working paper department of economics

Testing for Parameter Constancy
in Linear Regressions:
Empirical Distribution Function Approach

Jushan Bai

No. 93-9 July 1993

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

Testing for Parameter Constancy in Linear Regressions: Empirical Distribution Function Approach

Jushan Bai

No. 93-9 July 1993

Testing for Parameter Constancy in Linear Regressions: Empirical Distribution Function Approach

Jushan Bai
Department of Economics
Massachusetts Institute of Technology

July 29, 1993

Abstract

This paper proposes some tests for parameter constancy in linear regression models with possible infinite variance. Both dynamic and trending regressors are allowed. The tests are based on the empirical distribution function of estimated residuals and are shown to have non-trivial local power against a wide range of alternatives. Within a certain class of alternatives including simple shifts, the tests have higher power for testing the simple shift alternatives. These tests are formulated in such a way that the limiting variables are distribution-free. The residuals may be obtained based on any root-n consistent estimator (under the null) of regression parameters. As part of these results, some weak convergence for weighted sequential empirical processes of residuals is established.

Key words and phrases: structural change, empirical distribution function, weak convergence, nonparametric test, fluctuation test, CUSUM test.

1 Introduction

There are various sources in economics that could cause a parametric model to be unstable over a period of time. Changes in taste, technical progress, and changes in policies and regulations all are such examples. A change in the economic agent's expectation can induce a change in the reduced-form relationship among economic variables, even though no change in the parameters of the structural relationship is present, as envisioned by the Lucas critique. The shifts of the Phillips curve over time perhaps serve as the best illustration (Alogoskoufis and Smith, 1991). As a result, model stability has always been an important concern in econometric modeling. Earlier studies of parameter constancy include Chow (1960) and Quandt (1960). As perhaps a consequence of diagnostic failures, models capable of handling parameter instability have constantly spawned out. The random-coefficients model of Cooley and Prescott (1973), for instance, the switching-regression models of Goldfeld and Quandt (1973a,b) and numerous others find widespread use in economics. The purpose of this paper is to provide additional tools for the diagnosis of parameter instability in linear regressions.

Recent work in econometrics on this topic has been directed toward detecting parameter changes occurring at an unknown time, hardly a new problem given the large body of related literature in econometrics and statistics. Econometricians are particularly concerned with parameter instability in dynamic models with trending regressors, cointegrated variables and perhaps a unit root, and with serially correlated or heteroskedastic disturbances. Various test statistics that are capable of detecting changes in those situations have been developed; see, for example, Andrews (1990), Chu and White (1992), Hansen (1992), Perron (1991), and Ploberger, Kramer and Kontrus (1989). Empirical applications together with supporting theory can be found in Bai, Lumsdaine and Stock (1991), Banerjee, Lumsdaine and Stock (1992), Christiano (1992), Perron (1989), and Zivot and Andrews (1992), among others. In this paper, we propose some tests able to detect structural instability for some of these models. In addition, these tests are applicable to infinite variance regressions.

Two classes of tests are proposed, resembling the prototypical Kolmogorov-Smirnov two-sample test. The first class is based on non-weighted sequential empirical processes of residuals. This class was previously considered by Picard (1985) and Csörgő and Horváth (1988), among others. However, these authors only consider the case of i.i.d. observations under the null. We extend the tests to apply to regression models with estimated parameters.

The first class of tests has limited applicability in time series econometrics since if trending regressors are included in the regression model, the tests will no longer be asymptotically distribution-free. In this case, the second class of tests can be considered, obtained by constructing a weighted empirical process of residuals with weights equal to the regressors, apart from some weighting matrices. This class of tests is asymptotically distribution-free whether or not a trend regressor is present. Our procedure may be regarded as nonparametric, yet it is not fully nonparametric in light of the need to estimate the regression parameters. By way of construction, our tests are robust against heavy-tailed distributions and data aberrations. Recent work of Carlstein (1988) and Duembgen (1991), who consider the estimation of the shift point under the single shift alternative (two i.i.d. samples) and obtain good convergence rate, indicates that the tests of the Kolmogorov-Smirnov type may be powerful.

The classical statistical literature shows that goodness-of-fit tests based on empirical processes involving estimated parameters will depend upon both the estimated parameters and the underlying error-distribution function even in the limit (see Durbin (1973)). It is somewhat surprising that we can eliminate this dependence by choosing weighting vectors, in a natural way, in the construction of the empirical processes upon which our tests are based, whereas classical goodness-of-fit tests can be made asymptotically distribution-free merely by essentially abandoning part of the observations (see Durbin (1976)).

The tests proposed in this paper are quite general in the sense that we require no finite variance for the disturbances. Both dynamic and trending regressors are allowed in the regressions. Within certain classes of alternatives, we show the tests to be more powerful when used for testing simple shift alternatives, as expected. Moreover, these tests exhibit non-trivial local power.

As a related result that may be of independent interest, a weak convergence for randomly-weighted sequential empirical processes has been obtained. We then use this result to obtain the weak convergence for its counterpart for the regression residuals, laying the theoretical foundation of our tests.

This paper is organized as follows. Section 2 specifies the models and describes the assumptions. Section 3 defines the test statistics. Section 4 provides alternative expressions for the test statistics that are suitable for computation. Section 5 examines the local power of the tests. Trending regressors are considered in Section 6. Some comments and possible extensions are discussed in Section 7. Technical materials are collected in the appendix.

2 Models and Assumptions

The regression model under the null hypothesis is

$$y_t = x_t'\beta + \varepsilon_t \quad (t = 1, 2, \ldots, n) \tag{1}$$

where y_t is an observation of the dependent variable, x_t is a $p \times 1$ vector of observations of the independent variables, ε_t is an unobservable stochastic disturbance, and β is the $p \times 1$ vector of regression coefficients.

The non-null hypothesis specifies the following model:

$$y_t = x_t'\beta_t + \varepsilon_t^*.$$

where the β_t may not be constant over time and the disturbances ε_t^* may not be identically distributed. In particular, we are interested in the following local alternatives.

i) Changing regression parameters: $\beta_t = \beta(1 + \Delta_1 g(t/n)n^{-1/2})$.

ii) Changing variance: $\varepsilon_t^* = \varepsilon_t(1 + \Delta_2 h(t/n)n^{-1/2})$.

iii) Both i) and ii).

where Δ_1 and Δ_2 are two real numbers; the functions h and g are assumed to be bounded.

In what follows, the notation $o_p(1)$ ($O_p(1)$) is used to denote a sequence of random variables converging to zero in probability (being stochastically bounded). The norm $\|\cdot\|$ represents the Euclidean norm, i.e. $\|x\| = (\sum_{i=1}^p x_i^2)^{1/2}$ for $x \in R^p$. Finally, $[\cdot]$ denotes the greatest integer function.

We impose the following assumptions:

(A.1) Under the null hypothesis, the ε_t are i.i.d. with distribution function (d.f.) F, which admits a density function $f, f > 0$. Both $f(x)$ and $xf(x)$ are assumed to be uniformly continuous on the real line. Furthermore, there exists a finite number L such that $|xf(x)| < L$ and $|f(x)| < L$ for all x. The mean of ε_t is zero if this mean exists.

(A.2) The disturbances ε_t are independent of all contemporaneous and past regressors.

(A.3) The regressors satisfy

$$\text{plim}\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' = sQ \quad \text{for } s \in [0,1]$$

where Q is a $p \times p$ nonrandom positive definite matrix. The convergence is necessarily uniform in s, because the sum is "monotonic" in s.

(A.4)

$$\max_{1\le t\le n} n^{-1/2}\|x_t\| = o_p(1)$$

(A.5) For every fixed s_1, there exists a sequence of positive random variables $Z_n = O_p(1)$ such that

$$\frac{1}{n}\sum_{t=[ns_1]}^{[ns]} \|x_t\| \le (s - s_1)Z_n \qquad a.s.$$

for all $s \ge s_1$. In addition, the tail probability of Z_n satisfies, for some $\rho > 0$:

$$P(|Z_n| > C) \le M/C^{2(1+\rho)}.$$

Note that Z_n may be taken to be $\max_k k^{-1}\sum_{t=i}^{i+k}\|x_t\|$ provided the condition on the tail probability is also satisfied, where $i = [ns_1]$ is fixed.

(A.6) There exist $\gamma > 1$, $\alpha > 1$ and $K < \infty$ such that for all $0 \le s' \le s'' \le 1$, and for all n,

$$\frac{1}{n}\sum_{i<t\le j} E(x_t'x_t)^\gamma \le K(s''-s') \text{ and } E(\frac{1}{n}\sum_{i<t\le j} x_t'x_t)^\gamma \le K(s''-s')^\alpha,$$

where $i = [ns']$, $j = [ns'']$. The assumption is satisfied if the x_t are bounded regressors. Also if $E(x_t'x_t)^2 \le M$ for all t, then the assumption is satisfied with $\gamma = 2$ and $\alpha = 2$, because $E(\sum_{t=i}^{j} x_t'x_t)^2 \le \{\sum_{t=i}^{j}[E(x_t'x_t)^2]^{1/2}\}^2$ by the Cauchy-Schwartz inequality.

(A.7)

$$(X'X)^{1/2}(\hat{\beta} - \beta) = O_p(1)$$

where $X = (x_1, x_2, ..., x_n)'$. When the disturbances are i.i.d. and have finite variance, then least squares estimator satisfies this assumption. For infinite variance models, robust estimation such as LAD method has to be used to assure (A.7).

(A.8) There exist a $\delta > 0$ and an $M < \infty$ such that

$$E(\frac{1}{n}\sum_{t=1}^{n} \|x_t\|^{3(1+\delta)}) < M \text{ and } E(\frac{1}{n}\sum_{t=1}^{n} \|x_t\|^3)^{1+\delta} < M \quad \forall n.$$

(A.9) Finally

$$\text{plim}\frac{1}{n}\sum_{t=1}^{[ns]} x_t = s\bar{x} \quad \text{uniformly in } s \in [0,1]$$

where $\bar{x}$ is a $p \times 1$ constant vector. When a constant regressor is included, (A.9) is implied by (A.3).

Assumptions (A.3) and (A.9) exclude trending regressors, which will be discussed in Section 5.

