INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2004-46

Testing the Global Stability of a Linear Model

Ph. BARBE[1]
P. BERTAIL[2]

---

[1] CNRS.
[2] CREST, Laboratoire de Statistiques.

TESTING THE GLOBAL STABILITY OF A LINEAR MODEL

Philippe Barbe[1] and Patrice Bertail[2]
[1]CNRS and [2]CREST, Laboratoire de Statistiques

Abstract. We propose a test for assessing the stability of a temporal econometric linear model. The test is based on subsampling ideas more precisely to the consistency of the least square estimator whose behavior is studied on subintervals over time. Our test allows to detect breakpoints on the boundaries under reasonable assumptions. Under the null hypothesis, we derive the asymptotic distribution of some simple criteria compatible with the linearity of the model and study them by simulations. We also present a diagnostic plot which helps in determining the nature of the nonlinearities (outliers, structural changes, polynomial links) of the underlying model.

Résumé. Nous proposons un test de stabilité globale d'un modèle linéaire dans le temps. Ce test est basé sur des idées de sous-échantillonnage et l'estimation de l'estimateur des moindre carrés ordinaires indéxé par des intervalles de temps. Notre test permet en particulier de détecter des ruptures sur les bornes d'observations sous des hypothèses raisonnables. Nous obtenons la distribution asymptotique d'un critère simple compatible avec l'hypothèse nulle de linéarité et l'étudions par simulations. Nous présentons également un test graphique simple qui aide interpréter la nature des non linéarités (points abérrants, changements structurels, comportement polynomial) du modèle sous-jacent.

Keywords. Resampling, empirical process, linear model, structural changes, simulations.

AMS 2000 subject classification: Primary 62J20; secondary 62G09, 62J05, 62M10.

1. Introduction. Before drawing inference from a linear model, it seems important to assess the stability of the relationship. A first approach, coherent with various physical, biological or economical theories is to test for structural changes (see for instance Basseville and Nikiforov (1993) and the reference therein). This problem has been extensively studied in the statistical literature in the case of a single breakpoint τ and has been extended to various specification (non linear regression model, times series models, nonlinear simultaneous equations models etc…) and different stability problems (tests of finite multiple structural changes, tests of cross section consistency), see for instance Andrews and Ploberger (1994), Bai and Perron (1998), where references may be found. To avoid any misinterpretation, we stress that the tests we are interested in here do not focus on the problem of testing a linear model against a non linear model but rather intend to detect instabilities or structural changes in a given linear (or may be non-linear) model: relevant tests in this other direction may be found for instance in Härdle and Mammen (1993), Stute (1997), Spokoiny (2001).

When the breakdate(s) is(are) known, the tests proposed are more or less derived from the analysis of covariance in the well known two regressions problem, see Chow (1960). The basic ideas for testing parameter consistency in the single unknown breakpoint problem goes back to Quandt (1960) who proposed to take the largest Chow (1960) statistics over all possible breakdates, leading to the now classical change point estimation problem over an interval $[0, n]$. The intuition behind the proposed tests is that if we split the sample in two subsamples, the set of observations before and after a date t, then the difference between estimations (or monotone transformations) over period $[0, t]$ and period $[t, n]$ should be equal to 0 if there is no structural change and is maximal at the change point $t = \tau$ if it exits. Andrews (1993), Andrews and Ploberger (1994) have extended these ideas to various parametric models and proposed some optimal tests which are now currently used in the econometric literature. Extending ideas of Darkhovshk (1976) and Carlstein (1988), Dümbgen (1991) proposed a general non parametric approach of the change point problem estimation and gives various powerful Bayes exponential tests. From a technical point of view, the main results are proved by studying transformations of the weighted process (indexed by t) of the difference between the empirical distribution of the observations before and after t. These results may also be used to test the consistency over time of a smooth functional of the empirical distribution. However these tests are unable to detect structural change near the boundary of the interval of observation. One purpose of this paper is to propose some tests that will take care of this feature.

In this paper we extend some of the mentioned ideas to test the global stability of the model. Our tests are different from previous tests proposed in the statistical literature because, in some sense, they allow for an infinity of breakpoints and variations of the parameter. However our aim is not to locate the breakpoints; it is only to assess the consistency of the linear relationship. On the contrary to some econometric tests of structural change, for instance predictive tests, we do not make any prior assumption on the location of the change points. Indeed, as we shall see in the simulations, it may be very difficult to detect non-linearities and/or structural changes just by looking at the data.

The principle is based on the estimation of the parameter over (almost) all sub-intervals of $[0, n]$. The intuition behind the tests that we propose is that, under the hypothesis of global stability, all the estimations over subsets of observations must be close to the true parameter. This suggests to study an estimator of the parameter as a process indexed by intervals, see also Shorack and Wellner (1986, chapter 17). This approach is closely related to subsampling methods and more particularly to Jackknife techniques, used for instance in the detection of outliers, see Belsley, Kuh and Welsh (1980). However in our case, a departure from the model means misspecification of the model rather than inadequacy of the observations. The proposed procedure is also closely related to the subsampling scheme studied by Politis and Romano (1994), who show that it is possible to build first order correct confidence intervals under a very general set of assumptions, if the underlying model is true. However in our case, the resampling technique is not used to infer on the value of the parameters but to infer on the validity of the model itself.

Moreover on the contrary to most tests in the structural change literature, structural instability may be detected on the bounds of the intervals of observations, that is, at the very beginning of the sample or the end, but also on short period. Instability may indeed come from outliers but also from instantaneous shocks. We shall give some graphic tools to characterize and interpret the misspecification detected by the tests. Particularly, when we treat the parameter componentwise, we give a very useful statistics, closely related to Watson's (1961) test of uniformity on a circle, whose critical values may be directly deduced from the Kolmogorov-Smirnov test.

In part 2 we describe the main results and discuss some possible extensions particularly to non-linear model and to accommodate residuals which are not i.i.d.

In part 3, we give a simulation study of the tests of instability in various situations. The proofs of the results are postponed in part 4

2. Stability of linear specification. Consider the commonly used random design regression model in which the observation (X_i, Y_i) satisfy

$$Y_i = X_i'b + \epsilon_i \quad , \; i = 1, \ldots, n,$$

where the X_i's are $(p, 1)$ random vectors and the ε_i's are i.i.d. random variables. We write X for the $n \times p$ matrix whose i-th row is X_i' and Y for the column vector $(Y_i)_{1 \leq i \leq n}$.

The usual least square estimator (l.s.e.) of b is

$$\hat{b}_n = (X'X)^{-1}X'Y \, .$$

We assume that

A1. $E(\epsilon_i \,|\, X_i) = 0$
A2. $E\mathrm{Var}(\epsilon_i \,|\, X_i) = \sigma^2$ is positive and finite.
A3. $\lim_{n\to\infty} \max_{1\leq i\leq n} n^{-1} X_i^{(m)} \mathrm{Var}(\epsilon_i \,|\, X_i^{(m)}) = 0$ a.s.
A4. $\Sigma = E X_i X_i' \epsilon_i^2$ is finite and positive.
A5. $E|X_1|^\beta$ is finite for some $\beta > 6$.
A6. There exists a constant c such that for any unit vector u, and any ϵ positive, $P\{\, |u'X| \leq \epsilon \,\} \leq c\epsilon$.

A1–A4 are conditions which ensure the good behavior of the explanatory variables, and the convergence of $\widehat{b}_n$. They are a little weaker than the traditional ones. In particular, we do not require that the conditional variance of ϵ_j given X_j is constant. In our case we allow for a little heteroscedasticity of the residuals conditionally to the explanatory variables, an hypothesis which seems more appropriate in many models.

Assumption A6 is somewhat unusual. If X has density f, the density of $u'X$ at r is the Radon transform

$$g_u(r) = \int_{v:v'u=r} f(v) \, \mathrm{d}v \, .$$

If this is a bounded function in a neighborhood of the origin, then

$$\begin{aligned} P\{\, |u'X| \leq \epsilon \,\} &= \int_{-\epsilon}^{\epsilon} g_u(r) \, \mathrm{d}r \\ &\leq 2\epsilon \sup\{\, g_u(r) \,:\, |r| \leq \epsilon \,\} \, , \end{aligned}$$

and A5 holds. Thus, to verify A5, it suffices to check that for any unit vector u, the function

$$r \mapsto \int_{u^\perp} f(ru + y) \, \mathrm{d}y$$

is bounded independently of u in a neighborhood of the origin.

Now define Π to be the half unit square

$$\Pi = \{\, (s,t) \in \mathbf{R}^2 \,,\, 0 \leq s \leq t \leq 1 \,\} \,.$$

Let d be an integer such that

$$\beta\Big(\frac{d}{2} - (p+1)\Big) - 10(p-1) - 3d > 0 \tag{2.1}$$

Such integer exists under A5. Note that d must be larger than $2(p+1)$ and that the bigger β is, the smaller d can be chosen.

