002	5.126
	Slope post-break	−1.434	16.90	0.642	4.548
	LAD Intercept pre-break	−7.178	13.06	−5.603	5.893
	Intercept post-break	7.117	12.92	3.404	5.406
	Slope pre-break	−0.266	22.51	−0.252	7.899
	Slope post-break	−1.536	21.19	0.847	7.220
	Huber Intercept pre-break	−6.572	9.703	−4.972	4.249
	Intercept post-break	7.229	9.884	3.189	3.718
	Slope pre-break	0.176	18.42	0.044	5.417
	Slope post-break	−1.386	16.97	0.737	4.970
	τ-est. Intercept pre-break	−5.964	**8.582**	−4.413	**3.703**
	Intercept post-break	5.582	**8.621**	2.541	**3.388**
	Slope pre-break	0.842	**15.06**	0.054	**4.921**
	Slope post-break	−1.049	**13.51**	0.980	**4.543**

Bias and mean squared error (MSE) for LAD, Huber and τ-estimators (2000 replications). Model: $Y_t = 1 + X_t + I(t/n > \tau_0) + U_t$, with $t = 1, \ldots, n$, where $U_t \sim \mathrm{N}(0,1)$ and $X_t \sim F(X)$, such that $F = \mathrm{N}(0,1), \frac{1}{2}\exp(-|x|), t_3, t_5, N90$ and $N75$. Bold indicates the lowest MSE. Values must be divided by 10^3.

Acknowledgements

I am grateful to my thesis advisors, Miguel A. Delgado and Javier Hidalgo, for their support and suggestions. The research is funded by a postdoctoral fellowship from Comunidad de Madrid and European Union, No. 4323/2000 and by Dirección General de Enseñanza Superior, No. BEC2001-1270. This article is developed from a chapter of my doctoral dissertation at the Universidad Carlos III de Madrid.

Appendix A.

We shall consider the case of $\pi \leqslant \pi_0$, without loss of generality because of symmetry. Limits are taken as n, the sample size, increases to infinity. For notational convenience, we establish that $[n\pi] = k$, $[n\pi_0] = k_0$ and $\sum_{t=i}^{j}$ will be denoted by $\sum_{i}^{j}$. Also, two alternative sample partitions will be used: the first, $\{\mathscr{I}_t^j; j = 1,2\}$, already defined in Section 2, and the second, $\{I_t^h; h = 1,2,3\}$, with $I_t^1 = \mathscr{I}_t^1$, $I_t^2 = I\ (k < t \leqslant k_0)$ and $I_t^3 = I\ (t > k_0)$, such that $I_t^2 + I_t^3 = \mathscr{I}_t^2$. Finally, we establish $e_h = \mathrm{E}[I_t^h]$, for $h = 1,2,3$, and then, $e_1 = \pi$, $e_2 = \pi_0 - \pi$ and $e_3 = 1 - \pi_0$. The subscript n of the estimators will be omitted.

Proposition A.1. *Let the function $\mathscr{V}_n(\xi)$ be defined by (4) and (5). Then, for $j=1,2$,*

$$\frac{\partial \mathscr{V}_n(\xi)}{\partial \beta_j} = -s_n(\xi)\frac{1}{n}\sum_{t=1}^{n}\left(W_n(\xi)\psi_1\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right) + \psi_2\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\right)X_t\mathscr{I}_t^j \tag{A.1}$$

and

$$\frac{\partial^2 \mathscr{V}_n(\xi)}{\partial \beta_j \partial \beta_j'} = \frac{1}{n}\sum_{t=1}^{n}\left(W_n(\xi)\dot{\psi}_1\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right) + \dot{\psi}_2\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\right)X_tX_t'\mathscr{I}_t^j + W_{jn}(\xi), \tag{A.2}$$

where

$$W_n(\xi) = D_n^{-1}(\xi)\frac{1}{n}\sum_{j=1}^{2}\sum_{t=1}^{n}\tilde{\rho}_2\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\mathscr{I}_t^j,$$

$$W_{jn}(\xi) = \dot{s}_{jn}(\xi)\tilde{\gamma}_{jn}'(\xi) + \gamma_{jn}(\xi)\dot{s}_{jn}'(\xi) + \dot{s}_{jn}(\xi)\tilde{\gamma}_n(\xi)\dot{s}_{jn}'(\xi) \tag{A.3}$$

with

$$\gamma_{jn}(\xi) = \frac{1}{n}\sum_{t=1}^{n}\Upsilon_n\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)X_t\mathscr{I}_t^j, \qquad \tilde{\gamma}_{jn}(\xi) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\Upsilon}_n\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)X_t\mathscr{I}_t^j,$$

$$D_n(\xi) = \frac{1}{n}\sum_{j=1}^{2}\sum_{t=1}^{n}\tilde{\psi}_1\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\mathscr{I}_t^j,$$

$$\tilde{\gamma}_n(\xi) = \frac{1}{n}\sum_{j=1}^{2}\sum_{t=1}^{n} \tilde{\Upsilon}_n\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\frac{r_t(\beta_j)\mathscr{I}_t^j}{s_n(\xi)}$$

and

$$\dot{s}_{jn}(\xi) = \frac{\partial s_n(\xi)}{\partial \beta_j} = -\frac{1}{n}\sum_{t=1}^{n}\psi_1\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\frac{X_t\mathscr{I}_t^j}{D_n(\xi)},$$

such that $\Upsilon_n(z) = W_n(\xi)\Psi_1^-(z) - \Psi_2^-(z)$, $\tilde{\Upsilon}_n(z) = W_n(\xi)\Psi_1^+(z) - \Psi_2^-(z)$ *and* $\Psi_j^{\pm}(z) = \psi_j(z) \pm \tilde{\dot{\psi}}_j(z)$.