3 The Test Statistics

Let $\hat{\beta}$ be an estimator of β and put $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$. The test is based on the estimated residuals $\hat{\varepsilon}_t$. For each fixed k, define the empirical distribution function (e.d.f.) based on the the first k residuals as

$$\hat{F}_k(x) = \frac{1}{k}\sum_{t=1}^{k} I(\hat{\varepsilon}_t \leq x)$$

and the e.d.f. based on the last $n-k$ residuals as

$$\hat{F}^*_{n-k}(x) = \frac{1}{n-k}\sum_{t=k+1}^{n} I(\hat{\varepsilon}_t \leq x).$$

Further define

$$T_n(\frac{k}{n}, x) = \frac{k}{n}(1 - \frac{k}{n})\sqrt{n}\left(\hat{F}_k(x) - \hat{F}^*_{n-k}(x)\right)$$

and the test statistic

$$M_n = \max_k \sup_x |T_n(k/n, x)|$$

where the max is taken over $1 \leq k \leq n$ and the supremum with respect to x is taken over the entire real line. For each fixed k, the supremum of T_n with respect to the second argument gives the weighted Kolmogorov-Smirnov two-sample test with weights $[(k/n)(1-k/n)]^{1/2}$. Thus the test M_n looks for the maximum value of weighted Kolmogorov-Smirnov statistics for all possible sample splits.

We have the following identities:

$$T_n(\frac{k}{n}, x) = n^{-1/2}\sum_{t=1}^{k} I(\hat{\varepsilon}_t \leq x) - \frac{k}{n} n^{-1/2} \sum_{t=1}^{n} I(\hat{\varepsilon}_t \leq x) \tag{2}$$

$$= n^{-1/2}\sum_{t=1}^{k}\{I(\hat{\varepsilon}_t \leq x) - F(x)\} - \frac{k}{n} n^{-1/2}\sum_{t=1}^{n}\{I(\hat{\varepsilon}_t \leq x) - F(x).\} \tag{3}$$

Writing in the form (3) will be convenient for studying the limiting distribution of T_n and hence of M_n.

As will be shown, the test M_n has non-trivial local power against changes in the scale parameter of the disturbances. However, like the CUSUM test, it has no local power against shifts in the regression parameters if the mean regressor is zero. To circumvent this undesirable feature, we introduce a new class of tests based on weighted e.d.f. of residuals. Let $X_k = (x_1, ..., x_k)'$ and

$$A_k = (X'X)^{-1/2}(X_k'X_k)(X'X)^{-1/2}. \tag{4}$$

Define the $p \times 1$ vector process T_n^*,

$$T_n^*(\frac{k}{n}, x) = (X'X)^{-1/2}\sum_{t=1}^{k} x_t\{I(\hat{\varepsilon}_t \leq x) - \hat{F}_n(x)\}$$
$$-A_k(X'X)^{-1/2}\sum_{t=1}^{n} x_t\{I(\hat{\varepsilon}_t \leq x) - \hat{F}_n(x)\} \tag{5}$$

and the test statistic

$$M_n^* = \max_k \sup_x \|T_n^*(\frac{k}{n}, x)\|_\infty$$

where $\|y\|_\infty = \max\{|y_1|, ..., |y_p|\}$, the maximum norm. The process T_n^* and test M_n^* reduce to T_n and M_n, respectively, when the weights $x_t = 1$ for all t.

If there is a constant regressor, then the following identity holds,

$$(X'X)^{-1/2}\sum_{t=1}^{k} x_t - A_k(X'X)^{-1/2}\sum_{t=1}^{n} x_t = 0, \quad \forall\, k \tag{6}$$

so that the value of $T_n^*(k/n, x)$ will not change when $\hat{F}_n(x)$ in (5) is replaced by an arbitrary function of x. In particular, T_n^* can be written as

$$(X'X)^{-1/2}\sum_{t=1}^{k} x_t'\{I(\hat{\varepsilon}_t \leq x) - F(x)\} - A_k(X'X)^{-1/2}\sum_{t=1}^{n} x_t'\{I(\hat{\varepsilon}_t \leq x) - F(x)\}. \quad (7)$$

Equation (7) is a weighted version of (3). This expression cannot be used to compute the test M_n^*, as $F(x)$ is unknown; however, it will be useful in deriving the limiting process of T_n^*. When computing the test, one should omit $\hat{F}_n(x)$ if a constant regressor is included. However, whether or not there is a constant regressor, the two expressions for T_n^* (5) and (7) have the same null limiting process, because $n^{1/2}\{\hat{F}_n(x) - F(x)\} = O_p(1)$ uniformly in x, a well known result for residual e.d.f. (Shorack and Wellner 1986, Chapter 4), and

$$n^{-1/2}\{(X'X)^{-1/2}\sum_{t=1}^{k} x_t - A_k(X'X)^{-1/2}\sum_{t=1}^{n} x_t\} = o_p(1)$$

uniformly in k by assumptions (A.3) and (A.9). Finally, we remark here that if none of the regressors are trending, then we may substitute the scalar k/n for the matrix A_k.

Let $B(u,v)$ be a Gaussian process on $[0,1]^2$ with zero mean and covariance function

$$E\{B(s,u)B(t,v)\} = (\min(s,t) - st)(\min(u,v) - uv),$$

which we shall call a two-parameter Brownian bridge on $[0,1]^2$. In what follows, the notation " $\Rightarrow$ " is used to denote the weak convergence in the space of $D(T)$ or $D(T) \times D(T) \times \cdots \times D(T)$ where $T = [0,1]^2$ under the (extended) Skorohod J_1 topology.

Theorem 1 *Under model (1) and assumptions (A.1)–(A.9),*

(i) $$T_n(\frac{[n\cdot]}{n}, \cdot) \Rightarrow B(\cdot, F(\cdot))$$

and

(ii) $$T_n^*(\frac{[n\cdot]}{n}, \cdot) \Rightarrow B^*(\cdot, F(\cdot))$$

where $B^* = (B_1, B_2, ..., B_p)'$ *is a vector of* p *independent two-parameter Brownian bridges on* $[0,1]^2$.

Let $G(\cdot)$ denote the d.f. of the r.v. $\sup_{0\le u\le 1}\sup_{0\le v\le 1}|B(u,v)|$, which is tabulated in Picard (1985).

Corollary 1 *Under the assumptions of Theorem 1,*

$$\lim_{n\to\infty} P(M_n \le x) = G(x)$$

and

$$\lim_{n\to\infty} P(M_n^* \le x) = [G(x)]^p.$$

The proof of the theorem is based on the limiting behavior of the process K_n^*,

$$K_n^*(s,x) = (X'X)^{-1/2}\sum_{t=1}^{[ns]} x_t\{I(\hat{\varepsilon}_t \le x) - F(x)\}$$

which we shall call the weighted sequential empirical process of residuals (w.s.e.p.). Note

$$T_n^*(\frac{[ns]}{n}, z) = K_n^*(s,z) - A_{[ns]}K_n^*(1,z).$$

Introduce

$$H_n(s,x) = (X'X)^{-1/2}\sum_{t=1}^{[ns]} x_t\{I(\varepsilon_t \le x) - F(x)\}$$

which is a non-residual version of w.s.e.p. Theorem A.2 in the Appendix implies that $K_n^*(s,x)$ can be written as

$$H_n(s,x) + f(x)(X'X)^{-1/2}(X'_{[ns]}X_{[ns]})(\hat{\beta} - \beta) \tag{8}$$

plus an $o_p(1)$ term which is uniformly small in both s and x. The second term above is identical to the corresponding term of $A_{[ns]}K_n^*(1,x)$, so that

$$T_n^*(s,x) = H_n(s,x) - A_{[ns]}H_n(1,x) + o_p(1).$$

Corollary A.2 in the Appendix gives the limiting process of T_n^*.

The limiting process of K_n^*, if it exists, will depend on the limiting distribution of the estimated parameters and on the error density function f, as is easily seen from (8). However, parameter estimation does not affect the limiting process of T_n^*. The fact that the limiting process of T_n^* depends on F rather than f allows us to construct distribution-free tests. The sup-type test, for example, transforms out this dependence on F. Further, if the error ε_t has a symmetric distribution about zero, so that $F(0) = 1/2$, then tests based on $T_n^*(s, 0)$ will also be asymptotically distribution-free.

Besides the sup-type tests, the mean-type test can also be used. Let

$$A_n = \frac{1}{n^2} \sum_k \sum_j |T_n(\frac{k}{n}, \hat{\varepsilon}_j)|^2 \quad \text{and} \quad A_n^* = \frac{1}{n^2} \sum_k \sum_j \|T_n^*(\frac{k}{n}, \hat{\varepsilon}_j)\|^2.$$

The result of Theorem 1 implies that A_n converges in distribution to $\int_0^1 \int_0^1 B(s,t)^2 ds\, dt$ and A_n^* converges in distribution to $\int_0^1 \int_0^1 \sum_{i=1}^p B_i(s,t)^2 ds\, dt$, where $B_1, ..., B_p$ are independent copies of $B(\cdot, \cdot)$. Many other tests can be constructed based on the weak convergence of Theorem 1.

The weak convergence of empirical processes based on estimated residuals has been studied by many authors, see, for example, Mukantseva (1978), Boldin (1982, 1989), Pierce and Kopecky (1982), and Kreiss (1991). It appears that Koul (1970) is among the first who studied weighted empirical processes and followed by Withers (1972). Weighted empirical processes of residuals have been studied by Koul (1984, 1991). Shorack and Wellner (1986) give more references on residual empiricals. The weak convergence for the sequential version, which is essential for the structural change problem, has not been widely examined: Bai (1991) considered the sequential empirical process for ARMA residuals.