For any pair (s,t) in Π, we define the least square estimator of b on the set of all observations observed in the interval $[\,ns, nt\,]$,

$$b_n(s,t) = \begin{cases} \big(\sum_{ns<i\leq nt} X_i X_i'\big)^{-1} \sum_{ns<i\leq nt} X_i Y_i & \text{if } |s-t| > d/n \\ 0 & \text{otherwise.} \end{cases}$$

The definition of $b_n(s,t)$ is such that it is indeed defined on the whole set Π. If $|s-t| < p/n$, the matrix $\sum_{ns<i\leq nt} X_i' X_i$ is certainly non-invertible, for it has rank less than its dimension. The quantity $d \geq 2(p+1)$ is a measure of the smallest fraction of the data that one should retain to calculate the l.s.e. It essentially depends on the tail of the explanatory variables. If all moments of the explanatory variables are finite one can choose $d = 2(p+1)$.

The following theorem shows that the process $b_n(s,t)$, indexed by Π, correctly normalized, converges to a Gaussian process. To state it, we denote by $\mathrm{Bd}(\Pi)$ the set of all bounded functions on Π. We equipped $\mathrm{Bd}(\Pi)$ with the uniform topology induced by the uniform norm $||\cdot||_\infty$. In other words, a sequence of functions g_n, $n \geq 1$, defined and bounded on Π, converges to a function g in $\mathrm{Bd}(\Pi)$, if and only if

$$\lim_{n\to\infty} ||g_n - g||_\infty = \lim_{n\to\infty} \sup_{(s,t)\in\Pi} ||g_n(s,t) - g(s,t)|| = 0 \,.$$

It then makes sense to speak of weak convergence (or convergence in distribution) of (the distribution of) random variables in $\mathrm{Bd}(\Pi)$; see Pollard (1984). Define $U = E(X_1 X_1')$. Assumption A4 ensures that U is positive.

Theorem 2.1 *Assume that conditions A1 to A6 hold. Under the hypothesis of linearity,*

$$\sqrt{n}(t-s)\big(b_n(s,t) - b\big) \xrightarrow{w} U^{-1}\Sigma^{1/2}\big(W_p(t) - W_p(s)\big) \tag{2.2}$$

where W_p is a p-dimensional Brownian motion and the convergence holds in probability given $(X_i)_{i\geq 1}$. Therefore,

$$\sqrt{n}(t-s)\big(b_n(s,t) - \widehat{b}_n\big) \xrightarrow{w} U^{-1}\Sigma^{1/2}\big(B_p(t) - B_p(s)\big) \tag{2.3}$$

where B_p is a p-dimensional Brownian bridge and the convergence is in conditional probability given $(X_i)_{i\geq 1}$.

The proof of Theorem 2.1 is deferred to section 4. Theorem 2.1 is quite easy to understand. On the hypothesis of linearity, the estimators defined on subintervals are unbiased estimators of b and are asymptotically normally distributed with a covariance matrix which is a function of the length of the intervals. The covariance between two different estimators depends on the overlap of the two different intervals. The power of this result is that it is uniform on $[0,1]$, allowing a control of the stability of the parameter on the bound of the interval and for small values of $t-s$. The main difficulty of the proof is actually to control the behavior of the process for small increments (that will allow to detect transitory shocks or structural change on the boundaries) . Indeed if the difference between t and s is big, the estimator is typically constructed with a large number of observation, whereas is $t-s$ is of order k/n, then the estimator is constructed with a finite number of observations.

From a practical point of view, for a fixed n, the process is approximated by calculating all the possible estimators of b on subintervals of the form $[k/n,(k+l)/n]$ with $l=d+1,\dots,n-1$ and $k=1,\dots,n-l$. The number of calculations of the estimator of b needed to approximate the process $b_n(\cdot)$, is of order n^2. However, it is possible to reduce the computing cost in using recursive estimates. Knowing the value of the estimator on $[k/n,(k+l)/n]$, it is easy to obtain its value on $[k/n,(k+l+1)/n]$ (see for instance Belsley, Kuh and Welsh, 1980). Of course if n is very large it may be time (computer) consuming to calculate all the possible values of the estimators over all the possible subintervals. In that case, just like for the bootstrap or subsampling distribution (see Politis and Romano, 1994), we may replace the original process by a Monte Carlo approximation by drawing B sub-intervals at random and computing the corresponding estimators. Asymptotically, if B gets large the asymptotic performance will not be entailed by this Monte Carlo step.

Similar results may be obtained for other estimators, 2 stages or 3 stages least squares estimators based on some instrumental variables, under a similar set of assumptions on both the explanatory and instrumental variables.

This result suggests numerous ways of testing the hypothesis of global stability of the model. A first category of tests based on (2.3) may be seen as uniformity tests of the Cramer-Von Mises type (see also Watson, 1961, 1967). Indeed, if $|\cdot|_{\mathrm{Bd}(\Pi)}$ is any norm on $\mathrm{Bd}(\Pi)$, continuous w.r.t. $||\cdot||_\infty$, it is easy to obtain the limiting distribution of

$$D_n = \big|(t-s)\big(b_n(t,s) - \widehat{b}_n\big)\big|_{\mathrm{Bd}(\Pi)}\,.$$

Moreover, if $\widehat{\Sigma}$ and $\widehat{U}$ are convergent estimators of Σ and U, we can also derive the limiting distribution of

$$W_n = 2^{-1}\big|(t-s)\widehat{\Sigma}^{-1/2}\widehat{U}\big(b_n(s,t) - \widehat{b}_n\big)\big|_{\mathrm{Bd}(\Pi)}\,.$$

Examples of convergent estimators of Σ and U are

$$\widehat{\Sigma} = n^{-1}\sum_{1\le i\le n} X_i X_i'\hat{\epsilon}_i^2$$

and

$$\widehat{U} = n^{-1}\sum_{1\le i\le n} X_i X_i'$$

where $\hat{\epsilon}_i = Y_i - X_i'\hat{b}_n$ are the estimated residuals of the initial model .

The continuous mapping theorem and Theorem 2.1 readily imply the following:

Corollary. *Under the hypotheses of Theorem 1, the limiting distribution of W_n (resp. D_n) is that of $2^{-1}|B_p(t) - B_p(s)|_{\mathrm{Bd}(\Pi)}$ (resp. $|U^{-1}\Sigma^{1/2}\big(B_p(t) - B_p(s)\big)|_{\mathrm{Bd}(\Pi)}$)*

Remark 1. Since $\widehat{U}$ and $\widehat{\Sigma}$ may not be good estimates of U and Σ, and the asymptotic approximation a poor approximation of the true distribution it may be preferable to use robust estimators of the variance (as suggested in the simulations) or to approximate the true distributions of D_n and W_n by bootstrapping. However, it is clear that one should construct the bootstrap model assuming a linear relation if one wants to obtain a consistent estimator of the limiting distribution. The usual way to do this is to construct a pseudo model using the previous estimates and the centered (orthogonal to the instrumental variables if there are some) estimated residuals, see Freedman (1981, 1984). It is easy to see that such a procedure may be adapted in our context and may be used to approximate the true distribution. Of course this procedure would be very computationally expensive since the original tests is already computer intensive. We shall see in the simulations that even with a very crude asymptotic approximation, our subsampling procedure is very informative and allows to detect various misspecifications. However, it may be important to have a robust estimate of the variance, for instance for the detection of outliers (see simulations 7 and 8).

Remark 2. Notice that it is always possible to treat the problem componentwise. In the unidimensional case, if we choose a L_2-norm for $|\cdot|_{\mathrm{Bd}(\Pi)}$, then the statistics W_n has the same limiting distribution as the random variable

$$W = 2^{-1}\int_0^1\int_0^1 \big(B_1(t)-B_1(s)\big)^2\,\mathrm{d}t\,\mathrm{d}s = \int_0^1 B_1(t)^2\,\mathrm{d}t - \Big(\int_0^1 B_1(t)\,\mathrm{d}t\Big)^2 .$$

This is the limiting distribution of Watson's statistics, see Shorack and Wellner (1986, p.142 and p.220). Watson (1961, 1967) shows that W has the same distribution as

$$\Big(\sup_{0\le t\le 1}|B_1(t)|\Big)^2/\pi^2$$

It follows that the critical value of our test W_n may be directly obtained from the limiting distribution of the Kolmogorov-Smirnov test. This yields the following table of critical values for the test W_n, in the unidimensional case

$P\{\,W>w\,\}$	0.01	0.05	0.1	0.2	0.3
w	0.27	0.19	0.15	0.12	0.97

Table 1

Before concluding this theoretical section, let us examin shortly the important issue of the power of our tests. Under a nonlinear fixed alternative,

$$Y_j = m(X_j, j/n) + \epsilon_j\,,$$

where $m(\cdot,\cdot)$ is a nonlinear function, our proof shows that under suitable conditions $n^{-1}\sum_{ns<i\le nt} X_iY_i$ converges to $E\int_s^t X_1 m(X_1,u)\,\mathrm{d}u$. For fixed t and s, we then have

$$\lim_{n\to\infty} b_n(s,t) = U^{-1}(t-s)^{-1}\int_s^t E(Xm(X,u))\,\mathrm{d}u$$

in probability. This limit is constant if and only if the function $u \longrightarrow EXm(X,u)$ is constant. Our test will not detect such alternative (except in some particular cases, for instance if the explanatory variables are not i.i.d. and are for instance ordered). Hence, it is not really aimed at detecting nonlinearities but rather nonstability of the original specification. Perhaps more important than its power against a fixed alternative is the power against contiguous ones. So assume that

$$Y_j = X_jb + n^{-1/2}m(X_j, j/n) + \epsilon_j\,. \tag{2.4}$$

Our next result gives an idea about the local asymptotic power. In practical situation when one wants to protect himself or herself against a specific contiguous alternative, it allows to derive tests with good local properties. The conditions of the theorem are far from optimal but they are good enough to give an idea on the local power and allow a short and simple proof. We assume that

A7. $m(\cdot,\cdot)$ and its partial derivative with respect to the second variable are both continuous and bounded.