Proof. Consider $j = 1$. Then, differentiating (5) we get

$$\begin{aligned}\dot{s}_{1n}(\xi) &= -\left(\frac{1}{n}\sum_{j=1}^{2}\sum_{t=1}^{n}\tilde{\psi}_1\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\mathscr{I}_t^j\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}\psi_1\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)X_t\mathscr{I}_t^1\\ &= \frac{1}{n}\sum_{t=1}^{n}\psi_1\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)\frac{X_t\mathscr{I}_t^1}{D_n(\xi)}\end{aligned} \tag{A.4}$$

and result (A.1) follows after some operations. Now, consider the second derivative,

$$\begin{aligned}\frac{\partial^2\mathscr{V}_n(\xi)}{\partial\beta_1\partial\beta_1'} = {}&\frac{1}{n}\sum_{t=1}^{k}\left(W_n(\xi)\dot{\psi}_1\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right) + \dot{\psi}_2\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)\right)X_tX_t'\\ &-\frac{1}{n}\sum_{t=1}^{k}\left(W_n(\xi)\Psi_1^-\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right) + \Psi_2^-\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)\right)X_t\dot{s}_{1n}'(\xi)\\ &- s_n(\xi)\frac{1}{n}\sum_{t=1}^{k}\psi_1\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)X_t\dot{W}_{1n}'(\xi).\end{aligned}$$

Noting that,

$$\begin{aligned}\dot{W}_{1n}(\xi) = \frac{\partial W_n(\xi)}{\partial\beta_1} = {}&\frac{1}{n}\left(\sum_{j=1}^{2}\sum_{t=1}^{n}\tilde{\Upsilon}_n\left(\frac{r_t(\beta_j)}{s_n(\xi)}\right)\frac{r_t(\beta_j)\mathscr{I}_t^j}{s_n(\xi)}\right)\frac{\dot{s}_{1n}(\xi)}{D_n(\xi)}\\ &-\frac{1}{n}\sum_{t=1}^{k}\tilde{\Upsilon}_n\left(\frac{r_t(\beta_1)}{s_n(\xi)}\right)\frac{X_t}{s_n(\xi)D_n(\xi)},\end{aligned}$$

we get (A.2) for $j = 1$. The result for $j = 2$ is obtained similarly. □

Proof of Theorem 1. First, define for a given $k, \beta_t^0(k) = (\beta_{1t}^0{}'(k), \beta_{2t}^0{}'(k))'$, where $\beta_{1t}^0(k) = \beta_{10}I_t^1$ and $\beta_{2t}^0(k) = \beta_{10}I_t^2 + \beta_{20}I_t^3$, such that $\{U_t\}_{t=1}^n = \{r_t(\beta_{1t}^0(k) + \beta_{2t}^0(k))\}_{t=1}^n$. In order to get more clarity in the exposition, the explicit dependence on k which have previous parameters will be omitted throughout, such that $\beta_{jt}^0(k)=\beta_{jt}^0$. Then, from (3) and (4),

$$\mathscr{V}_n(\xi_0)=\frac{s_n^2(\xi_0)}{n}\sum_{t=1}^{n}\rho_2\left(\frac{U_t}{s_n(\xi_0)}\right)$$
$$=\frac{s_n^2(\xi_0)}{n}\left(\sum_{t=1}^{k}\rho_2\left(\frac{r_t(\beta_{1t}^0)}{s_n(\xi_0)}\right)+\sum_{t=k+1}^{n}\rho_2\left(\frac{r_t(\beta_{2t}^0)}{s_n(\xi_0)}\right)\right),$$

by A1.1-(i). From (5), $s_n(\xi_0)$, denoted by s_n^0 from now on, is such that $n^{-1}\sum_{t=1}^{n}\rho_1(U_t/s_n^0)=b$ and then,

$$\frac{1}{n}\sum_{t=1}^{k}\rho_1\left(\frac{r_t(\beta_{1t}^0)}{s_n^0}\right)+\frac{1}{n}\sum_{t=k+1}^{n}\rho_1\left(\frac{r_t(\beta_{2t}^0)}{s_n^0}\right)=b, \tag{A.5}$$

using the same arguments as above. Similarly,

$$W_n(\xi_0)=\frac{\sum_{t=1}^{n}\tilde\rho_2(\frac{U_t}{s_n^0})}{\sum_{t=1}^{n}\tilde\psi_1(\frac{U_t}{s_n^0})}=\frac{\sum_{t=1}^{n}\tilde\rho_2(\frac{r_t(\beta_{1t}^0+\beta_{2t}^0)}{s_n^0})}{\sum_{t=1}^{n}\tilde\psi_1(\frac{r_t(\beta_{1t}^0+\beta_{2t}^0)}{s_n^0})}$$
$$=\frac{\sum_{t=1}^{n}\tilde\rho_2(\frac{r_t(\beta_{1t}^0)}{s_n^0})+\sum_{t=k+1}^{n}\tilde\rho_2(\frac{r_t(\beta_{2t}^0)}{s_n^0})}{\sum_{t=1}^{n}\tilde\psi_1(\frac{r_t(\beta_{1t}^0)}{s_n^0})+\sum_{t=k+1}^{n}\tilde\psi_1(\frac{r_t(\beta_{2t}^0)}{s_n^0})}, \tag{A.6}$$

which will be also denoted by W_n^0. Therefore, by the mean value theorem (MVT) and Proposition A.1, $\mathscr{V}_n(\xi_0)-\mathscr{V}_n(\xi)$ is given by

$$\frac{s_n^0}{n}\sum_{t=1}^{k}(\beta_1-\beta_{1t}^0)'\psi_n^0\left(\frac{r_t(\beta_{1t}^0)}{s_n^0}\right)X_t+\frac{s_n^0}{n}\sum_{t=k+1}^{n}(\beta_2-\beta_{2t}^0)'\psi_n^0\left(\frac{r_t(\beta_{2t}^0)}{s_n^0}\right)X_t \tag{A.7}$$

$$-\frac{1}{2n}\sum_{t=1}^{k}(\beta_1-\beta_{1t}^0)'\left(\dot\psi_n^*\left(\frac{r_t(\beta_{1t}^*)}{s_n^*}\right)X_tX_t'+\frac{n}{k}W_{1n}^*\right)(\beta_1-\beta_{1t}^0) \tag{A.8}$$

$$-\frac{1}{2n}\sum_{t=k+1}^{n}(\beta_2-\beta_{2t}^0)'\left(\dot\psi_n^*\left(\frac{r_t(\beta_{2t}^*)}{s_n^*}\right)X_tX_t'+\frac{n}{n-k}W_{2n}^*\right)(\beta_2-\beta_{2t}^0), \tag{A.9}$$