4 Computing the tests

We now derive some alternative expressions for the test statistics M_n and M_n^* that are suitable for programmed computation. We shall focus on M_n^*; the test M_n is a special case. Now for each fixed k, $\|T_n^*(\frac{k}{n}, x)\|_\infty$ can only possibly change its value at $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ when x varies, therefore, the maximum value with respect to x can be found at $x = \hat{\varepsilon}_i$ $(i = 1, 2, ..., n)$ or equivalently, at $x = \hat{\varepsilon}_{(i)}$ $(i = 1, 2, ..., n)$, where $\hat{\varepsilon}_{(i)}$ is the i-th ordered statistic. Let $R_1, R_2, ..., R_n$ denote the ranks of $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ and $D_1, D_2, ..., D_n$ denote the anti-ranks so that $R_{D_i} = D_{R_i} = i$. For a fixed j, let us evaluate $T_n^*(\frac{k}{n}, x)$ at $x = \hat{\varepsilon}_{(j)}$. First assume there is a constant regressor, so that T_n^* is equivalent to the expression (5) with $\hat{F}_n(x) = 0$ omitted due to (6). Since $\sum_{t=1}^n x_t I(\hat{\varepsilon}_t \le \hat{\varepsilon}_{(j)})$ is the sum of those vectors x_t such that $\hat{\varepsilon}_t$ is not larger than $\hat{\varepsilon}_{(j)}$, it can be written as $\sum_{i=1}^j x_{D_i}$. Similarly, $\sum_{t=1}^k x_t I(\hat{\varepsilon}_t \le \hat{\varepsilon}_{(j)}) = \sum_{i=1}^j x_{D_i} I(D_i \le k)$. Thus if we define a sequence of numbers $Z_{t,k}$ $(t = 1, 2, ..., n)$ such that

$$Z_{t,k} = \begin{cases} 1 & \text{for } i = 1, 2, ..., k \\ 0 & \text{for } t = k+1, k+2, ..., n \end{cases}$$

Then $Z_{D_i,k} = 1$ if and only if $D_i \le k$. Thus

$$T_n^*(\frac{k}{n}, \hat{\varepsilon}_{(j)}) = (X'X)^{-1/2} \left(\sum_{i=1}^j x_{D_i} Z_{D_i,k} - (X_k' X_k)(X'X)^{-1} \sum_{i=1}^j x_{D_i} \right), \tag{9}$$

and

$$M_n^* = \max_k \max_j \left\| (X'X)^{-1/2} \left(\sum_{i=1}^j \{ Z_{D_i,k} I - (X_k' X_k)(X'X)^{-1} \} x_{D_i} \right) \right\|_\infty \tag{10}$$

where I is the $p \times p$ identity matrix. Taking $x_t = 1$ in the above formula for all t, we obtain

$$M_n = \max_k \max_j \frac{1}{\sqrt{n}} \left| \sum_{i=1}^j Z_{D_i,k} - j\frac{k}{n} \right|$$

yielding an easily computable formula, see similar formula in Hajek (1969, p. 62-63). When there is no constant regressor, the expression (9) has to be adjusted by subtracting the left hand side of (6) multiplied by j/n, which is the product of the left hand side of (6) and $\hat{F}_n(\hat{\varepsilon}_{(j)})$. The test M_n^* is adjusted accordingly and M_n stays the same.[1]

[1] A SAS program for computing the statistics is available upon request.

5 Local Power Analysis

We consider model (2) with the Kramer, Ploberger, and Alt (1988) (henceforth KPA) type local alternatives:

$$\beta_t = \beta + \Delta_1 g(t/n) n^{-1/2} \quad \text{and} \quad \varepsilon_t^* = \varepsilon_t (1 + \Delta_2 h(t/n) n^{-1/2})^{-1}. \tag{11}$$

where ε_t are i.i.d. with distribution function F and density function f. The function g and h are defined on $[0,1]$ and are integrable. Define the vector function

$$\lambda_g(s) = \int_0^s g(v)dv - s\int_0^1 g(v)dv \tag{12}$$

and the function

$$\lambda_h(s) = \int_0^s h(v)dv - s\int_0^1 h(v)dv. \tag{13}$$

If h is a simple shift function such that $h(x) = 0$ for $x \leq \tau$ and $h(x) = 1$ for $x > \tau$, where $\tau \in (0,1)$, then $\lambda_h(s) = (\tau \wedge s)(1 - \tau \vee s)$. Similar is true for λ_g.

Theorem 2 *Under assumptions (A.1)-(A.9) and the local alternatives (11), we have*

$$M_n \xrightarrow{d} \sup_{0\leq s\leq 1}\sup_{0\leq t\leq 1} |B(s,t) + \Delta_1 p(t)\bar{x}'\lambda_g(s) + \Delta_2 q(t)\lambda_h(s)| \tag{14}$$

and

$$M_n^* \xrightarrow{d} \sup_{0\leq s\leq 1}\sup_{0\leq t\leq 1} \|B^*(s,t) + \Delta_1 p(t)Q^{1/2}\lambda_g(s) + \Delta_2 q(t)Q^{-1/2}\bar{x}\lambda_h(s)\|_\infty \tag{15}$$

where $p(t) = f(F^{-1}(t))$ *and* $q(t) = f(F^{-1}(t))F^{-1}(t)$.

The tests have nontrivial local power as long as $\lambda_g(s) \neq 0$ or $\lambda_h(s) \neq 0$ for some s. In addition, $\lambda_g = \lambda_h = 0$ for all s if and only if g and h are constant functions, implying no change in the parameters.

Corollary 2 *(Changing regression parameters only). Under the assumptions of Theorem 2,*

$$M_n \xrightarrow{d} \sup_{0\leq s\leq 1}\sup_{0\leq t\leq 1} |B(s,t) + \Delta_1 p(t)\bar{x}'\lambda_g(s)|,$$

$$M_n^* \xrightarrow{d} \sup_{0\leq s\leq 1}\sup_{0\leq t\leq 1} \|B^*(s,t) + \Delta_1 p(t)Q^{1/2}\lambda_g(s)\|_\infty.$$

The corollary is obtained by simply taking $\Delta_2 = 0$ in Theorem 2. In testing for changes in the regression parameters, M_n behaves like the CUSUM test of Brown, Durbin, and Evans (1975) in the sense of lacking local power when the mean of regressors $\bar{x}$ is orthogonal to the vector function g, as shown by KPA. The test M_n^*, however, does have local power irrespective of the relationship between $\bar{x}$ and g. Thus it behaves like the fluctuation test of PKK. The drift term is also similar in form to the fluctuation test.

There is a danger of misinterpreting the result of Corollary 2. Let us examine the following model under the alternative hypothesis:

$$y_t = x_t'\beta + \Delta\ g(t/n)n^{-1/2} + \varepsilon_t, \tag{16}$$

where g is a scalar function. This model would be a special case of (11) provided there is a constant regressor, but for now assume there is no constant regressor and the mean regressor $\bar{x}$ is zero. Then it is M_n^* not M_n has no local power, which seemingly contradicts Corollary 2. This situation arises because there is change in the parameter of a regressor that is not considered under the null. To see this, we consider a more general situation:

$$y_t = x_t'\beta + \Delta\, z_t' g(t/n)n^{-1/2} + \varepsilon_t \tag{17}$$

where z_t is $q \times 1$ and g is a vector function. The x_t are the only regressors under the null of $\Delta = 0$. Suppose that $n^{-1}\sum_{t=1}^{[ns]} z_t \to s\bar{z}$ and $n^{-1}\sum_{t=1}^{[ns]} x_t z_t' \to sR_{xz}$ uniformly in s where R_{xz} is some $p \times q$ matrix. Then

$$M_n \xrightarrow{d} \sup_{0\le s\le 1}\ \sup_{0\le t\le 1} |B(s,t) + \Delta p(t)\bar{z}'\lambda_g(s)| \tag{18}$$

$$M_n^* \xrightarrow{d} \sup_{0\le s\le 1}\ \sup_{0\le t\le 1} \|B^*(s,t) + \Delta p(t)Q^{-1/2}R_{xz}\lambda_g(s)\|_\infty. \tag{19}$$

Now let $z_t = 1$, so that (17) reduces to (16). Moreover, $\bar{z} = 1$ and $R_{xz} = \bar{x}$. Thus if the mean value of the regressor is zero, M_n^* has no local power but M_n does. Of course, for $z_t = x_t$, (18) and (19) coincide with Corollary 2.

Now taking $\Delta_1 = 0$ in Theorem 2 yields:

Corollary 3 *(Changing scale only). Under the assumptions of Theorem 2,*

$$M_n \xrightarrow{d} \sup_{0 \le s \le 1} \sup_{0 \le t \le 1} |B(s,t) + \Delta_2 q(t) \lambda_h(s)|,$$

$$M_n^* \xrightarrow{d} \sup_{0 \le s \le 1} \sup_{0 \le t \le 1} \|B^*(s,t) + \Delta_2 q(t) Q^{-1/2} \bar{x} \lambda_h(s)\|_\infty.$$

When testing for a shift in the scale parameter, the situation is reversed from the test of a shift in regression parameters; M_n^* has no local power if the regressor mean, $\bar{x}$, is zero whereas M_n has local power irrespective of this mean value.

In summary, the test M_n has non-trivial local power when testing for changes in the disturbances regardless of the mean value of the regressors. The test M_n^* has non-trivial local power when testing for changes in the regression parameter β, regardless of the angle between the regressor and the structural shift, see KPA for comparison. Moreover, M_n also has local power for testing changes in β and M_n^* also has local power for testing changes in the disturbances, except for some special circumstances discussed earlier. Whereas the conventional CUSUM test only has local power against changes in regression parameters and CUSUM-SQ test only has local power against heteroskedasticity, see Ploberger and Kramer (1990). It is not clear how the fluctuation test performs for a change in the disturbances, as it is not intended to be used for this purpose.

The test statistics M_n and M_n^* are more powerful when used for testing the simple shift alternatives. To fix ideas, consider the scale change alternatives as in Corollary 3. Let H_a be the set of functions h defined by

$$H_a = \{h;\ 0 \le h \le 1,\ \int_0^1 h(v)dv = 1 - a\}$$

for some a satisfying $0 < a < 1$. The number $1 - a$ represents on average the deviation of h from 0.

Consider the test statistic M_n. Since $B(s,t)$ is uniformly bounded in probability, the value of M_n is mainly determined by the drift term $\Delta_2 q(t) \lambda_h(s)$, for large $|\Delta_2|$. Thus with high probability, in order for $T_n(s,x)$ to be maximized, the following quantity needs to be maximized with respect to s

$$\left| s \int_0^1 h(v)dv - \int_0^s h(v)dv \right|. \tag{20}$$

We determine the $h \in H_a$ and $s \in [0,1]$ that maximize the objective function (20). It is easy to show that there are two set of solutions, depending on whether the quantity inside the absolute value sign is positive or negative. One solution is given by

$$h^*(v) = \begin{cases} 0 & \text{if } v \le a \\ 1 & \text{if } v > a \end{cases} \tag{21}$$

and $s^* = a$. The other solution is given by

$$h^*(v) = \begin{cases} 1 & \text{if } v \le 1-a \\ 0 & \text{if } v > 1-a \end{cases}$$

and $s^* = 1 - a$. But both of these h^* imply a simple shift alternative, so the test is more powerful against a simple shift. Furthermore, the value of the objective function evaluated at the optimal solution is $a(1-a)$ in both cases. But $a(1-a)$ is maximized for $a = 1/2$, implying a higher power for detecting a shift that occurs near the middle of the observations.

To see (21) is a solution, consider the objective function (20) without the absolute value sign. For each fixed s, since the second term is non-positive, (20) will be maximized by choosing $h(v) = 0$ for $v \le s$. The objective function becomes $s\int_s^1 h(v)dv$ with $\int_s^1 h(v)dv = (1-a)$. To maximize the objective function, one needs to choose s as large as possible. In order to choose the largest s such that $\int_s^1 hdv = 1-a$, one needs to choose h as large as possible. Thus $h(v) = 1$. The constraint becomes $\int_s^1 dv = 1-a$ which implies $s^* = a$. The second solution is obtained by maximizing the negative function inside the absolute value sign.