We define the function

$$h(t) = \int_0^t EXm(X,u)\,\mathrm{d}u - t\int_0^1 EXm(X,u)\,\mathrm{d}u\,.$$

Theorem 2.2. *Under (A1)–(A7) and for the model (2.4), we have the convergence in distribution of the processes*

$$\sqrt{n}(t-s)\big(b_n(s,t)-\hat{b}_n\big) \mapsto U^{-1}\Big(\Sigma^{1/2}\big(B_p(s)-B_p(t)\big) - h(s) + h(t)\Big)$$

as n tends to infinity.

Theorem 2.2 tells us that for alternatives with large function $h(\cdot)$, the process $\sqrt{n}(t-s)\big(b_n(s,t)-\hat{b}_n\big)$ has a substantial drift. In this case, the test will tend to reject the null hypothesis as it should.

Some Extensions. The results may be extended to other non-linear models.

1) One would like for instance to test for the stability of the nonlinear model

$$Y_j = m(X_j,\theta) + \epsilon_j\ ,\ j = 1,\ldots,n$$

where m is a known function and θ a finite dimensional parameter. Assuming that under the null hypothesis, we have a convergent estimate of θ or that some suitable moments conditions hold, it is possible to study the process of the estimated values indexed by the subintervals of $[\,0,1\,]$. For instance if θ_n is the nonlinear least-square estimate of θ, the limiting behavior of θ_n is given by a straightforward linearization: Theorem 1 and its derived results may be adapted to that case without any effort.

2) It is also possible to generalize these ideas to non-i.i.d. residuals. Indeed the main tool used in the proof is the approximation of partial sums by a Wiener process and the ability to control the increments of the process. Many results on empirical processes have been obtained in that

direction in the recent year (see Doukhan, 1994) and may be used to obtain generalizations. In this situation, it is possible to mimic the proof of Theorem 1, at the cost of higher sophisticated notations, to obtain similar results. Of course the limiting distributions of D_n are more complicated and depend on the nature of the serial dependence. In that case it may be more difficult to obtain a convergent and robust estimate of the asymptotic variance. Thus, the subsampling techniques of Politis and Romano (1994) may prove to be very useful, to get a robust estimate of the variance (under the null hypothesis). It would be also very useful to use the same type of tools for integrated processes. One more time, this possibility mainly relies on the existence of a strong invariance principle (see Phillips and Durlauf, 1986) and the possibility to control for small increments, which may be more difficult in that case, unless we have additional assumptions on the explanatory variables.

3. Diagnostic tests: A simulation study. In this paragraph, we study the performance of the test W_n in the case of least-squares estimators and show how undersampling may be used to detect misspecifications. All the simulations are presented in the same way. In every simulations we try to fit the linear model $y_t = a + bx_t$, t=1,...,n. The first picture represents the link between the variables and the estimated linear relationship. It shows that it may be sometimes difficult to visually detect structural changes or non-linearities directly on the data. The second one represents the values of $(t-s)(b_n(t,s) - b_n)$ as a function of $t-s$, and two pointwise confidence bands respectively at level 95% and 99%. The bands are constructed in the following way. Let $\sigma_{\hat{b}_n}$ be the standard deviation of $\hat{b}_n$. For each values of $t-s$, the random variable $(t-s)(b_n(t,s)-\hat{b}_n)/\sigma_{\hat{b}_n}$ is asymptotically normal with variance $(t-s)(1-t+s)$. Let z_α be the quantile of order $1-\alpha$ of the standard Gaussian distribution. For each value $t-s = k/n, k > 0$, we build a confidence region of type

$$\left[\, -\sigma_{\hat{b}_n} z_\alpha k/n(1-k/n)\,,\, \sigma_{\hat{b}_n} z_\alpha k/n(1-k/n)\,\right],$$

Under the null hypothesis, for any fixed positive δ, all the values $(t-s)(b_n(t,s) - b_n)$, for $t-s \geq \delta$ should be inside the bounds at the given level. This is a very crude test which is of course less powerful than the previous tests that we studied in part 2. Nevertheless, it is very interesting to visually detect outliers, transitory shocks and/or structural changes. We also give the results of the linear regression and the value of test W_n.

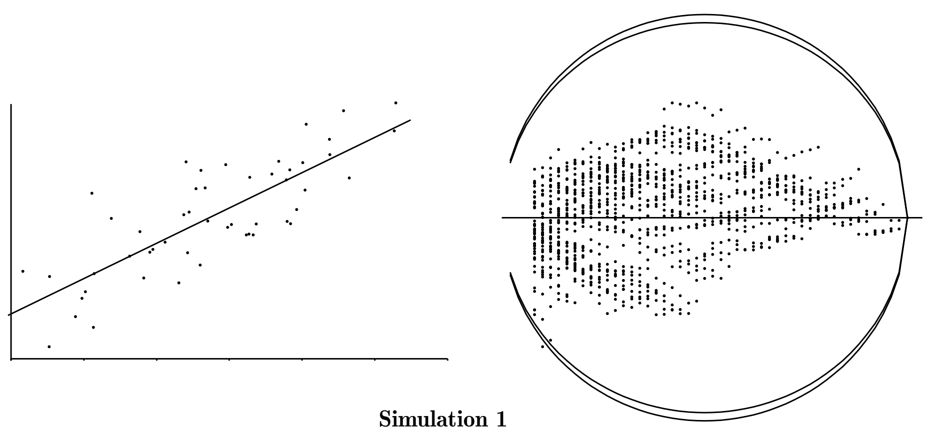
In our simulations the residuals are i.i.d. Gaussian r.v.'s with mean 0, standard deviation σ. The design points X_t's are randomly generated according to the rule $X_t = 2{*}U_{1,t}{+}U_{2,t}$, where $U_{i,t}$, $i = 1, 2$, $t = 1, \ldots, n$ are independent uniform random variables (except in simulations 11 and 12). It follows that the standard deviation of the X_t is 0.65. All the computations have been performed with Splus on an old bi-processor Pentium pro station. The routines are available on request. The test and the graphics may be obtained in less than 10 seconds for $n = 50$, 1 minutes for $n = 100$, and 5 minutes for $n = 500$, which is still reasonable. In the following, we give some details on the simulations and some short comments since the graphics speak by themselves. Note also that in nearly all the simulations, the adjusted R^2 is rather high, suggesting a good fit, even when the model is not well specified.

Simulations 1. We consider the true linear model

$$y_t = a + x_t b + \epsilon_t\,, \quad t = 1, \ldots, n,$$

with $a = 1,\ b = 1,\ \sigma = 1$, and $n = 50$.

The residual standard error (RSE) of the estimation is 0.51. The multiple R^2 (MR^2) is 0.65. The F-statistics is 89.7 with 1 on 48 degrees of freedom; its p-value is 0.00. The table bellow the plot summarizes the estimation. The column 'est' contains the estimated value of the coefficient; the column 'std' contains the estimated standard deviations of the estimators of the coefficients; the 'St t' column contains their Student t statistics; and the 'pv t' column indicates the p-value of the Student statistics.



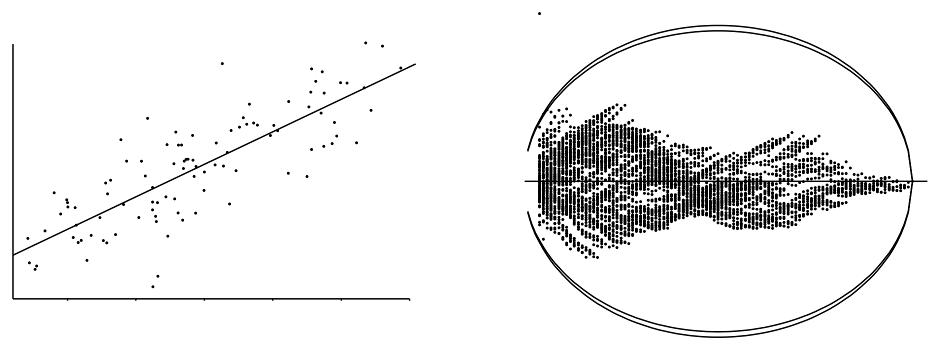
Simulation 1

	est	std	St t	pv t
a	1.0	0.28	3.6	0.00
b	1.1	0.11	9.5	0.00

For this realization of the model, the statistics W_n is 0.09. From table 1, we see that the quantile of order 95% of the limiting distribution of W_n is

0.19 while that of order 99% is 0.27. Thus, we accept the assumption that the model is linear, even at the low level 1%. On the plot, notice the form of the trajectories of the $b_n(t,s)$ which seem "random" compare to some of the next simulations.