where, as a matter of notation, for any sequence of functions $g_n:\Theta^2\times\Pi\to\mathbb{R}$, we denote $g_n^*=g_n((\beta_t^{*\prime},\pi)')$, with $\beta_t^*=(\beta_{1t}^{*\prime},\beta_{2t}^{*\prime})'$, such that $\beta_{jt}^*=\beta_{jt}^0+\delta_j(\beta_j-\beta_{jt}^0)$, for $j=1,2$, and $|\delta_j|<1$. Moreover, $\psi_n^0(z)=W_n^0\psi_1(z)+\psi_2(z)$, $\dot\psi_n^*(z)=W_n^*\dot\psi_1(z)+\dot\psi_2(z)$ and $W_{jn}^*=\dot s_{jn}^*\tilde\gamma_{jn}^{*\prime}+\gamma_{jn}^*\dot s_{jn}^{*\prime}+\dot s_{jn}^*\tilde\gamma_n^*\dot s_{jn}^{*\prime}$, for $j=1,2$. Note that $\{r_t(\beta_{1t}^*+\beta_{2t}^*)\}_{t=1}^n=\{U_t+X_t'(\delta_1(\beta_{10}-\beta_1)I_t^1+\delta_2(\beta_{10}-\beta_2)I_t^2+\delta_2(\beta_{20}-\beta_2)I_t^3)\}_{t=1}^n$. Then, s_n^* will be such that $n^{-1}\sum_{h=1}^{3}\sum_{t=1}^{n}\rho_1(s_n^{*-1}(U_t+\theta_h'X_t))I_t^h=b$, denoted by $s_n(\underline{\theta})$ in what follows, with $\underline{\theta}=(\theta_1',\theta_2',\theta_3')'\in\Theta^3$, such that $\theta_1=\delta_1(\beta_{10}-\beta_1)$, $\theta_2=\delta_2(\beta_{10}-\beta_2)$ and $\theta_3=\delta_2(\beta_{20}-\beta_2)$

for this particular case. Similarly,

$$W_n^* = \left(\sum_{h=1}^{3}\sum_{t=1}^{n}\tilde{\psi}_1\left(\frac{U_t+\theta_h' X_t}{s_n^*}\right)I_t^h\right)^{-1}\sum_{h=1}^{3}\sum_{t=1}^{n}\tilde{\rho}_2\left(\frac{U_t+\theta_h' X_t}{s_n^*}\right)I_t^h, \tag{A.10}$$

which will be represented by $W_n(\underline{\theta})$ throughout.

Next, define $M_n^i(j,l,\underline{\theta}) = n^{-1}\sum_{j+1}^{l}\dot{\psi}_n^*((U_t+\theta_i' X_t)/s_n(\underline{\theta}))X_t X_t'$, for $i=1,2,3$, and $N_n(j,l) = n^{-1}\sum_{j+1}^{l}\psi_n^0(U_t/s_n^0)X_t$. Then, given (A.7)–(A.9), $\mathscr{V}_n(\xi) - \mathscr{V}_n(\xi_0)$ will be equal to

$$-\frac{1}{2}(\beta_{10}-\beta_1)'(M_n^1(0,k,\underline{\theta})+W_{1n}(\underline{\theta}))(\beta_{10}-\beta_1) - (\beta_{10}-\beta_1)' s_n^0 \mathrm{N}_n(0,k) \tag{A.11}$$

$$-\frac{1}{2}(\beta_{20}-\beta_2)'(M_n^3(k_0,n,\underline{\theta})+W_{3n}(\underline{\theta}))(\beta_{20}-\beta_2) - (\beta_{20}-\beta_2)' s_n^0 \mathrm{N}_n(k_0,n) \tag{A.12}$$

$$-2\left(\frac{1}{2}\lambda' + \frac{1}{2}(\beta_{20}-\beta_2)'\right)(M_n^2(k,k_0,\underline{\theta})+W_{2n}(\underline{\theta}))\left(\frac{1}{2}\lambda + \frac{1}{2}(\beta_{20}-\beta_2)\right) \tag{A.13}$$

$$-(\lambda' + (\beta_{20}-\beta_2)')s_n^0 \mathrm{N}_n(k,k_0), \tag{A.14}$$

considering that $(\beta_{10}-\beta_2) = \lambda + (\beta_{20}-\beta_2)$ and that the components of $W_{hn}(\underline{\theta})$, for $h=1,2,3$ are the following:

$$\gamma_{hn}(\underline{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\Upsilon_n\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)X_t I_t^h, \quad \tilde{\gamma}_{hn}(\underline{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\Upsilon}_n\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)X_t I_t^h, \tag{A.15}$$

$$\gamma_n(\underline{\theta}) = \frac{1}{n}\sum_{h=1}^{3}\sum_{t=1}^{n}\tilde{\Upsilon}_n\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)I_t^h, \tag{A.16}$$

$$\dot{s}_{hn}(\underline{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}\psi_1\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)\frac{X_t I_t^h}{D_n(\underline{\theta})},$$

$$D_n(\underline{\theta}) = \frac{1}{n}\sum_{h=1}^{3}\sum_{t=1}^{n}\tilde{\psi}_1\left(\frac{U_t+\theta_h' X_t}{s_n(\underline{\theta})}\right)I_t^h. \tag{A.17}$$

The rest of the proof is similar to that of Fiteni (2002, Theorem 2). It suffices to prove that the corresponding Lemmas A.1–A.6 by Fiteni (2002) follows for the present estimation method, defined by an objective function which has an asymptotic linear approximation given by (A.11)–(A.14). This is obtained below, by Proposition A.2 (which prove the corresponding Lemma A.1), Proposition A.3 (proving Lemma A.2), Proposition A.4 (proving Lemma A.3) and Proposition A.5 (proving Lemmas A.4–A.6).