6 Trending Regressors

We consider the following model:

$$y_t = z_t'\alpha + \gamma_0 + \gamma_1(t/n) + \ldots + \gamma_q(t/n)^q + \varepsilon_t \tag{22}$$

where z_t is a $r \times 1$ vector of stochastic regressor and $\{z_s;\ s \leq t-1\}$ are independent of ε_t. Let $x_t = (z_t', 1, t/n, ..., (t/n)^q)'$ be a $p \times 1$ vector, with $p = r + q + 1$.

The polynomial trends $\{(t/n)^i; 1 \leq i \leq q\}$ could be written without dividing through by n. Writing in this fashion saves notations by eliminating the weighting matrix such as diag$(n^{-1/2}, ..., n^{-(q+1)/2})$ that would otherwise be needed. We shall maintain all assumptions (A.1)-(A.8) of Section 1, except changing (A.3) to

(A.3′)

$$\text{plim}\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' = \lim \frac{1}{n} E \sum_{t=1}^{[ns]} x_t x_t' = Q(s), \quad \text{uniformly in } s \in [0,1]$$

where $Q(s)$ is positive definite for $s > 0$ and $Q(0) = 0$. If each element of $Q(s)$ is a continuous function on $[0,1]$, then one can show that pointwise convergence in s implies uniform convergence in s. Assumption (A.3′) actually admits a much wider class of models than (22).

In the presence of trending regressor only the weighted version, M_n^*, is asymptotically distribution-free. We shall assume that there is a constant regressor. Let $X_k = (x_1', x_2' ..., x_k')'$ and and define A_k as in (4) and T_n^* as follows:

$$T_n^*(\frac{k}{n}, x) = (X'X)^{-1/2} \sum_{t=1}^{k} x_t I(\hat{\varepsilon}_t \leq x) - A_k (X'X)^{-1/2} \sum_{t=1}^{n} x_t I(\hat{\varepsilon}_t \leq x). \quad (23)$$

Again let $M_n^* = \max_k \sup_x \|T_n^*(k/n, x)\|_\infty$. The computation of M_n^* is given by (10). Note that $A_{[ns]} \xrightarrow{p} A(s) = Q(1)^{-1/2} Q(s) Q(1)^{-1/2}$ uniformly in s.

Theorem 3 *Under assumptions (A.1-A.8) with (A.3) replaced by (A.3′), we have*

$$T_n^*([n\cdot]/n, \cdot) \Rightarrow B^*(\cdot, F(\cdot))$$

where $B^*(s,u)$ is a vector Gaussian process defined on $[0,1]^2$ with zero mean and covariance matrix

$$E\{B^*(r,u) B^*(s,v)'\} = \{A(r \wedge s) - A(r)A(s)\}\{u \wedge v - uv\}.$$

Corollary 4 *Under assumptions of Theorem 3,*

$$M_n^* \xrightarrow{d} \sup_{0 \leq s,u \leq 1} \|B^*(s,u)\|_\infty.$$

The behavior of the test under the local alternatives (11) can again be analyzed. Extending Lemma 4 of KPA, we can show that

$$\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' g(t/n) \xrightarrow{p} \int_0^s \frac{dQ(v)}{dv} g(v)dv \tag{24}$$

and the convergence is uniform in s. The above integral exists if g has bounded variation on [0,1]. When $Q(v) = vQ(1)$, (24) reduces to the result of KPA. Let

$$\lambda_g^*(s) = \int_o^s \frac{dQ(v)}{dv} g(v)dv - Q(s)Q(1)^{-1}\int_0^1 \frac{dQ(v)}{dv} g(v)dv,$$

$$\lambda_h^*(s) = \int_o^s \frac{dQ(v)\mathrm{e}}{dv} h(v)dv - Q(s)Q(1)^{-1}\int_0^1 \frac{dQ(v)\mathrm{e}}{dv} h(v)dv,$$

where $\mathrm{e} = (1, 0, ..., 0)'$.

Theorem 4 *Under the local alternative (11),*

$$M_n^* \xrightarrow{d} \sup_{0\le s,u\le 1} \|B^*(s,u) + p(u)\Delta_1 Q(1)^{-1/2}\lambda_g^*(s) + q(u)\Delta_2 Q(1)^{-1/2}\lambda_h^*(s)\|_\infty$$

where $p(\cdot)$ and $q(\cdot)$ are given in Theorem 2.

Of course, when $Q(v) = vQ(1)$, the theorems and corollaries in previous sections can be derived from the results of this section. However, the limiting distribution obtained here is generally regressor-dependent, so critical values of the tests have to be found case by case, though leading cases can be tabulated. Also, the result of this section requires the existence of a constant regressor.

Tests allowing trending regressors have been proposed by MacNeill (1978), Sen (1980), Kim and Siegmund (1989), Hansen (1992), Chu and White (1992), Perron (1991), Vogelsang (1992). Those tests are connected with the partial sums, the likelihood ratio, or Wald-type statistics. It remains to be studied how the proposed tests in this paper perform relative to those in the literature.

7 Some Comments

We have maintained the assumption that the disturbances $\{\varepsilon_t\}$ are independent r.v.'s. This could be extended to ARMA models. Although further extension to more general dependence structure such as mixing is also possible in terms of weak convergence, critical values of the tests are difficult to obtain because the limiting process will have a complicated correlation structure.[2] For the ARMA models such as $\varepsilon_t = B(L)u_t$, where $B(L)$ is a ratio of two polynomials of the lag operator L, one can still estimate u_t. This could be done with a two-step procedure. The first step involves estimating the regression coefficients β and the second step estimating the coefficients of $B(L)$ using the first step residuals. The two-step procedure yields estimates of u_t from which empirical distribution function can be constructed. Although details remain to be worked out, the results of the previous sections is expected to hold. Bai (1991) obtained similar results for pure ARMA models for the test M_n.

The tests proposed in this paper are similar in form to the fluctuation test of PKK. In fact, consider regressing $I(\hat{\varepsilon}_t \leq x)$ on x_t for t=1, ..., k, and write

$$\hat{\theta}^{(k)}(x) = (X_k'X_k)^{-1}\sum_{t=1}^{k} x_t I(\hat{\varepsilon}_t \leq x).$$

Using the expression (23) for T_n^*, one obtains,

$$T_n^*(k/n, x) = A_k(X'X)^{1/2}\left(\hat{\theta}^{(k)}(x) - \hat{\theta}^{(n)}(x)\right).$$

The quantity $\hat{\theta}^{(k)}(x)$ can be considered to be an estimate of $Q^{-1}\bar{x}F(x)$ using a partial sample, and $\hat{\theta}^{(T)}(x)$ using the whole sample. The test here has one more dimension than PKK's, and thus can be viewed as a two-dimensional fluctuation test. Notice that PKK's test does not include trending regressors but T_n^* does. This comparison also suggests a way to extend PKK's test to trending regressors. Simply replace t/T in their notation by the matrix A_t and let

$$S_n = \max_{p \leq k \leq n} \hat{\sigma}^{-1} \| A_k(X'X)^{1/2}(\hat{\beta}^{(k)} - \hat{\beta}^{(n)}) \|_\infty$$

where $\hat{\sigma}$ and $\hat{\beta}^{(k)}$ $(k = p, ..., n)$ are defined in PKK. Then it is not difficult to show

$$S_n \xrightarrow{d} \sup_{0 \leq s \leq 1} \|\tilde{B}(s)\|_\infty$$

[2]There is a large literature on empirical process based on mixing sequences; see the early study of Billingsley (1968, p. 240) and the recent study of Andrews and Pollard (1990).

where $\tilde{B}$ is mean zero vector Gaussian process with covariance matrix

$$E\{\tilde{B}(u)\tilde{B}(v)'\} = A(u \wedge v) - A(u)A(v).$$

Other issues that are left unaddressed in this paper include cointegrated regressors, and size and power comparisons with other tests in the literature.

References

[1] Alogoskoufis, G. and Smith, R. (1991). The Phillips curve, the persistence of inflation, and the Lucas critique: Evidence from exchange rate regimes. *American Economic Review*, 81 1254-1275.

[2] Andrews, D.W.K. (1990). Tests for parameter instability and structural change with unknown change point. Manuscript, Cowles Foundation Discussion paper No. 943, Yale University.

[3] Andrews, D.W.K. and Pollard, D. (1990). A functional central limit theorem for strong mixing stochastic processes. Cowles foundation Discussion Paper No. 951, Cowles Foundation, Yale University.

[4] Bai, J. (1991). Weak convergence of sequential empirical processes of ARMA residuals. Manuscript, Department of Economics, U.C. Berkeley.

[5] Bai, J., Lumsdaine, R.L., and Stock, J.H. (1991). Testing for and dating breaks in integrated and cointegrated time series, manuscript, Kennedy School of Government, Harvard University.

[6] Banerjee, A., Lumsdaine, R.L., and Stock, J.H. (1992). Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence. *Journal of Business & Economic Statistics*, 10 271-287.

[7] Bickel, P.J. and Wichura, M.J. (1971). Convergence for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* 42 1656-1670.

[8] Billingsley, P. (1968). Convergence of Probability Measures. New York: Wiley.

[9] Boldin, M.V. (1982). Estimation of the distribution of noise in an autoregression scheme. *Theory Probab. Appl.* 27 866-871.

[10] Boldin, M.V. (1989). On testing hypotheses in sliding average scheme by the Kolmogorov-Smirnov and ω^2 tests. *Theory Probab. Appl.,* 34, 699-704.

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

[12] Carlstein, E. (1988). Nonparametric change point estimation. *Ann. Statist.*, 16, 188-197.

[13] Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions, *Econometrica,* 28, 591-605.

[14] Chu, C-S J. and H. White (1992). A direct test for changing trend, *Journal of Business & Economic Statistics*, 10 289-300.

[15] Cooley, T.F. and E.C. Prescott (1973). An adaptive regression model, *International Economic Review,* 14, 364-371.

[16] Christiano, L.J. (1992). Searching for a break in GNP, *Journal of Business & Economic Statistics*, 10 237-250.

[17] Csörgő, M. and Horváth L.(1988). Nonparametric methods for change-point problems, in *Handbook of Statistics,* Vol 7, ed. by P.R. Krishnaiah and C.R. Rao. New York: Elsevier.

[18] Duembgen, L. (1991). The asymptotic behavior of some nonparametric change point estimators, *Ann. Stat.*, 19, 1471-1495.

[19] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.*, 1, 279-290.