Simulation 2. The model is now the same as in simulation 1, but we have twice as many observations, that is $n = 100$.



Simulation 2

	est	std	St t	pv t
a	1.0	0.19	5.0	0.00
b	1.0	0.07	13.9	0.00

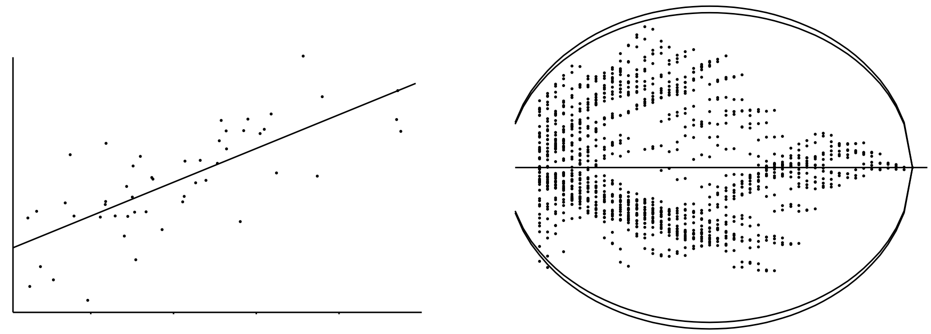
RSE	MR^2	F(1/98)	pv F	W_n
0.52	0.66	191.9	0.00	0.08

Again, we do not reject the null hypothesis, even at a very low level. The trajectories of the $b_n(t,s)$ are not that much different from that of simulation 1 if one factors out the scale (the scale changed as the number of observation did).

Simulation 3. We consider the model with a structural change at $[n/2]$

$$y_t = a_1 + x_t b_1 + \epsilon_t\,, \qquad t = 1, \ldots, [n/2]$$
$$y_t = y_{[n/2]} + b_2(x_t - x_{[n/2]}) + \epsilon_t\,, \qquad t = [n/2]+1, \ldots, n$$

with $a_1 = 1$, $b_1 = 1$ and respectively $n = 50$, $\sigma = 0.1$ and with $b_2 = 1.5$.



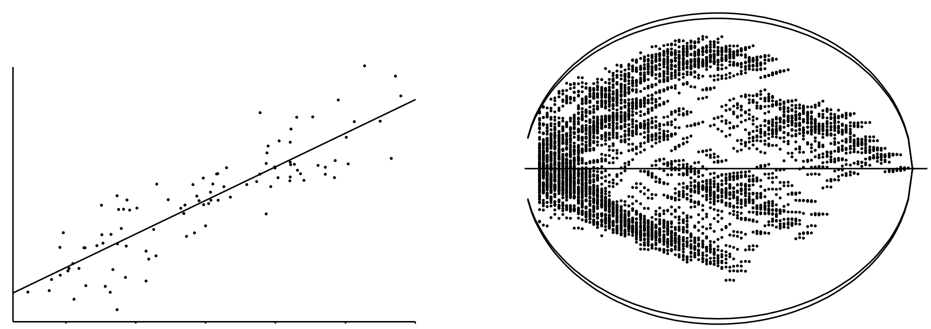
Simulation 3

	est	std	St t	pv t
a	0.9	0.3	3.1	0.00
b	0.9	0.12	7.8	0.00

RSE	MR^2	F(1/48)	pv F	W_n
0.49	0.56	60.6	0.00	0.20

The test based on W_n detects the structural change at the level 5%. Note the two distinct directions on the graphic due to the different slopes in the model.

Simulation 4. The model is the same as in simulation 3 but we double the number of observations, i.e. $n = 100$.



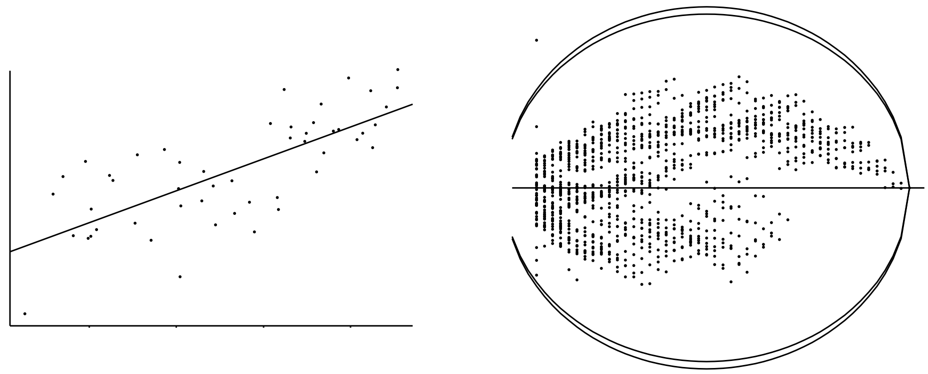
Simulation 4

	est	std	St t	pv t
a	0.4	0.19	2.5	0.01
b	1.2	0.07	18.1	0.00

RSE	MR^2	F(1/98)	pv F	W_n
0.47	0.77	328.5	0.00	0.31

As we should expect, the conclusion in this simulation is clearer than in the previous one.

Simulation 5. The model is the same as in simulation 3, but the structural change is smaller since now $b_2 = 1.1$. The number of observation is $n = 50$.



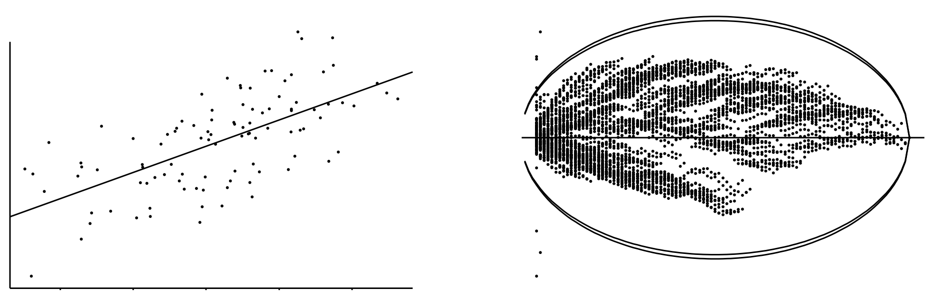
Simulation 5

	est	std	St t	pv t
a	0.2	0.44	0.5	0.63
b	1.2	0.15	7.5	0.00

RSE	MR^2	F(1/48)	pv F	W_n
0.66	0.54	57.3	0.00	0.14

For such a small sample size and small change, the test based on W_n fails to detect the erroneous specification at the level 5%. However, the plot of the process $b_n(\cdot)$ is still very informative and allows to suspect the structural change.

Simulation 6. The models and parameters are those of simulation 5, but we double the number of observations; that is $n = 100$ now.



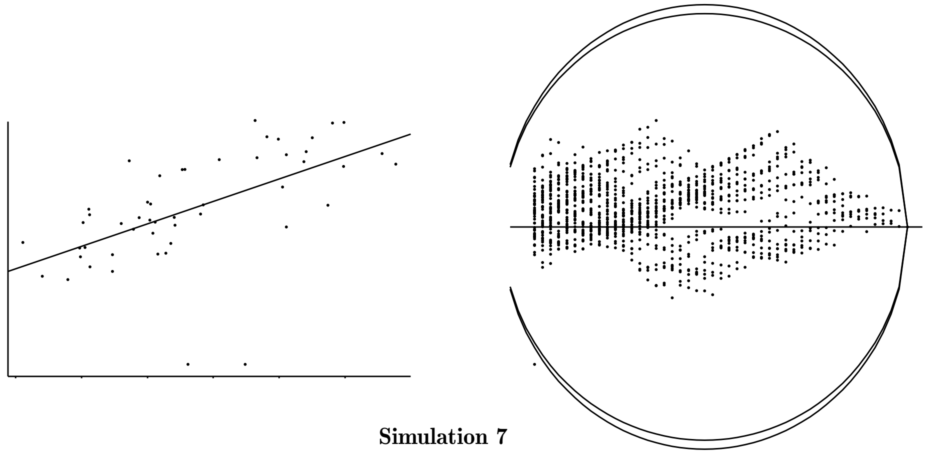
Simulation 6

	est	std	St t	pv t
a	0.8	0.29	2.9	0.00
b	0.9	0.10	8.4	0.00

RSE	MR^2	F(1/98)	pv F	W_n
0.62	0.42	69.8	0.00	0.20

The test based on W_n detects the wrong specification of the model.

Simulation 7. We consider the same model as in simulation 1, but with two outliers, for $t = 8$ and $t = 9$.