Proposition A.2. *Under* A1 *and* A2 (*except* A2.9), *is holds that*, (i) $\sup_{\pi\in\Pi,\theta\in\Theta}\|n^{-1}\sum_{t=1}^{[n\pi]}\dot{\eta}_{it}(\theta;\sigma_0)-\pi M_i(\theta)\|\xrightarrow{\text{p}}0$, *where* $M_i(\theta)$ *is defined by* A2.7, (ii) $\sup_{\underline{\theta}\in\Theta^3}\|W_n(\underline{\theta})-W(\underline{\theta})\|\xrightarrow{\text{p}}0$ *and* (iii) $\sup_{\underline{\theta}\in\Theta^3}\|W_{hn}(\underline{\theta})-W_h(\underline{\theta})\|\xrightarrow{\text{p}}0$, *for* $i=1,2$ *and* $h=1,2,3$, *with* $W(\underline{\theta})=(\sum_{h=1}^{3}e_hH_{\bar{\psi}_1}(\theta_h,s(\underline{\theta})))^{-1}\sum_{h=1}^{3}e_hH_{\bar{\rho}_2}(\theta_h,s(\underline{\theta}))$ *and* $W_h(\underline{\theta})=\dot{s}_j(\underline{\theta})\tilde{\gamma}_j'(\underline{\theta})+\gamma_j(\underline{\theta})\dot{s}_j'(\underline{\theta})+\dot{s}_j(\underline{\theta})\gamma(\underline{\theta})\dot{s}_j'(\underline{\theta})$, *such that*

$$\dot{s}_j(\underline{\theta})=-e_jH^X_{\psi_1}(\theta_j,s(\underline{\theta}))/D(\underline{\theta}),\quad D(\underline{\theta})=\sum_{h=1}^{3}e_hH^X_{\tilde{\psi}_1}(\theta_h,s(\underline{\theta})),$$

$$\gamma_j(\underline{\theta})=e_j(W(\underline{\theta})H^X_{\psi_1-\tilde{\psi}_1}(\theta_j,s(\underline{\theta}))-H^X_{\psi_2-\tilde{\psi}_2}(\theta_j,s(\underline{\theta}))),$$

$$\tilde{\gamma}_j(\underline{\theta})=e_j(W(\underline{\theta})H^X_{\psi_1+\tilde{\psi}_1}(\theta_j,s(\underline{\theta}))-H^X_{\psi_2-\tilde{\psi}_2}(\theta_j,s(\underline{\theta}))),$$

$$\gamma(\underline{\theta})=\sum_{h=1}^{3}e_h(W(\underline{\theta})H^X_{\tilde{\psi}_1-\tilde{\tilde{\psi}}_1}(\theta_h,s(\underline{\theta}))-H^X_{\tilde{\psi}_2-\tilde{\tilde{\psi}}_2}(\theta_h,s(\underline{\theta}))),$$

where $H^X_{\psi_k\pm\tilde{\psi}_k}(\theta_j,s(\underline{\theta}))=H^X_{\psi_k}(\theta_j,s(\underline{\theta}))\pm H^X_{\tilde{\psi}_k}(\theta_j,s(\underline{\theta})),H^X_{\tilde{\psi}_k-\tilde{\tilde{\psi}}_k}(\theta_j,s(\underline{\theta}))=H^X_{\tilde{\psi}_k}(\theta_j,s(\underline{\theta}))-H^X_{\tilde{\tilde{\psi}}_k}(\theta_j,s(\underline{\theta}))$, *defined by* A2.8 *for* $k=1,2$ *and* $j=1,2,3$.

Proof. Result (i) is immediate applying the triangle inequality, Lemma A.3 and Assumption A2.7. Lemmas A.2 and A.5 and the Slutzky Theorem establish results (ii) and (iii). □

Proposition A.3. *Let* $\Theta_0^3\subset\Theta^3$ *be a compact subset of* $\mathbb{R}^{3p}$, *containing neighborhoods of* $\underline{\theta}_0$, *such that* $s(\underline{\theta}_0)=\sigma_0$. *Consider a sequence* $\{\underline{\theta}_n,\ n\geqslant 1\}\in\Theta_0$, *with* $\underline{\theta}_n\rightarrow_{n\uparrow\infty}\underline{\theta}_0$. *Then*, *under* A1 *and* A2 (*except* A2.9), *it holds that* (i) $\sup_{\pi\in\Pi}\|n^{-1}\sum_{t=1}^{[n\pi]}\dot{\eta}_{jt}(\theta_{in};s_n(\underline{\theta}_n))-\pi M_j(\theta_0)\|\xrightarrow{\text{p}}0$, *where* $M_j(\cdot)$ *is defined by* A2.7, (ii) $\|W_n(\underline{\theta}_n)-W(\underline{\theta}_0)\|\xrightarrow{\text{p}}0$ *and* (iii) $\|W_{hn}(\underline{\theta}_n)-W_h(\underline{\theta}_0)\|\xrightarrow{\text{p}}0$, *for* $i,h=1,2,3$ *and* $j=1,2$.

Proof. Result (i) follows by the triangle inequality, Lemma A.7 an Assumption A2.7. Results (ii) and (iii) are obtained using Lemma A.7 and the Slutzky Theorem. □

Proposition A.4. *Let* $\{a_k\}_{k\geqslant 1}$ *be a sequence of decreasing positive constants and* $S_k^j=\sum_{t=1}^{k}\psi_j(U_t/\sigma_0)X_t$, *for* $j=1,2$. *Then*, *under* A1 *and* A2 (*except* A1.3, A2.5–A2.8), *there exists a* $K<\infty$ *such that*, *for every* $\varepsilon>0$ *and* $m>0$ *it holds that* $\Pr\{\max_{m\leqslant k\leqslant n}\}\sup_{\alpha\in\mathbb{R}^p,\|\alpha\|=1}\alpha_k|\alpha' S_k|>\varepsilon\}\leqslant\varepsilon^{-2}K(ma_m^2+\sum_{j=m+1}^{n}a_j^2)$.

Proof. By A2.4, the result follows from Lemma A.3-(3.2) by Fiteni (2002). □

Proposition A.5. *Consider* $\{\eta_t^0=W_0\eta_{1t}+\eta_{2t},t\leqslant n\}$ *and* $v_n(\pi)=n^{-1/2}\sum_{t=1}^{[n\pi]}\eta_t^0$, *such that* $\{v_n(\pi),\ n\geqslant 1\}$ *belongs to the bounded cadlag functions space in* $\mathbb{R}^p$ *and is defined on* $\Pi\subset[0,1]$. *Under* A1 *and* A2 (*except* A1.3, A2.5–A2.8), $v_n(\pi)\Rightarrow S^{1/2}B(\pi)$, *where "$\Rightarrow$" denotes weak convergence in the $D([0,1])$ space under the Skorokhod metric and $B(\pi)$ is a p-vector of independent Brownian process.*

Proof. By A2.4, $\{\eta_t^0\}$ is L_2-*NED* of size $-1/2$. Lemma 4 by Fiteni (2002) establishes the result. $\square$

Proposition A.6. *Under* A1 *and* A2, $\sup_\pi \| n^{-1/2}\sum_{t=1}^{[n\pi]} \psi_j(U_t/s_n^0)X_t - n^{-1/2}\sum_{t=1}^{[n\pi]}\psi_j(U_t/\sigma_0)X_t\| \xrightarrow{\text{p}} 0$ *holds.*