[20] Durbin, J. (1976). Kolmogorov-Smirnov tests when parameters are estimated, in *Empirical Distributions and Processes*, Springer-Verlad Lecture Notes in Mathematics.

[21] Goldfeld, S.M. and Quandt, R.E. (1973a). The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement*, 2475-485.

[22] Goldfeld, S.M. and Quandt, R.E. (1973b). A Markov model for switching regressions, *Journal of Econometrics*, 1 3-16.

[23] Hajek, J. (1969). Nonparametric Statistics. Holden-Day, San Francisco.

[24] Hall, P. and C.C. Heyde (1980). Martingale Limit Theory and its Applications. New York: Academic Press.

[25] Hansen, B.E. (1992). Tests for parameter instability in regressions with I(1) processes, *Journal of Business & Economic Statistics*, 10 321-335.

[26] Kim, H.J. and D. Siegmund (1989). The likelihood ratio test for a change point in simple linear regression, *Biometrika*, 76, 409-423.

[27] Koul, H. L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* 41, 1768-1773.

[28] Koul, H. L. (1984). Tests of goodness-of-fit in linear regression. *Colloaquia Mathematica Societatis Janos Bolyai.*, 45. *Goodness of fit,* Debrecen, Hungary. 279-315.

[29] Koul, H. L. (1991). A weak convergence result useful in robust autoregression. *Statist. Plann. & Inference*, 29, 1291-308.

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

[31] Kreiss, P. (1991). Estimation of the distribution of noise in stationary processes. *Metrika,* 38, 285-297.

[32] MacNeill, I.B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to test for change of regression at unknown times, *Annals of Statistics*, 6 422-433.

[33] Mukantseva, L.A. (1977). Testing normality in one-dimensional and multidimensional linear regression. *Theory Prob. Appl.* 22 591-602.

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

[35] Perron, P. (1991). A test for changes in a polynomial trend function for a dynamic time series, mimeo, Princeton University.

[36] Picard, D. (1985). Testing and estimating change-point in time series. *Adv. Appl. Prob.* 17 841-867.

[37] Pierce, D.A. and Kopecky, K.J. (1979). Testing goodness of fit for the distribution of errors in regression model. *Biometrika,* 66 1-5.

[38] Ploberger, W. (1989). The local power of the CUSUM -SQ test against heteroskedasticity. In P. Hackel (ed.) *Statistical Analysis and Forecasting of Economic Structural Change.* Heidelberg: Springer, 127-133.

[39] Ploberger, W. and Kramer, W. (1990). The local power of the CUSUM and CUSUM of squares tests. *Econometric Theory*, 6 335-347.

[40] Ploberger, W., Kramer, W., and Kontrus, K. (1988). A new test for structural stability in the linear regression model, *Journal of Econometrics*, 40 307-318.

[41] Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes, *J. American Stat. Asso.,* 55 324-330.

[42] Sen, P.K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitsch. Wahrsch. verw. Gebiete,* 52 203-218.

[43] Shorack, G.R. and Wellner, J.A. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

[44] Vogelsang, T.J. (1992). Wald-type tests for detecting shifts in the trend function of a dynamic time series. Manuscript, Department of Economics, Princeton University.

[45] Withers, C.S. (1975). Convergence of empirical processes of mixing rv's on $[0,1]$. *The Annals of Statistics*, 3 1101-1108.

[46] Zivot, E. and Andrews, D.W.K. (1992). Further evidence on the great crash, the oil price shock and the unit root hypothesis, *Journal of Business & Economic Statistics*, 10 251-271.

A Appendix

In view of the mathematical structure of parameter changes in regression models, we shall first present and prove a series of results concerning the weak convergence of weighted sequential empirical processes. These results are of independent interest and will be used subsequently in proving the theorems stated in the body of the paper.

Let $D_q^p[0,1]$ be the set of functions $f = (f_1, ..., f_p)$ defined on $[0,1]^q$ that are right continuous and have left limits. Endowed with the extended Skorohod J_1 topology, $D_p^q[0,1]$ is a separable and complete metric space, so that finite dimensional convergence plus tightness implies weak convergence for a sequence of random elements of $D_q^p[0,1]$; see Bickel and Wichura (1971). The space $D_p^q[0,1]$ for $p = q = 1$ is extensively studied by Billingsley (1968).

Theorem A.1 *Let $U_1, U_2, \ldots, U_n$ be a sequence of i.i.d. uniformly distributed random variables on $[0,1]$ and x_i $(i = 1, 2, ..., n)$ be a sequence of random vectors satisfying assumptions (A.5) and (A.6). Assume that U_i is independent of x_j for $j \leq i$. Then the process $Y_n(s,u)$ defined as*

$$Y_n(s,u) = n^{-1/2} \sum_{t=1}^{[ns]} x_t \{I(U_t \leq u) - u\}$$

with $Y_n(0,u) = Y_n(s,0) = 0$ is tight in $D_2^p[0,1]$.

Remarks: The process Y_n is a multivariate and multiparameter process. The requirement of uniform distribution is only for convenience. The theorem holds for arbitrary i.i.d. random variables ε_t. In this case, $I(U_t \leq u) - u$ is replaced by $I(\varepsilon_t \leq x) - F(x)$ where F is the distribution function of ε_t. Then $Y_n(s,u)$ (with $u = F(x)$) is tight. In addition, the i.i.d. assumption on U_i can be relaxed to a triangular array such that $U_{n1}, ..., U_{nn}$ are independent variables on $[0,1]$ with U_{ni} having a *d.f.* F_{ni} such that $\max_{1 \leq i \leq n} |F_{ni}(u_2) - F_{ni}(u_1)| \leq C|u_2 - u_1|$, where C is generic constant. This claim can be easily seen from the proof.

Lemma A.1 *Assume the conditions of Theorem A.1 hold. Then there exists* $K < \infty$, *such that for all* $s_1 < s_2$ *and* $u_1 < u_2$, *where* $0 \le s_i, u_i \le 1$ $(i = 1, 2)$

$$E\|Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)\|^{2\gamma}$$
$$\le K(u_2-u_1)^\alpha(s_2-s_1)^\alpha + n^{-(\gamma-1)}K(u_2-u_1)(s_2-s_1).$$

Without the loss of generality, one can assume that $\alpha \le \gamma$, since $|u_2 - u_1| \le 1$ and $|s_2 - s_1| \le 1$. Moreover, when

$$\tau n^{-(\gamma-1)/2(\alpha-1)} \le u_2 - u_1 \quad \text{and} \quad \tau n^{-(\gamma-1)/2(\alpha-1)} \le s_2 - s_1 \tag{25}$$

for $\tau > 0$, then the lemma implies

$$E\|Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)\|^{2\gamma}$$
$$\le K[1+\tau^{-2(\alpha-1)}](u_2-u_1)^\alpha(s_2-s_1)^\alpha. \tag{26}$$

This inequality is analogous to (22.15) of Billingsley (1968, p. 198).

Proof. Write $\eta_t = I(u_1 < U_t \le u_2) - u_2 + u_1$ and $Y_n^* = Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)$ for the moment. Then $Y_n^* = n^{-1/2}\sum_{i<t\le j} x_t\eta_t$ with $i = [ns_1]$ and $j = [ns_2]$. Note that $\{x_t\eta_t, \mathcal{F}_t\}$ is a sequence of (nonstationary) vector martingale differences, where $\mathcal{F}_t$ is the σ-field generated by $..., x_t, x_{t+1}; ..., U_{t-1}, U_t$. By the inequality of Rosenthal (Hall and Hedye, 1980 p. 23), there exists a constant $M < \infty$ only depending on γ and p such that

$$\begin{aligned} E\|Y_n^*\|^{2\gamma} &= E\left\{\left(\frac{1}{n}[\sum_{i<t\le j} x_t\eta_t]'\sum_{i<h\le j} x_h\eta_h\right)^\gamma\right\} \\ &\le ME\left(\frac{1}{n}\sum_{i<t\le j} E\{(x_t'x_t)\eta_t^2|\mathcal{F}_{t-1}\}\right)^\gamma + Mn^{-\gamma}\sum_{i<t\le j} E\{(x_t'x_t)^\gamma\eta_t^{2\gamma}\}. \end{aligned} \tag{27}$$

Note that x_t is measurable with respect to $\mathcal{F}_{t-1}$ and η_t is independent of $\mathcal{F}_{t-1}$. In addition, $E\eta_t^2 \le u_2 - u_1$ and $E\eta_t^{2\gamma} \le u_2 - u_1$. These results together with assumption (A.6) provide bounds for the two terms on the right of (27). The first term is bounded by

$$M(u_2-u_1)^\gamma E\left(\frac{1}{n}\sum_{i<t\le j}(x_t'x_t)\right)^\gamma \le MK(u_2-u_1)^\gamma(s_2-s_1)^\alpha$$

and the second term is bounded by

$$Mn^{-(\gamma-1)}(u_2-u_1)\frac{1}{n}\sum_{i<t\leq j}E(x_t'x_t)^\gamma \leq MKn^{-(\gamma-1)}(u_2-u_1)(s_2-s_1).$$

Renaming MK as K, the lemma then follows from $(u_2-u_1)^\gamma \leq (u_2-u_1)^\alpha$, for $\gamma \geq \alpha$.