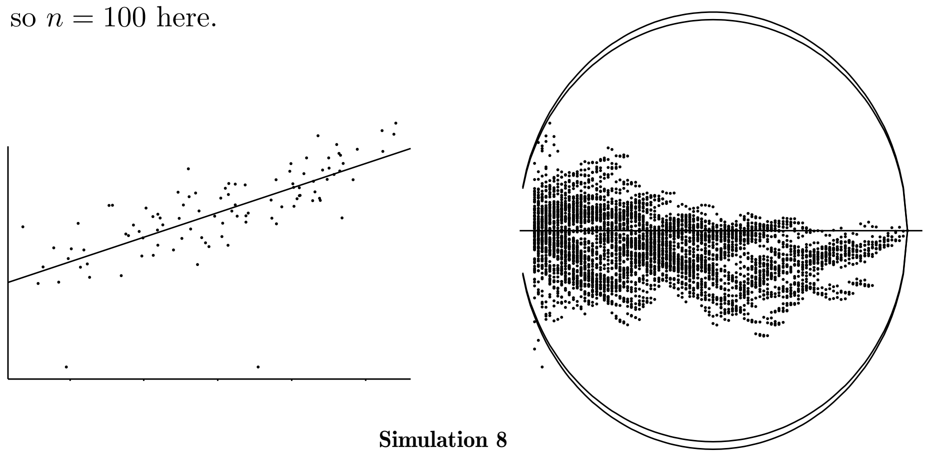
Simulation 7

	est	std	St t	pv t
a	1.1	0.45	2.3	0.02
b	0.9	0.18	5.1	0.00

RSE	MR^2	F(1/48)	pv F	W_n
0.94	0.35	25.6	0.00	0.05

The test based on W_n does not detect the misspecification. This may be due to the largest estimator of the variance created by the outliers. This simulation suggests that it may be more appropriate to use a robust estimator of the variance to build the test. Indeed, using a truncated estimator of the variance (eliminating the 5% higher observations) , we obtain $W_n = 0.31$ and we reject the linear specification.

Simulation 8. We double the number of points compare to simulation 7; so $n = 100$ here.



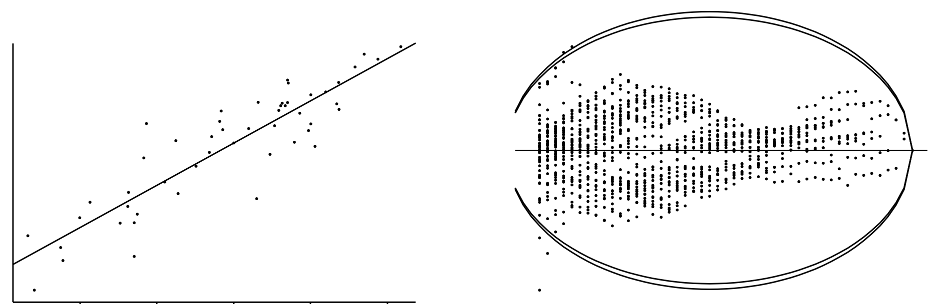
Simulation 8

	est	std	St t	pv t
a	0.7	0.27	2.6	0.01
b	1.1	0.10	10.8	0.00

RSE	MR^2	F(1/98)	pv F	W_n
0.65	0.54	115.6	0.00	0.06

The conclusion of simulation 7 apply to simulation 8. However, the graph of the process $b_n(\cdot)$ is very informative. Note the direction corresponding to the estimation of the slope when the outliers are not in $[\,s,t\,]$ and the extra trajectories close the the vertical axis at $t-s=0$; these extra trajectories correspond to the impact of outliers. When using the truncated estimator of the variance, we find $W_n = 0.48$ and reject a proper specification of the model.

Simulation 9. In this simulation we introduce a structural change on the boundary at $t = n/10$, with $n = 50$ observations.

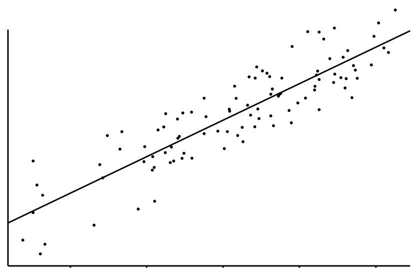


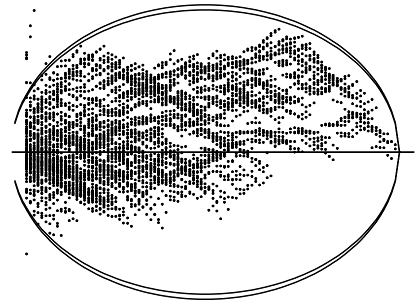
Simulation 9

	est	std.	St t	pv t	RSE	MR^2	F(1/48)	pv F	W_n
a	0.1	0.30	0.2	0.85	0.5	0.80	197.2	0.00	0.10
b	1.7	0.12	14.0	0.00					

The change is too close to the boundary to be detected by the W_n statistics calculated on a so small number of observations. But again, the plot of the process $b_n(\cdot)$ for $s-t$ near 0 let us suspect that something may be wrong with the linear model.

Simulation 10. As in the previous simulation, we consider a linear model with structural change on the boundary at $t = n/20$ but for $n = 100$ observations.





Simulation 10

	est	std	St t	pv t
a	-0.3	0.22	-1.2	0.24
b	1.55	0.08	18.9	0.00

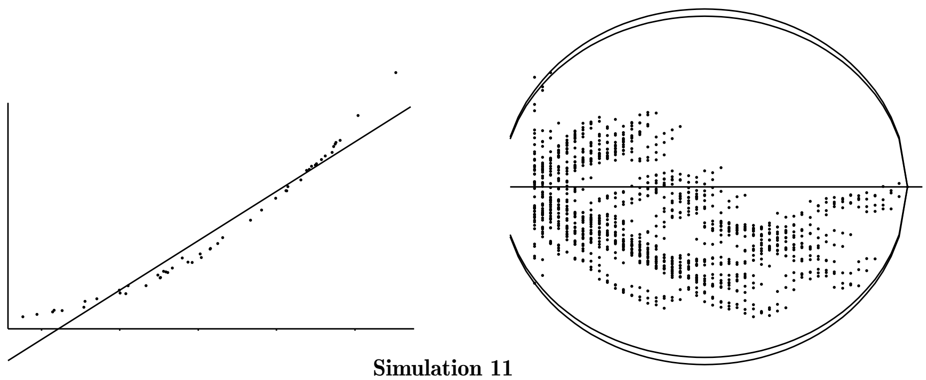
RSE	MR^2	F(1/98)	pv F	W_n
0.5	0.78	357.4	0.00	0.24

Compare to simulation 9, some more points allow to detect the misspecification, even though the change is relatively closer to the boundary. The conclusion is even clearer on the plot of the process $b_n(\cdot)$.

Simulation 11. In this simulation as well as in simulation 12, we introduce some non-linearities in the model with dependent data X_t having a temporal trend, $X_t = t + U_1, t$. The true model is

$$y_t = a + x_t^3 b + \epsilon_t \,, \qquad t = 1, \ldots, n$$

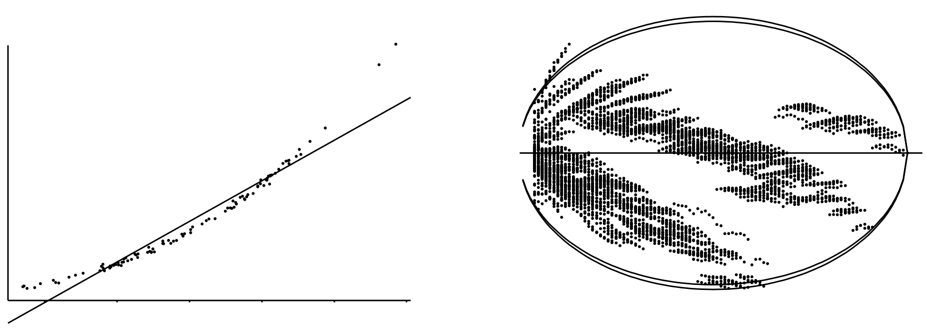
with the parameters of simulation 1. In that case we may see this non-linear model as a model with an infinity of structural break. This specification is directly covered by our theoretical analysis but is still interesting from a practical point of view.



Simulation 11

	est	std	St t	pv t	RSE	MR^2	F(1/48)	pv F	W_n
a	-31	1.78	-17.5	0.00	2.9	0.95	933.1	0.00	0.19
b	20.3	0.67	30.5	0.00					

The test based on W_n rejects the linearity hypothesis; this seems quite obvious if we look at the plot of the original data. The trajectories of $b_n(\cdot)$ clearly show that the link is polynomial. Indeed, we see the non-linear relationship as a succession of changes in the slopes since on short period, the curve is almost linear. However it does not mean that our test will systematically detect non-linear relationship, which is another problem. In this case it is detected because the observations are (almost) ordered.
Simulation 12. Same model as in simulation 11, but with $n = 100$.



Simulation 12

	est	std	St t	pv t	RSE	MR^2	F(1/98)	pv F	W_n
a	-29	1.29	-22.5	0.00	2.9	0.94	1443.9	0.00	0.26
b	19.2	0.50	38.0	0.00					

Other simulations have been performed in similar situations but also in combined situations (structural changes plus outliers, structural change plus non-linearities, outliers and non-linearities,...) leading to the same kind of conclusions. W_n easily test for the stability of the linear relationship, except for very small sample size and the graphic of the $b_n(t,s)$, among other things identify the source of the instability.