Proof. The results follows as in Corollary 2 by Fiteni (2002). $\square$

Proof of Theorem 2. The parameter estimator (2) can also be defined as $\hat{\xi}_n = \arg\min_{\xi\in\Theta^2\times\Pi}(\mathscr{V}_n(\xi) - \mathscr{V}_n(\xi_0))$. The limiting distribution of the estimators will be obtained analyzing the local behavior of the objective function on a compact set determined by Theorem A.1. Therefore, we reparametrize the objective function in (3) and (4), such that $\Lambda_n(v) = \mathscr{V}_n(\xi_0 + (n^{-1/2}v_1', n^{-1/2}v_2', n^{-1}P_\lambda v_3)') - \mathscr{V}_n(\xi_0)$, for $v=(v_1', v_2', v_3)' \in V_N \subset \mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}$, with $V_N = \{v: \|v_j\| < N,\ j=1,2,3\}$, a compact set defined for an arbitrary constant $N>0$. $P_\lambda = \mathrm{O}(\|\lambda\|^{-2})$, with $P_{\lambda_n} = \mathrm{O}(\|\lambda_n\|^{-2})$ for the decreasing case and $P_\lambda = 1$ for a constant λ. The weak convergence result follows taking into account that $\sqrt{n}(\hat{\beta}_{jn} - \beta_{j0}) = \hat{v}_j$, for $j=1$ and 2, and $n(\hat{\pi}_n - \pi_0) = P_\lambda \hat{v}_3$, such that $(\hat{v}_1', \hat{v}_2', \hat{v}_3)' = \arg\min_{v\in V_N}\Lambda_n(v)$, defined on a compact set for $N<\infty$. Again, we only consider the case of $v_3 < 0$, without loss of generality because of symmetry. For notational convenience $[v_3P_\lambda]$ will be denoted by v_3P_λ. First, observe that, by Lemma A.6, $s_n^0 \to^{\text{p}} \sigma_0$ and $s_n(\underline{\theta}_n) \to^{\text{p}} \sigma_0$ for $\underline{\theta}_n = (\delta_1 v_1'/\sqrt{n}, \delta_2(v_2'/\sqrt{n} + \lambda'), \delta_3 v_2'/\sqrt{n})' \to (0_p', \delta_2\lambda', 0_p')$ as $n\to\infty$. Similarly, $W_n^0 \to^{\text{p}} W_0$ by the Slutzky theorem and $W_n(\underline{\theta}_n) \to^{\text{p}} W_0$ by Proposition A.3. Then, considering (23)–(26), we obtain that

$$n\Lambda_n(v) = -v_1'\frac{s_n^0}{\sqrt{n}}\sum_{t=1}^{k_0}\eta_{tn}^0 - v_2'\frac{s_n^0}{\sqrt{n}}\sum_{t=k_0+1}^{n}\eta_{tn}^0 + v_1'\frac{s_n^0}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0}\eta_{tn}^0 \tag{A.18}$$

$$-\lambda' s_n^0\sum_{t=k_0+v_3P_\lambda+1}^{k_0}\eta_{tn}^0 - v_2'\frac{s_n^0}{\sqrt{n}}\sum_{t=k_0+v_3P_\lambda+1}^{k_0}\eta_{tn}^0 \tag{A.19}$$

$$-\tfrac{1}{2}v_1'\tau_0 M v_1 - \tfrac{1}{2}v_2'(1-\tau_0)Mv_2 - \tfrac{1}{2}\lambda' M(\lambda)\lambda P_\lambda v_3 + \mathrm{op}(1), \tag{A.20}$$

using again Proposition A.3 and noting that $\eta_{tn}^0 = W_n^0\eta_{1t}(0_p; s_n^0) + \eta_{2t}(0_p; s_n^0)$, for each t, and $n^{-1}P_\lambda = \mathrm{O}(n^{-1}\|\lambda\|^{-2})$, an o(1) term in both cases of λ, fixed and decreasing with n. Next, from Propositions A.5 and A.6, we have the following convergence results:

$$v_1'\frac{s_n^0}{\sqrt{n}}\sum_{t=1}^{k_0}\eta_{tn}^0 \;\Rightarrow\; v_1' S^{1/2}B(\tau_0), \qquad v_2'\frac{s_n^0}{\sqrt{n}}\sum_{t=k_0+1}^{n}\eta_{tn}^0 \;\Rightarrow\; v_2' S^{1/2}B(1-\tau_0)$$

and

$$v_j' \frac{s_n^0}{\sqrt{n}} \sum_{t=k_0+v_3 P_\lambda+1}^{k_0} \eta_{tn}^0 = \mathrm{op}(1),$$

for $j = 1, 2$, with S defined by A2.9. From here the result follows as in Theorem 2 by Fiteni (2002) using Propositions A.3, A.5 and A.6. $\square$

Proof of Corollary 1. Immediate from Theorem 2. $\square$

A.1. Lemmata

Lemma A.1. *Let Θ be a compact set of $\mathbb{R}^p$ and $[h_1, h_2]$ a closed interval with $0 < h_1, h_2 < \infty$. Under* A1.1, A2.1–A2.3 *and* A2.8, $\sup_{\pi\in\Pi} \sup_{\theta_i\in\Theta, s\in[h_1,h_2]} |n^{-1} \sum_{t=1}^{[n\pi]} \rho_1(s^{-1}(U_t + \theta_i' X_t)) - \pi H_{\rho_1}(\theta_i, s)| \xrightarrow{\text{a.s.}} 0$ *holds.*

Proof. By the triangular inequality, the left-hand side of above expression is upper bounded by

$$\sup_{\pi\in\Pi} \sup_{\theta_i\in\Theta, s\in[h_1,h_2]} \left| \frac{1}{n} \sum_{t=1}^{[n\pi]} \left(\rho_1\left(\frac{U_t + \theta_i' X_t}{s}\right) - \mathrm{E}\left[\rho_1\left(\frac{U_t + \theta_i' X_t}{s}\right)\right]\right)\right| \tag{A.21}$$

$$+ \sup_{\pi\in\Pi} \sup_{\theta_i\in\Theta, s\in[h_1,h_2]} \left| \frac{1}{n} \sum_{t=1}^{[n\pi]} \mathrm{E}\left[\rho_1\left(\frac{U_t + \theta_i' X_t}{s}\right)\right] - \pi H_{\rho_1}(\theta_i, s)\right|. \tag{A.22}$$