Lemma A.2 *Under (A.5), we have for $s_1 \leq s \leq s_2$ and $u_1 \leq u \leq u_2$,*

$$\|Y_n(s,u)-Y_n(s_1,u_1)\| \leq \|Y_n(s_2,u_2)-Y_n(s_1,u_1)\| + O_p(1)n^{1/2}[(u_2-u_1)+(s_2-s_1)]$$

where the term $O_p(1)$ is uniform in s ($s \geq s_1$), does not depend on u and u_1 and satisfies

$$P(|O_p(1)| > C) < M/C^{2(1+\rho)}, \qquad \forall C > 0, \quad \textit{for some } \rho > 0.$$

Proof. First notice that all components of x_t can be assumed to be nonnegative. Otherwise write $x_t = \sum_{i=1}^p x_t^+(i) - \sum_{i=1}^p x_t^-(i)$ where $x_t^+(i) = (0,..0,x_{ti},0,...,0)'$ if $x_{ti} \geq 0$ and $x_t^+(i) = (0,..0,-x_{ti},0,...,0)'$ if $x_{ti} < 0$. In this way, Y_n can be written as a linear combination (with coefficients 1 or -1) of at most $2p$ processes with each process having nonnegative weighting vectors. In addition, $\|x_t^+(i)\| \leq \|x_t\|$ and $\|x_t^-(i)\| \leq \|x_t\|$. So assumptions (A.5) and (A.6) are satisfied for $x_t^+(i)$ and $x_t^-(i)$. It is thus enough to assume that the x_t are nonnegative. A new piece of notation, for vectors a and b, take $a \leq b$ to mean $a_i \leq b_i$ for all components. Since $x_t \geq 0$, the vector functions $x_t I(U \leq u)$ and $x_t u$ are nondecreasing in u. It is easy to show

$$\begin{aligned} Y_n(s,u)-Y_n(s_1,u_1) &\leq Y_n(s_2,u_2)-Y_n(s_1,u_2) \\ &+n^{1/2}(\frac{1}{n}\sum_{t=1}^{[ns]}x_t)(u_2-u)+n^{1/2}\left(\frac{1}{n}\sum_{t=[ns]}^{[ns_2]}x_t\{I(U_t\leq u_2)-u_2\}\right) \end{aligned}$$

and

$$Y_n(s_1,u_1)-Y_n(s,u) \leq n^{1/2}(\frac{1}{n}\sum_{t=1}^{[ns]}x_t)(u-u_1)+n^{1/2}\left(\frac{1}{n}\sum_{t=[ns_1]}^{[ns]}x_t\{I(U_t\leq u)-u_1\}\right).$$

The lemma follows from the boundedness of the indicator function and (A.5). □

Remarks: Bickel and Wichura (1971) provided a general framework for showing the tightness of a sequence of multiparameter stochastic processes. Their conditions are hard to verify and probably do not hold because of the dependence and unboundedness of x_t. Although there are empirical process theories for mixing and nonstationary variables, (see Andrews and Pollard (1990) and the references therein), none of them are directly applicable. Also, the presence of the $O_p(1)$ term in our Lemma A.2 seems to make it necessary for us to evaluate directly the modulus of continuity. A direct proof is also instructive. The arguments of Bickel and Wichura inspire the ideas used in the remaining proof.

Proof of Theorem A.1. Define

$$\omega_\delta(Y_n) = \sup\{\|Y_n(s',u') - Y_n(s'',u'')\|;\ |s'-s''| < \delta, |u'-u''| < \delta, s', s'', u', u'' \in [0,1]\}.$$

We shall show that for any $\epsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an integer n_0, such that

$$P(\omega_\delta(Y_n) > \epsilon) < \eta, \qquad n > n_0.$$

Since $[0,1]^2$ has only about δ^{-2} squares with side length δ, it suffices to show that for every point $(s_1,u_1) \in [0,1]^2$, every $\epsilon > 0$ and $\eta > 0$, there exist a $\delta \in (0,1)$ and an integer n_0 such that

$$P(\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\epsilon) < 2\delta^2\eta, \qquad n > n_0. \tag{28}$$

where $\langle\delta\rangle = \{(s,u);\ s_1 \le s \le s_1+\delta, u_1 \le u \le u_1+\delta\} \cap [0,1]^2$.

For given $\delta > 0$ and $\eta > 0$, choose C large enough so that for the $O_p(1)$ in Lemma A.2

$$P(|O_p(1)| > C) < \delta^2\eta. \tag{29}$$

By Lemma A.2 (see also (22.18) of Billingsley, 1968, p. 199), when $|O_p(1)| \le C$,

$$\sup_{[\delta]} \|Y_n(s,u) - Y_n(s_1,u_1)\| \le 3 \max_{1\le i,j\le m} \|Y_n(s_1+i\epsilon_n, u_1+j\epsilon_n) - Y_n(s_1,u_1)\| + 2\epsilon$$

where $\epsilon_n = \epsilon/(n^{1/2}C)$ and $m = [n^{1/2}C\delta/\epsilon] + 1$. Write

$$X(i,j) = Y_n(s_1+i\epsilon_n, u_1+j\epsilon_n) - Y_n(s_1,u_1).$$

Then

$$P(\sup_{[\delta]} \|Y_n(s,u)-Y_n(s_1,u_1)\| > 5\epsilon) \le P(|O_p(1)| > C)+P(\max_{1\le i,j\le m} \|X(i,j)\| > \epsilon). \quad (30)$$

Now for fixed i and k ($i \ge k$) write $Z(j) = X(i,j) - X(k,j)$. Notice that

$$(\epsilon/C)n^{-(\gamma-1)/(\alpha-1)} \le \epsilon/(Cn^{1/2}) = \epsilon_n \le j\epsilon_n, \qquad j \ge 1,$$

which follows from $n^{-(\gamma-1)/2(\alpha-1)} \le n^{-1/2}$ because $1 < \alpha \le \gamma$. By (25) and (26),

$$E\|Z(j) - Z(l)\|^{2\gamma} \le KC_\epsilon[(i-k)\epsilon_n]^\alpha[(j-l)\epsilon_n]^\alpha, \qquad 1 \le l \le j \le m$$

where, from (26) with $\tau = \epsilon/C$,

$$C_\epsilon = [1 + (C/\epsilon)^{2(\alpha-1)}] \le 2(C/\epsilon)^{2(\alpha-1)} \qquad \text{for small } \epsilon. \quad (31)$$

Thus by Theorem 12.2 of Billingsley (1968, p. 94), we have

$$P(\max_{1\le j\le m} \|Z(j)\| > \epsilon) \le \frac{K_1KC_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha(m\epsilon_n)^\alpha \le \frac{K_2C_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha\delta^\alpha \quad (32)$$

where K_1 is a generic constant and $K_2 = 2^\alpha K_1K$. The last inequality follows from $(m\epsilon_n) \le 2\delta$ for large n. Because

$$\left|\max_j \|X(i,j)\| - \max_j \|X(k,j)\|\right| \le \max_j \|X(i,j) - X(k,j)\| = \max_j \|Z(j)\|,$$

if we let $V(i) = \max_j \|X(i,j)\|$, then (32) implies

$$P(|V(i) - V(k)| > \epsilon) < \frac{K_2C_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha\delta^\alpha, \qquad 1 \le k \le i \le m.$$

Thus by Theorem 12.2 of Billingsley once again [let $\xi_h = V(h) - V(h-1)$, so that $V(i)$ is the partial sum S_i of random variables ξ_h in Billingsley's notation], we obtain

$$P(\max_{1\le i\le m} |V(i)| > \epsilon) \le \frac{K_1'K_2C_\epsilon}{\epsilon^{2\gamma}}(m\epsilon_n)^\alpha\delta^\alpha \le \frac{K_3C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha}$$

where K_1' is a generic constant and $K_3 = 2^\alpha K_1'K_2$. Note that $\max_i |V(i)| = \max_i \max_j \|X(i,j)\|$. Thus by (30)

$$P(\sup_{[\delta]} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\epsilon) \le \delta^2\eta + \frac{K_3C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha}.$$

By (31), the second term on the right hand side above is bounded by

$$\frac{K_3 C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha} \leq \delta^2 \frac{2K_3}{\epsilon^{2(\gamma+\alpha-1)}}(C\delta)^{2(\alpha-1)}. \tag{33}$$

By Lemma A.2, one can choose $C = (M/\eta)^{2(1+\rho)}\delta^{-(\frac{1}{1+\rho})}$ to assure (29) and the left hand side (33) becomes $K(\epsilon,\eta)\delta^a$, where $K(\epsilon,\eta)$ is a constant and $a = \frac{2(\alpha-1)\rho}{1+\rho} > 0$. Choose δ such that $K(\epsilon,\eta)\delta^a \leq \eta$, then (28) follows. The proof of the theorem is completed. □

Corollary A.1 *Under assumptions (A.2), (A.3′), (A.5)-A.6), the process H_n defined as*

$$H_n(s,x) = (X'X)^{-1/2}\sum_{t=1}^{[ns]} x_t\{I(\varepsilon_t \leq x) - F(x)\}$$

converges weakly to a Gaussian process H with zero mean and covariance matrix

$$E\{H(r,x)H(s,y)'\} = Q(1)^{-1/2}Q(r\wedge s)Q(1)^{-1/2}[F(x\wedge y) - F(x)F(y)]. \tag{34}$$

Proof. $H_n(s,x) = (X'X/n)^{-1/2}Y_n(s,F(x))$ if one lets $U_i = F(\varepsilon_i)$. Since $(X'X/n)$ converges in probability to the matrix $Q(1)$, the tightness of H_n follows from Theorem A.1. The finite dimensional convergence to a normal distribution is obvious. To verify the covariance matrix, consider for $r < s$ and $u = F(x) < v = F(y)$ and utilize the martingale property,

$$E\{Y_n(r,u)Y_n'(s,v)\} = \frac{1}{n}E\left(\sum_{t=1}^{[nr]} x_t x_t'\right)(u - uv) \tag{35}$$

which tends to $Q(r)(u - uv)$ by (A.3′). □

Corollary A.2 *Under the assumptions of the previous corollary, the process V_n defined as*

$$V_n(s,x) = H_n(s,x) - A_{[ns]}H_n(1,x)$$

converges weakly to a Gaussian process V with mean zero and covariance matrix

$$E\{V(r,u)V(s,v)'\} = \{A(r\wedge s) - A(r)A(s)\}\{u\wedge v - uv\}. \tag{36}$$

Proof. The tightness of V_n follows from the tightness of H_n and the convergence of $A_{[ns]}$ to a deterministic matrix $A(s)$ uniformly in s. The limiting process of V_n is, by Corollary A.1,

$$V(s,x) = H(s,x) - A(s)H(1,x).$$

Now (36) follows easily from (34). □

Note that (A.3) is a special case of (A.3′). When $Q(s) = sQ$ for some $Q > 0$, the covariance matrix of V then becomes $(r \wedge s - rs)\{F(x \wedge y) - F(x)F(y)\}I$ where I is the $p \times p$ identity matrix, yielding a multivariate Kiefer process with independent components.

We next study the asymptotic behavior of the residual empirical process. Under model (1), $\hat{\varepsilon}_t \le z$ if and only if $\varepsilon_t \le z + x_t'(\hat{\beta} - \beta)$, thus the residual s.e.p. K_n^* is given by

$$K_n^*(s,z) = (X'X)^{-1/2} \sum_{t=1}^{[ns]} x_t\{I(\varepsilon_t \le z + x_t'(\hat{\beta} - \beta)) - F(x)\}.$$

Under the local alternative of (11), $\hat{\varepsilon}_t \le z$ if and only if

$$\varepsilon_t \le z\{1 + \Delta_2 h(t/n)n^{-1/2}\} + x_t'\{(\hat{\beta} - \beta) + \Delta_1 x_t' g(t/n)n^{-1/2}\}\{1 + \Delta_2 h(t/n)n^{-1/2}\}.$$

Thus K_n^* becomes

$$K_n^*(s,z) = (X'X)^{-1/2} \sum_{t=1}^{[ns]} x_t\{I(\varepsilon_t \le z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(x)\} \qquad (37)$$

where

$$a_t = \Delta_2 h(t/n), \text{ and } b_t = x_t'\{\sqrt{n}(\hat{\beta} - \beta) + \Delta_1 x_t' g(t/n)\}\{1 + \Delta_2 h(t/n)n^{-1/2}\}. \qquad (38)$$

Choosing the weights $x_t = 1$ in (37), then K_n^* is just the non-weighted s.e.p. of residuals. We shall introduce a more general process that can accommodate all above cases, and examine the asymptotic behavior of this general process.