4. Proof of Theorem 2.1. In order to prove Theorem 2.1, we define the $[nt] \times d$ matrix $X_n(t)$ whose i-th row is X_i'. Similarly we define $\epsilon_n(t)$ to be the $[nt] \times 1$ vector whose i-th component is ϵ_i. Set

$$Z_n(t) = n^{-1/2} X_n(t)' \varepsilon_n(t)\,.$$

The proof is captured in several steps. We first show that the process $Z_n(\cdot)$ may be approximated by a suitably standardized Brownian motion. Then, we show that the approximation can be strengthened into a weighted approximation for the process $Z_n(t) - Z_n(s)$ defined over Π. Defining

$$X_n(s,t) = (X'_{[ns]}, \ldots, X'_{[nt]})'\,.$$

we then prove that the process

$$U_n(t,s) = \frac{X_n(s,t)' X_n(s,t)}{n(t-s)}$$

suitably normalized converges to a deterministic process. Since under the linearity assumption

$$n^{1/2}(t-s)(b_n(s,t) - b) = n^{-1/2} U_n(s,t)^{-1} \big(Z_n(t) - Z_n(s) \big) + o_p(1)\,,$$

this will give a good approximation for $b_n(s,t)$ from which the theorem will be derived.

Lemma 1. *As random variables in* $(\mathrm{Bd}(\Pi), ||\cdot||_\infty)$, *the processes* $Z_n(\cdot)$ *converge weakly to* $\Sigma^{1/2} W_p(\cdot)$ *in probability conditionally on* $(X_i)_{i\geq 1}$.

Proof. The proof of the convergence of the finite dimensional distributions is straightforward. To prove the tightness, it suffices to prove that the components of the vector $Z_n(\cdot)$ are tight in probability conditionally on $(X_i)_{i\geq 1}$. Since we are dealing with the components, we can assume without any loss of generality that $p = 1$. Define

$$\sigma_n^2 = n^{-1} \sum_{1\leq i\leq n} \mathrm{Var}(\epsilon_i | X_i)\,,$$

and set $\zeta_{i,n} = X_i \epsilon_i / \sigma_n$. With this notation,

$$Z_n(t) = \frac{\sigma_n}{\sqrt{n}} \sum_{i\leq nt} \zeta_{i,n}\,.$$

Assumption A1 and the strong law of large numbers ensure that

$$\lim_{n\to\infty} \sigma_n^2 = E\mathrm{Var}(\epsilon_1|X_1)$$

is finite. Moreover, A2 implies

$$\lim_{n\to\infty} \max_{1\leq i\leq n} n^{-1}\mathrm{Var}(\epsilon_i|X_i) = 0 \quad \text{a.s.}$$

Let

$$g_n(\epsilon) = n^{-1} \sum_{1\leq i\leq n} E\big(\mathbf{1}_{[\sqrt{n}\epsilon,\infty)}(X_i^2\epsilon_i^2)X_i^2\epsilon_i^2|X_i\big)\,.$$

Clearly, for any positive M and n large enough,

$$g_n(\epsilon) \leq n^{-1} \sum_{1\leq i\leq n} E\big(\mathbf{1}_{[M\epsilon,\infty)}(X_i^2\epsilon_i^2)X_i^2\epsilon_i^2|X_i\big)\,.$$

This upper bound converges almost surely to

$$E\big(\mathbf{1}_{[M\epsilon,\infty)}(X_1^2\epsilon_1^2)X_1^2\epsilon_1^2|X_1\big) = E\big(X_1^2 E(\mathbf{1}_{[M\epsilon,\infty)}(X_1\epsilon_1^2)\epsilon_1^2|X_1)\big)\,,$$

which converges to 0 as M tends to infinity. Therefore, $\lim_{n\to\infty} g_n(\epsilon) = 0$ a.s. Apply Theorem 42.2.C in Loève (1977) to obtain that the distribution of $Z_n(\cdot)$ converges weakly to that of a Wiener process in probability conditionally on $(X_i)_{i\geq 1}$, and thus is tight in conditional probability. ∎

The following is an immediate consequence of Lemma 1.

Lemma 2. *As random variables in* $(\mathrm{Bd}(\Pi), |\cdot|_\infty)$*, the processes* $Z_n(t) - Z_n(s)$ *converge weakly to* $\Sigma^{1/2}\big(W_p(t) - W_p(t)\big)$*, in probability conditionally on* $(X_i)_{i\geq 1}$*.*

Observe now that for any fixed s and t, $U_n(s,t)$ converges almost surely to U. Hence we can expect to approximate $n^{1/2}(t-s)(b_n(s,t) - b)$ by $U^{-1}\Sigma^{1/2}\,(W_d(t) - W_d(s))$. Unfortunately for any $1 \leq k \leq n$, $U_n(k/n, (k+1)/n) = V_k X_k'$ does not converges to U, so that the convergence of $U_n(s,t)$ to U is not uniform in s and t. To overcome this problem, write for any positive η less than $1/2$,

$$\begin{aligned} &U_n(s,t)^{-1}\big(Z_n(t) - Z_n(s)\big) \\ &\qquad = (t-s)^{1/2-\eta} U_n(s,t)^{-1}\Big(\big(Z_n(t) - Z_n(s)\big)/(t-s)^{1/2-\eta}\Big)\,. \end{aligned} \tag{4.1}$$

We shall prove that $(t-s)^{1/2-\eta}U_n(s,t)^{-1}$ behaves nicely, as well as $(Z_n(t) - Z_n(s))/(t-s)^{1/2-\eta}$.

Lemma 3. *Whenever η is positive less than $1/2$,*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{|t-s|<\delta}\frac{|Z_n(t)-Z_n(s)|}{(t-s)^{1/2-\eta}}=0 \quad \textit{in probability,}$$

in probability conditionally on $(X_i)_{i\geq 1}$.

Proof. It suffices to prove the result on every coordinate of the vector valued process Z_n. Thus, we can assume without any loss of generality that $p=1$. Write

$$\omega(Z_n,\delta)=\sup_{|t-s|<\delta}\frac{|Z_n(t)-Z_n(s)|}{(t-s)^{1/2-\eta}}\,.$$

The weighted approximation of Csörgő, Csörgő, Horváth and Mason (1986) implies

$$\lim_{\delta\to 0}\limsup_{n\to\infty}P\{\,\omega(Z_n,\delta)\geq\epsilon\,\}=0\,.$$

Since

$$P\{\,\omega(Z_n,\delta)\geq\epsilon\,\}=EP\{\,\omega(Z_n,\delta)\geq\epsilon\,|\,X\,\}$$

the random variables

$$P\{\,\omega(Z_n,\delta)\geq\epsilon\,|\,X\,\}$$

converge to 0 in L^1, and therefore in probability. ∎

Lemma 4. *For a symmetric matrix M, write $\lambda_{\min}(M)$ its smallest eigenvalue. Assume that $E|X|^\beta$ is finite. If there exists two real numbers c and δ such that for any unit vector u*

$$F_u(\epsilon)=P\{\,|u'X_1|\leq\epsilon\,\}\leq c\epsilon^\rho\,,$$

then for any $d>p$,

$$P\{\,\lambda_{\min}(X_1'X_1+\cdots+X_d'X_d)\leq\epsilon\,\}\leq c\epsilon^{\beta(\rho d+2-2p)/(2\beta+4(p-1))}\,.$$

Proof. Let S_{p-1} denote the unite sphere centered at the origin in $\mathbf{R}^p$. The left hand side in the conclusion of the Lemma is equal to

$$P\{\,\exists u\in S_{p-1}\,:\,\sum_{1\leq i\leq d}u'X_iX_i'u\leq\epsilon\,\}\,.$$

Since $u'X_iX_i'u=|u'X_i|^2$, for any nonnegative number M, the above probability is less than

$$P\{\,\exists u\in S_{p-1}\,:\,\sum_{1\leq i\leq d}|u'X_i|^2\leq\epsilon\,;\,\max_{1\leq i\leq d}|X_i|\leq M\,\}+P\{\,\max_{1\leq i\leq d}|X_i|\geq M\,\}\,.$$

The moment assumption in the Lemma yields

$$P\{\max_{1\leq i\leq d}|X_i|\geq M\}\leq dP\{|X_i|\geq M\}\leq d\frac{E|X|^\beta}{M^\beta}\,.$$

If $|X_i|$ is less than M, observe that for any unit vectors u and v,

$$\begin{aligned}|(u_i'X_i)^2-(v'X_i)^2| &= |(u'-v')X_i(u'+v')X_i|\\ &\leq |u-v||X_i|(|u|+|v|)|X_i|\\ &\leq 2|u-v|M^2\end{aligned}$$

Therefore, if u and v are distant apart of $\epsilon/4M^2$, then $(u'X_i)^2$ and $(v'X_i)^2$ are apart of at most $\epsilon/2$.