By A2.8, (34)$\to$ 0. By A2.1–A2.3 and A2.8, (33)$\xrightarrow{\text{p}}$0 follows from Lemma A3 of Andrews (1993). $\square$

Lemma A.2. *Under* A1.1–A1.2, A2.1–A2.3 *and* A2.8, *it holds that*, $\sup_{\underline{\theta}\in\Theta^3} |s_n(\underline{\theta}, \pi) - s(\underline{\theta}, \pi)| \xrightarrow{\text{a.s.}} 0$, *where, for* $\underline{\theta} = (\theta_1', \theta_2', \theta_3')' \in \Theta^3, s_n(\underline{\theta}, \pi)$ *and* $s(\underline{\theta}, \pi)$ *are such that*, $B_n(\underline{\theta}, \pi, s_n(\underline{\theta}, \pi)) = b$ *and* $B(\underline{\theta}, \pi, s(\underline{\theta}, \pi)) = b$, *respectively, with* $B_n(c, \pi, s) = n^{-1} \sum_{h=1}^3 \sum_{t=1}^n \rho_1(s^{-1}(U_t + c_h' X_t)) I_t^h$ *and* $B(c, \pi, s) = \sum_{h=1}^3 e_h H_{\rho_1}(c_h, s)$, *for* $c = (c_1', c_2', c_3')' \in \mathbb{R}^{3p}$, $s > 0$, $\pi \in \Pi \subset (0, 1)$ *and* $H_{\rho_1}(c_h, s)$ *defined by* A2.8.

Proof. For each π, define $h_1 = \inf_{\underline{\theta}\in\Theta^3} s(\underline{\theta}, \pi)$ and $h_2 = \sup_{\underline{\theta}\in\Theta^3} s(\underline{\theta}, \pi)$, such that $h_1 > 0$ and $h_2 < \infty$. From Lemma 1, it follows that

$$\sup_{\pi\in\Pi} \sup_{\underline{\theta}\in\Theta^3, s\in[h_1,h_2]} |B_n(\underline{\theta}, \pi, s) - B(\underline{\theta}, \pi, s)| \xrightarrow{\text{a.s.}} 0. \tag{A.23}$$

Let ε be such that $0 \leqslant \varepsilon \leqslant h_1/2$ and define, for each $\pi, g_1(\underline{\theta}, \pi) = B(\underline{\theta}, \pi, s(\underline{\theta}, \pi) + \varepsilon)$ and $g_2(\underline{\theta}) = B(\underline{\theta}, \pi, s(\underline{\theta}, \pi) - \varepsilon)$. Then, $g_1(\underline{\theta}) < b < g_2(\underline{\theta})$. Given $g_1(\cdot)$ and $g_2(\cdot)$, continuous functions in $\underline{\theta}$, we obtain that $\gamma_1 = \sup_{\underline{\theta}\in\Theta^3} g_1(\underline{\theta}) < b < \inf_{\underline{\theta}\in\Theta^3} g_2(\underline{\theta}) = \gamma_2$.

Let $\delta = \min\{b - \gamma_1, \gamma_2 - b\}$. If (A.23) holds, then there exists a n_0 such that $\forall n \geqslant n_0$,

$$\sup_{\pi\in\Pi}\ \sup_{\underline{\theta}\in\Theta^3, s\in[h_1/2,2h_2]} |B_n(\underline{\theta},\pi,s) - B(\underline{\theta},\pi,s)| < \frac{\delta}{2}. \tag{A.24}$$

Next, observe that, for each $\pi\in\Pi$,

(a) Given that $s(\underline{\theta},\pi)-\varepsilon \geqslant h_1-h_1/2=h_1/2$, then $s(\underline{\theta},\pi)-\varepsilon\in[h_1/2,2h_2]$, and therefore, by (A.24) $\sup_{\underline{\theta}\in\Theta^3}|B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)-B(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)| \leqslant \delta/2$, which also holds for the infimum. Because $\inf|A-B| > |\inf(A)-\sup(B)|$, $S_{\sup 1} \leqslant \inf_{\underline{\theta}\in\Theta^3} B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon) \leqslant S_{\sup 2}$, where $S_{\sup 1} = \sup_{\underline{\theta}\in\Theta^3} B(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)-\delta/2$ and $S_{\sup 2} = \sup_{\underline{\theta}\in\Theta^3} B(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)+\delta/2$. Hence, we obtain that $S_{\inf 1} \leqslant \inf_{\underline{\theta}\in\Theta^3}B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)$, where $S_{\inf 1}$ is defined as $S_{\sup 1}$, with the infimum instead of the supremum. However, given that $S_{\inf 1}=\gamma_2-\delta/2 \geqslant b+\delta-\delta/2=b+\delta/2$, it holds that $b+\delta/2 \leqslant \inf_{\underline{\theta}\in\Theta^3}B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)$.

(b) Given that $s(\underline{\theta},\pi)+\varepsilon \leqslant h_2+h_1/2 \leqslant 2h_2$, then $s(\underline{\theta},\pi)+\varepsilon\in[h_1/2,2h_2]$ and therefore, by (A.24) $\sup_{\underline{\theta}\in\Theta^3}|B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon)-B(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon)| \leqslant \delta/2$. Noting that $\sup|A-B| \geqslant |\sup(A-B)| \geqslant |\sup(A)-\sup(B)|$, $S'_{\sup 1} \leqslant \sup_{\underline{\theta}\in\Theta^3} B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon) \leqslant S'_{\sup 2}$, with $S'_{\sup 1} = \sup_{\underline{\theta}\in\Theta^3} B(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon)-\delta/2$ and $S'_{\sup 2} = \sup_{\underline{\theta}\in\Theta^3}B(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon)+\delta/2$. Hence, $S'_{\sup 2} = \gamma_1+\delta/2 \leqslant b-\delta+\delta/2 = b-\delta/2$.