Let $a = (a_1, a_2, ..., a_n)$, $b = (b_1, b_2, ..., b_n)$ be two $1 \times n$ random vectors, and $C = (c_1, c_2, ..., c_n)'$ be a $n \times q$ random matrix ($q \ge 1$). Define

$$K_n(s,z,a,b) = (C'C)^{-1/2} \sum_{t=1}^{[ns]} c_t \left\{I(\varepsilon_t \le z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z)\right\}.$$

For $c_t = x_t, a_t = 0$, and $b_t = x_t' n^{1/2}(\hat{\beta} - \beta)$, or a_t and b_t in (38), we have

$$K_n(s,z,a,b) = K_n^*(s,z) \quad \text{and moreover,} \quad K_n(s,z,0,0) = H_n(s,z). \tag{39}$$

Define

$$Z_n(s,z,a,b) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} c_t \left\{ I(\varepsilon_t \le z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) \right\}.$$

Assume

(B.1) The variable ε_t is independent of $\mathcal{F}_{t-1}$, where

$$\mathcal{F}_{t-1} = \sigma - \text{field}\{a_{s+1}, b_{s+1}, c_{s+1}, \varepsilon_s;\ s \le t-1\}.$$

(B.2) $n^{-1} \sum_{t=1}^{n} \|c_t\| = O_p(1)$.

(B.3) $n^{-1/2} \max_{1 \le i \le n} |\eta_i| = o_p(1)$, for $\eta_i = a_i, b_i$.

(B.4) There exist a $\gamma > 1$ and $A < \infty$ such that for all n

$$E\{\frac{1}{n} \sum_{t=1}^{n} \|c_t\|^2 (|a_t| + |b_t|)\}^\gamma < A \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^{n} E\{\|c_t\|^2 (|a_t| + |b_t|)\}^\gamma < A.$$

(B.5) Condition (B.3) and (B.4) with $|b_t|$ replaced by $\|x_t\|$.

Note that under (B.1) the summands in Z_n are conditionally centered.

Theorem A.2 *Under the assumptions of (A.1), (B.1)-(B.5)*

$$K_n(s,z,a,b) = K_n(s,z,0,0) +$$
$$(C'C/n)^{-1/2} \left\{ f(z) z \left(\frac{1}{n} \sum_{t=1}^{[ns]} c_t a_t \right) + f(z) \left(\frac{1}{n} \sum_{t=1}^{[ns]} c_t b_t \right) \right\} + o_p(1)$$

where the $o_p(1)$ is uniform in s and in z, and for $b_t = x_t'\alpha$, the $o_p(1)$ is also uniform in $\alpha \in D$, an arbitrary compact set of R^p. In particular, the result holds for $b_t = x_t' n^{1/2}(\hat{\beta} - \beta)$ as long as $n^{1/2}(\hat{\beta} - \beta) = O_p(1)$.

Proof: By adding and subtracting terms,

$$\begin{aligned} K_n(s,z,a,b) &= K_n(s,z,0,0) \\ &+ Z_n(s,z,a,b) - Z_n(s,z,0,0) \\ &+ (C'C)^{-1/2} \sum_{t=1}^{[ns]} c_t \left\{ F(z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z) \right\}. \end{aligned}$$

Theorem A.2 now follows from Theorem A.3(i) and (ii) below and the Taylor series expansion. □

Theorem A.3 *(i) Under assumptions (A.1) and (B.1)-(B.4),*

$$\sup_{0\leq s\leq 1, z\in R} \|Z_n(s,z,a,b) - Z_n(s,z,0,0)\| = o_p(1).$$

(ii) Let $b_t = x_t'\alpha$ for α in a compact set D of R^p and denote $b(\alpha) = (x_1'\alpha, ..., x_n'\alpha)$. Then under assumptions (A.1) and (B.1), (B.2) and (B.5)

$$\sup_{\alpha\in D}\sup_{0\leq s\leq 1, z\in R} \|Z_n(s,z,a,b(\alpha)) - Z_n(s,z,0,0)\| = o_p(1).$$

(iii) Let $a_t = r_t'\tau$; $r_t, \tau \in R^\ell$ for some $\ell \geq 1$; $\tau \in S$, a compact set. Denote $a(\tau) = (r_1'\tau, ..., r_n'\tau)$. Assume (B.3) and (B.4) hold with $|a_t| = \|r_t\|$. Then under (A.1), (B.1) and (B.2)

$$\sup_{\tau\in S}\sup_{\alpha\in D}\sup_{0\leq s\leq 1, z\in R} \|Z_n(s,z,a(\tau),b(\alpha)) - Z_n(s,z,0,0)\| = o_p(1).$$

Note that part (i) is a special case of part (ii). Similarly, (ii) is a special case of (iii). However, each of the latter is also a consequence of its former, as will be shown. Part (ii) allows b_t to depend, in a particular way, on the entire data set. An example is $b_i = x_i'\sqrt{n}(\hat\beta - \beta)$ as long as $\sqrt{n}(\hat\beta - \beta) = O_p(1)$. Similarly, part (iii) allows scale parameter to be estimated. In our application, part (ii) is all that is required. To prove the theorem, we need the following lemma.

Lemma A.3 *Under assumption (A.1) and (B.1)-(B.4), for every $d \in (0, 1/2)$*

$$\sup_{y,z} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|c_t F(y_t^*) - c_t F(z_t^*)\| = o_p(1)$$

where $y_t^* = y(1 + a_t n^{-1/2}) + b_t n^{-1/2}$, $z_t^* = z(1 + a_t n^{-1/2}) + b_t n^{-1/2}$ and the supremum extends over all pair of (y, z) such that $|F(y) - F(z)| \leq n^{-1/2-d}$.

Proof: Follows from the mean value theorem. □

Proof of (i). Let $N(n)$ be an integer such that $N(n) = [n^{1/2+d}] + 1$, where d is defined in Lemma A.3. Following the arguments of Boldin (1982), divide the real line into $N(n)$ parts by points $-\infty = z_0 < z_1 < \cdots < z_{N(n)} = \infty$ with $F(z_i) = iN(n)^{-1}$. As explained in the proof of Lemma A.2, there is no loss of generality by assuming $c_j \geq 0$. Then $c_t I(\varepsilon_t \leq z)$ and $c_t F(z)$ are nondecreasing. Thus when $z_r < z < x_{r+1}$, we have

$$
\begin{aligned}
&Z_n(s,z,a,b) - Z_n(s,z,0,0)\\
&\leq Z_n(s,z_{r+1},a,b) - Z_n(s,z_{r+1},0,0)\\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} c_t\{I(\varepsilon_t \leq z_{r+1}) - F(z_{r+1}) - I(\varepsilon_t \leq z) + F(z)\}\\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} c_t\{F(z_{r+1}(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z(1 + a_t n^{-1/2}) + b_t n^{-1/2})\}.
\end{aligned}
$$

The reverse inequality holds when z_{r+1} is replaced by z_r. Therefore, by the inequality $|y| \leq \max(|c|, |d|)$ for $c \leq y \leq d$,

$$
\begin{aligned}
\sup_{s,z} \|Z_n(s,z,a,b) - Z_n(s,z,0,0)\| &\leq \max_r \sup_s \|Z_n(s,z_r,a,b) - Z_n(s,z_r,0,0)\|\\
&+ \sup_{s,|u-v|\leq N(n)^{-1}} \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{[ns]} c_t\{I(\varepsilon_t \leq F^{-1}(u)) - u - I(\varepsilon_t \leq F^{-1}(v)) + v\} \right\|\\
&+ \sup_s \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{[ns]} c_t\{F(z_{r+1}(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z_r(1 + a_t n^{-1/2}) + b_t n^{-1/2})\} \right\|
\end{aligned}
$$

Because $\| \sum_{t=1}^{[ns]} \cdot \| \leq \sum_{t=1}^{n} \| \cdot \|$ and $|F(z_{r+1}) - F(z_r)| \leq n^{-1/2-d}$ by construction, the last term on the right is $o_p(1)$ by Lemma A.3. The second last term is $o_p(1)$ because of Theorem A.1. It remains to show

$$
\max_{0\leq r\leq N(n)} \max_{1\leq j\leq n} \|Z_n^*(j/n, z_r)\| = o_p(1) \tag{40}
$$

where $Z_n^*(j/n, z_r) := Z_n(s,z_r,a,b) - Z_n(s,z_r,0,0)$. But

$$
P(\max_{0\leq r\leq N(n)} \max_{1\leq j\leq n} \|Z_n^*(j/n,z_r)\| > \epsilon) \leq N(n) \max_r P(\max_j \|Z_n^*(j/n,z_r)\| > \epsilon).
$$

The remaining task is to bound the above probability. Let

$$
\xi_t = c_t \left\{ I\left(\varepsilon_t \leq z_r(1 + \frac{1}{\sqrt{n}}a_t) + \frac{1}{\sqrt{n}}b_t\right) - F\left(z_r(1 + \frac{1}{\sqrt{n}}a_t) + \frac{1}{\sqrt{n}}b_t\right) - I(\varepsilon_t \leq z_r) + F(z_r) \right\}
$$

then $(\xi_t, \mathcal{F}_t)$ is an array of martingale differences and

$$Z_n^*(j/n, z_r) = n^{-1/2}\sum_{t=1}^{j}\xi_t.$$

By the Doob inequality,

$$P(\max_j \|n^{-1/2}\sum_{t=1}^{j}\xi_t\| > \epsilon) \le \epsilon^{-2\gamma}M_1 E\|n^{-1/2}\sum_{t=1}^{n}\xi_t\|^{2\gamma}, \tag{41}$$

where M_1 is a constant only depending on p and γ. By the Rosenthal inequality (Hall and Heyde, 1980, p. 23), there exists $M_2 > 0$, such that

$$E(\|\sum_{t=1}^{n}\xi_t\|)^{2\gamma} \le M_2 E\{\sum_{t=1}^{n}E(\|\xi_t\|^2|\mathcal{F}_{t-1})\}^{\gamma} + M_2\sum_{t=1}^{n}E\|\xi_t\|^{2\gamma} \tag{42}$$

for all n. Because (a_i, b_i, c_i) is measurable with respect to $\mathcal{F}_{i-1}$ and ε_i is independent of $\mathcal{F}_{i-1}$ by (B.1),

$$E(\|\xi_i\|^2|\mathcal{F}_{i-1}) \le \|c_i\|^2\{F(z_r(1+a_i n^{-1/2}) + b_i n^{-1/2}) - F(z_r)\} \le \frac{1}{\sqrt{n}}\|c_i\|^2 L(|a_i| + |b_i|)$$

where L is an upper bound for both $|f(x)|$ and $|xf(x)|$ for all x. Using the above inequality and $E\|\xi_i\|^{2\gamma} = E\{E(\|\xi_i\|^{2\gamma}|\mathcal{F}_{i-1})\}$, we have

$$E\|\xi_i\|^{2\gamma} \le n^{-\gamma/2}L^{\gamma}E\{\|c_i\|^2(|a_i| + |b_i|)\}^{\gamma}.$$