Let $N(S_{p-1},\eta)$ be the smallest number of points in S_{p-1} such that caps of radius η at those points cover S_{p-1}. We then have

$$\begin{aligned}&P\{\exists u\in S_{p-1}\,:\,\sum_{1\leq i\leq d}|u'X_i|^2\leq\epsilon\,;\,\max_{1\leq i\leq d}|X_i|\leq M\}\\ &\qquad\leq N(S_{p-1},\epsilon/4M^2)\max_{u\in S_{p-1}}P\{\sum_{1\leq i\leq d}(u'X_i)^2\leq\epsilon/2\}\,.\end{aligned}$$

Since the inequality $\sum_{1\leq i\leq d}(u'X_i)^2\leq\epsilon/2$ implies that every summand $(u'X_i)^2$ is less than $\epsilon/2$ as well, we conclude that

$$\begin{aligned}&P\{\lambda_{\min}(X_1X_1'+\cdots+X_pX_p')\leq\epsilon\}\\ &\qquad\leq N(S_{p-1},\epsilon/4M^2)\max_{u\in S_{p-1}}P\{|u'X_1|\leq\sqrt{\epsilon/2}\}^d+\frac{dE|X_1|^\beta}{M^\beta}\,.\end{aligned}$$

We now claim that there exists a constant c such that $N(S_{p-1},\eta)\leq c\eta^{p-1}$. Indeed, the parameterizations of the northern hemisphere of S_{p-1} in polar coordinates realizes a diffeomorphism with $[0,2\pi]^{p-2}\times[0,\pi]$. The parameter set can be covered by $O(\epsilon^{p-1})$ cubes of size ϵ (as ϵ tends to 0), and the parameterizations has uniformly bounded Jacobian over the parameter set (because the sine and cosine functions have bounded derivative). The claim follows.

Using our claim, the previous inequality and the assumption of the Lemma, we conclude that for a constant c,

$$P\{\lambda_{\min}(X_1X_1'+\cdots+X_pX_p'|\leq\epsilon\}\leq c\Big(\Big(\frac{M^{2(p-1)}}{\epsilon^{p-1}}\Big)\epsilon^{\rho d/2}+\frac{1}{M^\beta}\Big)\,.$$

Optimizing in M yields the result. ∎

The key point of the Lemma is that for $d>2(p-1)/\rho$, the lower tail of the distribution of the smallest eigenvalue of the sum decays like a power of ϵ. In typical application we will require $\beta > 3$ and we will have $\rho = 1$. In this case, the exponent of ϵ is at least $3\big(d - 2(p-1)\big)/\big(6 + 4(p-1)\big)$, which is positive as soon as $d > 2(p-1)$.

We now show that $(t-s)^{1/2-\eta} U_n(s,t)^{-1}$ is uniformly bounded. In the following $\|\cdot\|$ denotes the operator norm, that is the largest singular value of the operator.

Lemma 5. *If the conclusion of Lemma 4 holds and*

$$\Big(\Big(\frac{3}{2} - \eta\Big)\frac{2}{\beta} - \frac{1}{2}\Big)\beta \frac{\delta d + 2 - 2p}{2\beta + 4(p-1)} < -1 \,. \tag{4.2}$$

For any positive δ,

$$\limsup_{n\to\infty} \sup_{|t-s|<\delta} \|(t-s)^{1/2-\eta} U_n(s,t)^{-1}\|$$

is a stochastically bounded sequence.

Proof. Let ϵ be a positive real number and define the event

$$A_n(\epsilon) = \{ \sup_{d/n \le t-s \le \delta} (t-s)^{1/2-\eta} \|U_n(s,t)^{-1}\| \ge 1/\epsilon \} \,.$$

Recall that if M is invertible, then the norm of its inverse is the inverse of its smallest eigenvalue, that is

$$\|M^{-1}\| = 1/ \inf_{|u|=1} |Mu| = 1/\lambda_{\min}(M) \,.$$

This implies that the event $A(\epsilon)$ is

$$\big\{ \exists s,t \,,\, d/n \le t-s \le \delta n \,,\, \lambda_{\min}\big(X_n(s,t)' X_n(s,t)\big) \le \epsilon n (t-s)^{3/2-\eta} \big\} \,.$$

Hence, we have the equality

$$\begin{aligned} A(\epsilon) = \Big\{ & \exists 1 \le i \le n \,, \exists d \le k \le n\delta \,,\, i+k \le n \\ & \qquad \lambda_{\min}\Big(\sum_{i \le \ell \le i+k} X_\ell X_\ell' \Big) \le \epsilon k^{3/2-\eta}/n^{1/2-\eta} \Big\} \,. \end{aligned}$$

In the sequel, when it is convenient, we write $\Gamma_{p,q}(M)$ the (p,q) entry of a matrix M.

From the strong invariance principle of Komlós, Major and Tusnády (1975, 1976) and Major (1976), we infer that up to enlarging the probability space, there exists Wiener processes $W_{i,j}$, such that

$$\sup_{0 \le t \le n} \Big| \Gamma_{p,q}\Big(\sum_{\ell \le nt} X_\ell X_\ell' \Big) - t\Gamma_{p,q} E X_1 X_1' - W_{p,q}(t) \Big| = O(n^{2/\beta})$$

almost surely. This implies that

$$\Gamma_{i,j}\Big(\sum_{i\leq \ell\leq i+k} X_\ell X_\ell'\Big) = k\Gamma_{p,q}(EX_1X_1') + W_{p,q}(i+k) - W_{p,q}(i) + O(n^{2/\beta})$$

uniformly in $1\leq i\leq n$ and $1\leq k\leq n$ with $i+k\leq n$. Therefore, since the matrices $X_\ell X_\ell'$ are of fixed finite size and the map associating to a matrix its smallest eigenvalue is continuous, the probability of the event

$$A_1(\epsilon) = \Big\{ \exists 1\leq i\leq n\,,\ \exists M n^{2/\beta}\leq k\leq n\delta\,,\ i+k\leq n \\ \lambda_{\min}\Big(\sum_{i\leq \ell\leq i+k} X_\ell X_\ell'\Big) \leq \epsilon k^{3/2-\eta}/n^{1/2-\eta}\Big\}$$

can be made arbitrary small by choosing M large enough. Consider the event

$$A_2(\epsilon) = \Big\{ \exists 1\leq i\leq n \,:\, \exists d\leq k\leq M n^{2/\beta}\,,\ i+k\leq n\,, \\ \lambda_{\min}\Big(\sum_{i\leq \ell\leq i+k} X_\ell X_\ell'\Big) \leq \epsilon k^{3/2-\eta}/n^{1/2-\eta}\Big\}\,.$$

Clearly, $A(\epsilon) = A_1(\epsilon)\cup A_2(\epsilon)$. So it suffices to prove that in choosing ϵ large enough, no matter what M is, we can make the probability of $A_2(\epsilon)$ as small as desired. Observe that $P\big(A_2(\epsilon)\big)$ is less than

$$nP\Big\{ \exists d\leq k\leq M n^{2/\beta} \,:\, \lambda_{\min}\Big(\sum_{1\leq \ell\leq k} X_\ell X_\ell'\Big) \leq \epsilon k^{3/2-\eta}/n^{1/2-\eta}\Big\}\,.$$

Define $Z_j = \sum_{1\leq \ell\leq d} X_{mj+\ell}X_{mj+\ell}'$. Write any integer k as $k = md + r$, with $r < d$. Then

$$\sum_{1\leq \ell\leq k} X_\ell X_\ell' = \sum_{0\leq j\leq m} Z_j + \sum_{1\leq \ell\leq r} X_{md+\ell}X_{md+\ell}'\,.$$

Observe that the Z_j's are i.i.d. symmetric matrices almost surely nonnegative. Now, if M_1 and M_2 are two symmetric matrices, the smallest eigenvalue of their sum is larger than the sum of their smallest eigenvalue, *viz.*

$$\lambda_{\min}(M_1+M_2)\geq \lambda_{\min}(M_1)+\lambda_{\min}(M_2)$$

as the variational formula

$$\lambda_{\min}(M_1+M_2) = \inf_{|u|=1} u'(M_1+M_2)u$$

shows. Therefore

$$
\begin{aligned}
&P\Big\{\,\exists\, k \le Mn^{2/\beta} \,:\, \lambda_{\min}\Big(\sum_{1\le \ell\le k} X_\ell X_\ell'\Big) \le \epsilon k^{3/2-\eta}/n^{1/2-\eta}\,\Big\} \\
&\quad \le P\big\{\,\exists 1 \le m \le Mn^{2/\beta} \,:\, \sum_{1\le j\le m} \lambda_{\min}(Z_j) \le \epsilon\big((m+1)d\big)^{3/2-\eta}/n^{1/2-\eta}\,\big\}\,.
\end{aligned}
$$

For $m \ge 1$, the ratio $(m+1)/m$ is at most 2. Thus, the above probability is less than

$$
\sum_{1\le m\le Mn^{2/\beta}} P\Big\{\,\sum_{1\le j\le m} \lambda_{\min}(Z_j) \le 4\epsilon d^{3/2-\eta} m^{3/2-\eta}/n^{1/2-\eta}\,\Big\}\,. \tag{4.3}
$$

If the sum of the positive r.v.'s $\lambda_{\min}(Z_j)$ is less than a given number, then certainly all the variables in the sum are less than that number. Therefore (4.3) is less than

$$
\begin{aligned}
&\sum_{1\le m\le Mn^{2/\beta}} P\big\{\,\lambda_{\min}(Z_1) \le 4\epsilon d^{3/2-\eta} m^{3/2-\eta}/n^{1/2-\eta}\,\big\}^m \\
&\qquad \le \sum_{1\le m\le Mn^{2/\beta}} P\big\{\,\lambda_{\min}(Z_1) \le 4\epsilon (dM)^{(\frac{3}{2}-\eta)} n^{(\frac{3}{2}-\eta)\frac{2}{\beta}-\frac{1}{2}+\eta}\,\big\}^m\,. \\
&\qquad \le 2P\big\{\,\lambda_{\min}(Z_1) \le 4\epsilon (dM)^{\frac{3}{2}-\eta} n^{(\frac{3}{2}-\eta)\frac{2}{\beta}-\frac{1}{2}+\eta}\,\big\}
\end{aligned}
$$

the last inequality coming from summing a geometric series which is less than $1/2$ thanks to Lemma 4. Applying Lemma 4, we see that this upper bound is of order

$$
n^{((\frac{3}{2}-\eta)\frac{2}{\beta}-\frac{1}{2}+\eta)\beta\frac{\delta d+2-2p}{2\beta+4(p-1)}}\,.
$$

The condition on β in the statement of the Lemma ensures that the above is $o(n^{-1})$. Therefore, $P\big(A_2(\epsilon)\big) = o(1)$ as n tends to infinity, no matter what M. This concludes the proof. ∎

We now conclude the proof of (2.2). For $\eta = 0$ and $\delta = 1$, inequality (4.2) is equivalent to (2.2). Thus, we can find a positive η such that (4.2) holds. Combining (4.1), Lemmas 2, 3 and 5 yields (2.2). Then (2.3) follows since $\hat{b}_n = b_n(0,1)$.