As a consequence of (a) and (b), we obtain that, for $\varepsilon > 0$, there exists a n_0 such that $\forall n \geqslant n_0$,

$$\sup_{\underline{\theta}\in\Theta^3} B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)+\varepsilon) \leqslant b-\frac{\delta}{2} \leqslant b \leqslant b+\frac{\delta}{2} \leqslant \inf_{\underline{\theta}\in\Theta^3} B_n(\underline{\theta},\pi,s(\underline{\theta},\pi)-\varepsilon)$$

and, noting that $b = B_n(\underline{\theta},\pi,s_n(\underline{\theta},\pi))$, we conclude the proof of this lemma because $\forall\varepsilon > 0$, there will exist a n_0 such that $\forall n \geqslant n_0$, $s(\underline{\theta},\pi)-\varepsilon \leqslant s_n(\underline{\theta},\pi) \leqslant s(\underline{\theta},\pi)+\varepsilon$, uniformly in $\underline{\theta}\in\Theta^3$. □

For notational convenience, throughout this Lemmata, $s_n(\underline{\theta},\pi)$ and $s(\underline{\theta},\pi)$ will be denoted by $s_n(\underline{\theta})$ and $s(\underline{\theta})$, respectively.

Lemma A.3. *Suppose that*: (a) *Assumption* A2.1–A2.3 *hold*; (b) $f(z;\theta;s)$ *is an* $\mathbb{R}^q$*-valued function on* $Z\times\Theta\times[h_1;h_2]$, $0 < h_1, h_2 < \infty$, *that is continuous in* z *for all* $(\theta;s)\in\Theta\times[h_1;h_2]$ *and is continuous in* θ *uniformly over* $(z;\theta;s)\in C\times\Theta\times\mathbb{R}^+$ *for all compact sets* $C\subset Z$; (c) *For some* $\varepsilon > 0$, *it holds that* $\lim_{n\to\infty} n^{-1}\sum_1^n \mathrm{E}[\sup_{\theta\in\Theta;s\in[h_1;h_2]}|f(Z_t;\theta;s)|^{1+\varepsilon}] < \infty$. *Then,*

$$\sup_{\pi\in\Pi}\ \sup_{\theta\in\Theta;s\in[h_1,h_2]} \left\|\frac{1}{n}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta;s)-\mathrm{E}[f(Z_t;\theta;s)])\right\| \xrightarrow{\mathrm{p}} 0.$$

Proof. From Lemma A3 by Andrews (1993). □

Lemma A.4. *Suppose that assumptions of Lemma* A.2 *and* A.3 *hold. Let* $\Theta_0\subset\Theta$ *be a compact subset of* $\mathbb{R}^p$, *containing neighborhoods of* θ_0 *and consider a sequence* $\{\theta_n,\ n\geqslant 1\}\in\Theta_0$ *such that* $\theta_n \to_{n\uparrow\infty} \theta_0$. *Then*, $\sup_{\pi\in\Pi}\sup_{s\in[h_1,h_2]}\|n^{-1}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta_n;s)-\mathrm{E}[f(Z_t;\theta_0;s)])\| \to^{\mathrm{p}} 0$.

proof. Observe that, by the triangle inequality the left-hand side is upper bounded by

$$\sup_{\pi\in\Pi}\sup_{s\in[h_1,h_2]}\left\|\frac{1}{n}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta_n;s)\mathrm{E}f(Z_t;\theta_n;s))\right\|$$

$$+\sup_{\pi\in\Pi}\sup_{s\in[h_1,h_2]}\left\|\frac{1}{n}\sum_{t=1}^{[n\pi]}(\mathrm{E}f(Z_t;\theta_n;s)-\mathrm{E}f(Z_t;\theta_0;s))\right\| = (\mathrm{I})+(\mathrm{II}).$$

(I) is upper bounded by $\sup_{\pi\in\Pi},\sup_{s\in[h_1,h_2],\theta\in\Theta_0}\|n^{-1}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta;s)-\mathrm{E}f(Z_t;\theta;s))\| \xrightarrow{p} 0$, from Lemma A3 by Andrews (1993). To study (II), note that: (i) by the tightness condition of $\{F_n,\ n\geqslant 1\}$, we obtain that $n^{-1}\sum_{t=1}^{n}P(Z_t\notin C_j)\to 0$ as $j\to\infty$, for some sequence of compact sets $\{C_j,\ j\geqslant 1\}$ in Z, and (ii) $\forall j\geqslant 1$,

$$\sup_{n\geqslant 1}\sup_{\pi\in\Pi}\sup_{s\in[h_1,h_2]}\left|\frac{1}{n}\sum_{t=1}^{[n\pi]}\mathrm{E}[f(Z_t;\theta_n;s)-f(Z_t;\theta_0;s)]I\ (Z_t\in C_j)\right\|$$

$$\leqslant \sup_{z\in C_j}\sup_{s\in[h_1h_2]}\|f(Z_t;\theta_n;s)-f(Z_t;\theta_0;s)\|\to 0,$$

for $\theta_n\to\theta_0$, for function $\dot{\eta}(\cdot)$ defined in $(z,\theta)\in Z\times\Theta$, continuous by A1.1, and thus, uniformly continuous in the compact set C_j. (iii) $\sup_{\pi\in\Pi,s\in[h_1,h_2]}\|n^{-1}\sum_{t=1}^{[n\tau]}\mathrm{E}[f(Z_t;\theta_n;s)-f(Z_t;\theta_0;s)]\|$ converges to zero as $\theta_n\to\theta_0$ by results (i) and (ii). □

Lemma A.5. *Suppose that Assumptions of Lemma* A.4 *holds and for* $\underline{\theta}=(\theta_1',\theta_2',\theta_3')'\in\Theta^3$, *let* $s_n(\underline{\theta})$ *and* $s(\underline{\theta})$ *be defined by Lemma* A.2. *Then, for* $i=1,2,3$,

$$\sup_{\pi\in\Pi}\sup_{\underline{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta_i;s_n(\underline{\theta}))-\mathrm{E}[f(Z_t;\theta_i;s(\underline{\theta}))])\right\|\xrightarrow{p}0.$$

Proof. The proof is similar to that of Lemma A.4. The result follows by Lemmas A.2 and A.3. □

Lemma A.6. *Let* $\Theta_0^3\subset\Theta^3$ *be a compact subset of* $\mathbb{R}^{3p}$, *containing neighborhoods of* $\underline{\theta}_0$. *Consider a sequence* $\{\underline{\theta}_n,\ n\geqslant 1\}\in\Theta_0^3$ *such that* $\underline{\theta}_n\to_{n\uparrow\infty}\underline{\theta}_0$. *Then, under* A1.1–A1.2, A2.1–A2.3 *and* A2.8, *it holds that* $\|s_n(\underline{\theta}_n)-s(\underline{\theta}_0)\|\xrightarrow{p}0$.

Proof. By A2.9 and Lemma A.4, $\sup_{\pi\in\Pi}\sup_{s\in[h_1,h_2]}|B_n(\underline{\theta}_n,\pi,s)-B(\underline{\theta}_0,\pi,s)|\xrightarrow{p}0$, which proves the result by Lemma A.2. □

Lemma A.7. *Suppose that assumptions of Lemma* A.4 *hold. Then, for* $i=1,2,3$,

$$\sup_{\pi\in\Pi}\left\|\frac{1}{n}\sum_{t=1}^{[n\pi]}(f(Z_t;\theta_{in};s_n(\underline{\theta}_n))-\mathrm{E}[f(Z_t;\theta_{i0};s(\underline{\theta}_0))])\right\|\xrightarrow{p}0.$$

Proof. From Lemmas A.6 and A.3 by Andrews (1993), the result follows as in Lemma A.4. □

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

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., 1998. Estimation of multiple-regime regressions with least absolutes deviation. Journal of Statistical Planning 74, 103–134.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Bekaert, G., Gray, S.F., 1998. Target zones and exchange rates: and empirical investigation. Journal of International Economics 45, 1–35.

Ben-David, D., Papell, D.H., 1997. International trade and structural change. Journal of International Economics 43, 13–523.

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.

Bierens, H.J., 1981. Robust Methods and Asymptotic Theory. In: Lecture Notes in Economics and Mathematical System, Vol. 192. Springer, Berlin.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Corbae, D., Ouliaris, S., 1991. A test of long-run purchasing power parity allowing for structural breaks. The Economic Record 64, 26–33.

Csörgo, M., Horváth, L., 1997. Limit theorems in change-point analysis. Wiley Series in Probability and Statistics. Wiley, New York.

Davidson, J., 1994. Stochastic limit theory. Advanced Texts in Econometrics. Oxford University Press, Oxford.

Delgado, M.A., Hidalgo, J., 2000. Nonparametric inference on structural breaks. Journal of Econometrics 96, 113–144.

Feldstein, M., Stock, J.H., 1994. The use of a monetary aggregate to target nominal GDP. Monetary Policy. University of Chicago Press, Chicago.

Fiteni, I., 2002. Robust estimation of structural break points. Econometric Theory 18, 349–386.

Friedman, B.M., Mieselman, D., 1963. The relative stability of the investment multiplier and monetary velocity in the United States, 1897–1958. Stabilization Policies, Commission on Money and Credit. Prentice-Hall, Englewood Cliffs, NJ, pp. 165–268.

Garcia, R., Perron, P., 1996. An analysis of the real interest rate under regime shifts. Review of Economic Studies 78, 111–125.

Hackl, P., Westlund, A., 1989. Statistical analysis of structural change: an annotated bibliography. Empirical Economics 143, 167–192.

Hackl, P., Westlund, A. (Eds.) 1991. Economic Structural Change: Analysis and Forecasting. Springer, Berlin, pp. 95–119.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust statistics: the approach based on influence functions. Wiley, New York.

Hansen, B.E., 1991. Strong laws for dependent heterogeneous processes. Econometric Theory 7, 213–221.

Hinkley, D.V., 1970. Inference about the change point in a sequence of random variables. Biometrika 57, 1–17.

Huber, P.J., 1977. Robust methods of estimation of regression coefficients. Mathematische Operationsforschung Statistik. Series Statistics 8, 41–53.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Huber, P.J., Dutter, R., 1974. Numerical solutions of robust regression problems. In: G. Brushmann (Ed.), COMPSTAT, Proceedings Symposium on Computational Statistics. Physike-Verlag, Berlin, pp. 165–172.

Ibramigov, I.A., 1962. Some limit theorems for stationary processes. Theory of Probability and its Applications 7, 349–382.

Koenker, R.W., D'Orey, V., 1987. Computing regression quantiles. Royal Statistical Society Series Applied Statistics 8, 383–393.

Krämer, W., Sonnberger, H., 1986. The Linear Regression Model under Test. Physica-Verlag, Heidelberg.

Krishnaiah, P.R., Miao, B.Q., 1988. Review about estimation of change points. Handbook of Statistics, Vol. 7. Elsevier, New York.

Maddala, G.S., Kim, I.-M., 1998. Unit Roots, Cointegration and Structural Change. Themes in Modern Econometrics. Cambridge University Press, Cambridge.

McLeish, D.L., 1975a. A maximal inequality and dependent strong laws. Annals of Probability 3, 829–839.

McLeish, D.L., 1975b. Invariance principles for dependent variables. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite 32, 165–178.

Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomics time series: some evidence and implications. Journal of Monetary Economics 4, 458–467.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361–1401.

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. Journal of Business and Economic Statistics 8, 153–162.

Perron, P., Vogelsan, T.J., 1992. Testing for a unit root in a time series with a changing mean: corrections and extensions. Journal of Business and Economic Statistics 10, 467–470.

Picard, D., 1985. Testing and estimating change-points in time series. Advances in Applied Probability 17, 841–867.

Pollard, D., 1984. Convergence of Stochastic Processes. Springer, New York.

Pötscher, P.M., Prucha, I.R., 1991. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: consistency and approximation concepts. Econometric Reviews 10, 125–216.

Rappoport, P., Reichlin, L., 1989. Segmented trends and non-stationary time series. Economic Journal 99, 168–177.

Reyes, M.G., 1999. Size, time-varying beta, and conditional heterocedasticity in UK stock returns. Review of Financial Economics 8, 1–10.

Rousseauw, P.J., Yohai, V.J., 1984. Robust regressions by means of S-estimators. In: Franke, J., Härdle, W., Martin, R.D. (Eds.), Robust and Non-Linear Time Series, Lecture Notes in Statistics, No. 26. Springer, New York, pp. 256–272.

Stock, J.H., 1997. Unit roots, structural breaks and trends. In: Engle R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol 4. Elsevier, New York, pp. 2739–2841.

Wooldridge, J.M., White, H., 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. Econometric Theory 4, 210–230.

Yao, Y.-C., 1987. Approximating the distribution of the change-point in a sequence in independent r.v.'s. Annals of Statistics 15, 1321–1328.

Yohai, V.J., Zamar, R., 1988. High breakdown-point estimates of regression by means of the minimization of an efficient scale. Journal of the American Statistical Association 83, 406–413.