5. Proof of Theorem 2.2. Define

$$
\begin{aligned}
\tilde{Z}_n(t) &= n^{-1/2} \sum_{i\le nt} \big(n^{-1/2} X_i' m(X_i, i/n) + X_i'\epsilon_i\big) \\
&= n^{-1} \sum_{i\le nt} X_i' m(X_i, i/n) + Z_n(t)\,.
\end{aligned}
$$

Using the arguments of Lemma 1, the process $n^{1/2}(t-s)\big(b_n(s,t)-b\big)$ has the representation

$$n^{-1/2}U_n(s,t)^{-1}\Big(Z_n(t)-Z_n(s)+n^{-1}\sum_{ns<i\leq nt}X_i'm(X_i,i/n)\Big)+o_P(1)\,.$$

The same argument used to prove Theorem 1 give Theorem 2 provided we can show that

$$\lim_{n\to\infty}\sup_{0\leq s<t<1}\Big|n^{-1}\sum_{ns<i\leq nt}X_im(X_i,i/n)-\int_s^t EX_1m(X_1,u)\,\mathrm{d}u\Big|=0 \quad (5.1)$$

in probability, and (compare with Lemma 3)

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{d/n\leq t-s\leq\delta}\Big|n^{-1}(t-s)^{\eta-1/2}\sum_{ns<i\leq nt}X_im(X_i,i/n)\Big|=0\,. \quad (5.2)$$

in probability. Define $\Sigma_{m,u}=EX_1X_1'm(X,u)^2$ the covariance matrix of $X_1m(X_1,u)$. This covariance is well defined since $m(\cdot,\cdot)$ is bounded and X_1 is square integrable. Set $\Sigma_m=\int_0^1\Sigma_{m,u}\,\mathrm{d}u$. Since the function $m(\cdot,\cdot)$ is bounded and A5 hold, the random variables $X_im(X_i,i/n)$ have moment of order 6. This garantees that they satisfy the Lindeberg condition. Then Prokhorov's (1965) version of Donsker's theorem implies that the process

$$n^{-1}\sum_{1\leq i\leq nt}\big(X_im(X_i,i/n)-EX_im(X_i,i/n)\big)$$

converge in distribution to $\Sigma_mW_p(\cdot)$. We also have

$$\begin{aligned}n^{-1/2}\sum_{1\leq i\leq nt}&\Big|EX_im(X_i,i/n)-\int_{(i-1)/n}^{i/n}EX_1m(X_1,u)\,\mathrm{d}u\Big|\\&\leq n^{-1/2}\sum_{1\leq i\leq n}n^{-1}E|X_1\sup_{0\leq u\leq 1}\partial_um(X_1,u)|=o(1)\end{aligned}$$

as n tends to infinity, thanks to (A7). This proves (5.1).

To prove (5.2), we write

$$\begin{aligned}&\frac{1}{n(t-s)^{1/2-\eta}}\sum_{ns<i\leq nt}X_im(X_i,i/n)\\&\quad=\frac{1}{n^{1/2}(t-s)^{1/2-\eta}}\frac{1}{\sqrt{n}}\sum_{ns<i\leq nt}\big(X_im(X_i,i/n)-EX_im(X_i,i/n)\big)\\&\qquad+\frac{1}{n(t-s)^{1/2-\eta}}\sum_{ns<i\leq nt}EX_im(X_i,i/n)\,. \quad (5.3)\end{aligned}$$

Since $t-s \geq d/n$, the inequality $n^{1/2}(t-s)^{1/2-\eta} \geq n^{\eta} d^{1/2-\eta}$ holds. Hence, the first term in the right hand side of (5.3) is at most $n^{-\eta} O_P(1)$ and converges to 0 uniformly in s and t with $t-s > d/n$. The second term is

$$\frac{1}{(t-s)^{1/2-\eta}} \int_s^t EX_1 m(X_1, u)\, \mathrm{d}u \\ + \frac{O(1)}{n^2 (t-s)^{1/2-\eta}} \sum_{ns < i \leq nt} E|X_i \sup_{0 \leq u \leq 1} \partial_u m(X_i, u)| \\ = \frac{1}{(t-s)^{1/2-\eta}} \int_s^t EX_1 m(X_1, u)\, \mathrm{d}u + O(1)(t-s)^{1/2+\eta}\,.$$

The convergence in (5.2) follows. ∎

References.

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

Andrews, D.W.K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, **62**, 1383–1414.

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

Basseville, M and Nikiforov, I. (1993). *Detection of abrupt changes: theory and application*. Information and system science series. Prentice Hall, Englewood Cliffs.

Belsley, D.A., Kuh E., Welsh, R.E. (1980). *Regression diagnostics: identifying influential data and sources of colinearity*, Wiley.

Carlstein, E. (1988). Nonparametric estimation of a change-point, *Ann. Statist.*, **16**, 188–197.

Chow, G.C. (1960). Tests of the equality between two sets of coefficients in two linear regressions, *Econometrica*, **28**, 561–605.

Csörgö, M., Csögö, S., Horvaŕh, L., Mason, D.M. (1976). Weighted empirical and quantile processes, *Ann. Probab.*, 14, 31–85.

Darkhovshk, B.S. (1976). A non-parametric method for the a posteriori detection of the disorder time of a sequence of independent random variables, *Theor. Probab. Appl.*, **21**, 178–183.

Doukhan, P. (1994). Mixing: properties and examples, *Lecture Notes in Statistics*, 85, Springer-Verlag.

Dümbgen, L. (1991). The asymptotic behavior of some non-parametric change-point estimators, *Ann. Statist.*, **19**, 1471–1495.

Freedman, D. (1981). Bootstrapping regression models, *Ann. Statist.*, **9**, 1218–1228.

Freedman, D. (1984). On Bootstrapping two-stage least squares estimates in stationary linear model, *Ann. Statist.*, **12**, 827–842.

Hrdle, W. and Mammen, E. (1993). Comparing Nonparametric versus Parametric regression fit. *Ann. Statist.*, **21**, 1926-1947.

Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample df., I, *Z. Wahrsch. verw. Geb.*, **32**, 111–131.

Komlós, J., Major, P., Tusnády, G. (1976). An approximation of partial sums of independent r.v.'s and the sample df., II, *Z. Wahrsch. verw. Geb.*, **34**, 33–58.

Loève, M. (1977). *Probability Theory I, II*, 4th ed., Springer.

Major, P. (1976). The approximation of partial sums of independent r.v.'s, *Z. Wahrsch. verw. Geb.*, **35**, 213–220.

Phillips, P.C.B. and Durlauf, S.N. (1986). Multiple Time series regression with integrated processes, *Review of Economic Studies*, **53**, 473–495.

Politis, D.N. and Romano, J.P.(1994). Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions, *Ann. Statist.*, **22**, 2031–2050.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer.

Prokhorov, Y.V. (1956). Convergence of random processes and limit theorems in probability theory, *Theor. Probab. Appl.*, **1**, 157–214.

Quandt, R. (1960). Tests of the hypothesis that a linear regression obeys two separate regimes, *J.A.S.A.*, **55**, 324–330.

Shorack G.R., Wellner J. (1986). *Empirical Processes with Applications to Statistics*, Wiley.

Spokoiny, V. (2001). Data driven testing the fit of linear models. *Math. Methods Statist.*, **10**, no. 4 (2001), pp. 465 - 497.

Stute, W. (1997). Nonparametric Model Checks for regression. *Ann. Statist.*, **25**, 613-641.

Watson, G.S. (1961) Goodness of fit tests on circle, *Biometrika*, **48**, 109–114.

Watson, G.S. (1967) Another test for the uniformity of a circular distribution, *Biometrika*, **54**, 675–676

Ph. Barbe	P. Bertail
90, Rue de Vaugirard	CREST, LS Timbre J340
75006 PARIS	3, av. Pierre Larousse
FRANCE	92245 MALAKOFF
	FRANCE
	Patrice.Bertail@ensae.fr