By (42), for $M_3 = M_2 L^{\gamma}$,

$$\begin{aligned} E(n^{-1/2}\|\sum_{t=1}^{n}\xi_t\|)^{2\gamma} &\le M_3 n^{-\gamma/2}E\{\frac{1}{n}\sum_{t=1}^{n}\|c_t\|^2(|a_t| + |b_t|)\}^{\gamma} \\ &+ M_3 n^{-\gamma/2-(\gamma-1)}\frac{1}{n}\sum_{t=1}^{n}E\{\|c_t\|^2(|a_t| + |b_t|)\}^{\gamma} \\ &\le 2M_3 A n^{-\gamma/2}. \end{aligned}$$

The last inequality follows from assumption (B.4). The above bound does not depend on z_r. Thus for $M_4 = 2M_1 M_3 A$,

$$P(\max_r \max_j |Z_n^*(j/n, z_r)| > \epsilon) \le \epsilon^{-2\gamma}M_4 N(n) n^{-\gamma/2} = \epsilon^{-2\gamma}M_4 n^{-(\gamma-1)/2+d}$$

because $N(n) = n^{1/2+d}$. The above is $o(1)$ if we choose $d \in (0, (\gamma - 1)/2)$ in Lemma A.3. The proof of (i) is completed. □

Proof of (ii). This really follows from the compactness of D. The proof is standard, see Koul (1991), for example. Since D is compact, for any $\delta > 0$, the set D can be partitioned into finite number of subsets such that the diameter of each subset is not greater than δ. Denote these subsets by $D_1, D_2, ..., D_{m(\delta)}$. Fix k and consider D_k. Pick $\alpha_k \in D_k$. For all $\alpha \in D_k$

$$(x_t'\alpha_k - \delta\|x_t\|) \leq x_t'\alpha \leq (x_t'\alpha_k + \delta\|x_t\|)$$

because $\|\alpha_k - \alpha\| \leq \delta$. Thus if we define the vector $b(k, \lambda) = (x_1'\alpha_k + \lambda\|x_1\|, ..., x_n'\alpha_k + \lambda\|x_n\|)$ then assuming again $c_t \geq 0$ for all t, we have for all $\alpha \in D_k$, by the monotonicity of $c_t I(\varepsilon_t \leq z)$,

$$Z_n(s, z, a, b(\alpha)) \leq Z_n(s, z, a, b(k, \delta)) +$$
$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} c_t \left\{F\left(z(1 + a_t n^{-1/2}) + (x_t'\alpha_k + \delta\|x_t\|)n^{-1/2}\right) - F\left(z(1 + a_t n^{-1/2}) + x_t'\alpha n^{-1/2}\right)\right\}$$

and a reversed inequality holds when δ is replaced by $-\delta$. Using the mean value theorem and assumption (A.1), it is easy to verify that the second term on the right is bounded (with respect to the norm $\|\cdot\|$) by $\delta O_p(1)$, where the $O_p(1)$ is uniform in all $s \in [0, 1]$, all $z \in R$, and all $\alpha \in D$. Thus

$$\begin{aligned} \sup_\alpha \sup_{s,z} \|Z_n(s, z, a, b(\alpha)) - Z_n(s, z, 0, 0)\| \\ \leq \max_k \sup_{s,z} \|Z_n(s, z, a, b(k, \delta)) - Z_n(s, z, 0, 0)\| \\ + \max_k \sup_{s,z} \|Z_n(s, z, a, b(k, -\delta)) - Z_n(s, z, 0, 0)\| + \delta O_p(1) \end{aligned}$$

where the supremums are taken over $\alpha \in D$, $s \in [0, 1]$, $z \in R$, and $k \leq m(\delta)$, respectively. The term $\delta O_p(1)$ can be made arbitrarily small in probability by choosing a small δ. Once δ is chosen, $m(\delta)$ will be a bounded integer. The first two terms on the right hand side are then $o_p(1)$ by part (i). □

Proof of (iii). Follows from the same type of arguments as in the proof of (ii). Instead of using the result of part (i), one uses the result of part (ii). The proof of the theorem is now completed. □

We now in the position to prove Theorems 1 through 4. Conditions required for the preliminary results, Theorem A.1 to Theorem A.3 and their corollaries, are all satisfied under (A.1)-(A.9) for various choices of a_t, b_t and c_t below. Conditions (A.3) and (A.9) can be replaced by (A.3$'$) when weighted empirical processes are under consideration.

Proof of Theorem 1 and Theorem 3. Under the null hypothesis, $\hat{\varepsilon}_t = \varepsilon_t - x_t'(\hat{\beta}-\beta)$ so $\hat{\varepsilon}_t \le z$ if and only if $\varepsilon_t \le z + x_t'(\hat{\beta}-\beta)$. Apply Theorem A.2 with $a_t = 0, b_t = x_t'\sqrt{n}(\hat{\beta}-\beta)$, and $c_t = x_t$; in view of (39),

$$K_n^*(s,z) - A_{[ns]}K_n^*(1,z) = H_n(s,z) - A_{[ns]}H_n(1,z) \tag{43}$$

$$+f(z)(X'X/n)^{-1/2}\frac{1}{n}\sum_{t=1}^{[ns]} x_t b_t - f(z)A_{[ns]}(X'X/n)^{-1/2}\frac{1}{n}\sum_{t=1}^{n} x_t b_t \tag{44}$$

$$+o_p(1). \tag{45}$$

Expression (44) is identically zero for all $s \in [0,1]$ when $b_t = x_t'\sqrt{n}(\hat{\beta}-\beta)$. That is, the drift terms of $K_n^*(s,z)$ and $A_{[ns]}K_n^*(1,z)$ are canceled out. Theorem 3 now follows from Corollary A.2. Theorem 1(ii) follows as a special case. To prove Theorem 1(i), take $x_t = 1$ and $A_{[ns]} = [ns]/n$ in the above proof, then (44) becomes

$$f(z)\frac{1}{n}\sum_{t=1}^{[ns]} b_t - f(z)\frac{[ns]}{n}\sum_{t=1}^{n} b_t = f(z)\left(\frac{1}{n}\sum_{t=1}^{[ns]} x_t - \frac{[ns]}{n}\sum_{t=1}^{n} x_t\right)\sqrt{n}(\hat{\beta}-\beta), \tag{46}$$

which is $o_p(1)$ under assumptions (A.7) and (A.9). The limiting process of $H_n(s,z) - A_{[ns]}H_n(1,z)$ reduces to the one stated in Theorem 1(i) when $x_t = 1$ for all t. □

Proof of Theorem 2. Under the local alternatives (11), K_n^* is given by (37) with a_t and b_t given by (38). Note that under these local alternatives, the root-n consistency of $\hat{\beta}$ generally prevails. For example, assuming the ε_t have a finite variance, least squares estimator of β is still root-n consistent. The root-n consistency allows us to obtain a non-explosive limit (otherwise the tests will be consistent even for local changes). Note that b_t is dominated by $x_t'\sqrt{n}(\hat{\beta}-\beta)+\Delta_1 x_t'g(t/n)$, with the remaining term being negligible in the limit. Moreover, when $b_t = x_t'\sqrt{n}(\hat{\beta}-\beta)$, from the previous proof, the drift term of $K_n^*(s,z) - A_{[ns]}K_n^*(1,z)$ is negligible for either $c_t = x_t$ or $c_t = 1$. We can thus assume $b_t = \Delta_1 x_t' g(t/n)$. Let $c_t = x_t$. Now by Theorem A.2, for $a_t = \Delta_2 h(t/n)$

$$K_n^*(s,z) - A_{[ns]}K_n^*(1,z) = H_n(s,z) - A_{[ns]}H_n(1,z)$$
$$+f(z)z\Delta_2\left\{(\frac{X'X}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{[ns]} x_t h(t/n) - A_{[ns]}(\frac{X'X}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{n} x_t h(t/n)\right\} \quad (47)$$
$$+f(z)\Delta_1\left\{(\frac{X'X}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' g(t/n) - A_{[ns]}(\frac{X'X}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{n} x_t x_t' g(t/n)\right\} (48)$$
$$+o_p(1)$$

By the results of KPA, under (A.3) and (A.9)

$$\frac{1}{n}\sum_{t=1}^{[ns]} x_t h(t/n) \xrightarrow{p} \bar{x}\int_0^s h(v)dv, \quad (49)$$
$$\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' g(t/n) \xrightarrow{p} Q\int_0^s g(v)dv. \quad (50)$$

Furthermore, under assumption (A.3),

$$A_{[ns]}(X'X/n)^{-1/2} \xrightarrow{p} sQ^{-1/2}. \quad (51)$$

From these results, (47) converges to $f(z)z\Delta_2 Q^{-1/2}\bar{x}\lambda_h(s)$ where λ_h is given by (13); and similarly, (48) converges to $f(z)\Delta_1 Q^{1/2}\lambda_g(s)$ where λ_g is given by (12). Thus (15) is obtained and (14) is obtained similarly by choosing $c_t = 1$. The proof is completed.

□

Proof of (18) and (19). Now $\hat{\varepsilon}_t \le z$ if and only if $\varepsilon_t \le x_t'(\hat{\beta}-\beta) + \Delta z_t' g(t/n) n^{-1/2}$. Let $a_t = 0$, $b_t = x_t'\sqrt{n}(\hat{\beta}-\beta) + \Delta z_t' g(t/n)$. Again we can ignore $x_t'\sqrt{n}(\hat{\beta}-\beta)$ in b_t and assume $b_t = \Delta\, z_t' g(t/n)$. For $c_t = 1$ or $c_t = x_t$, the drift term of $K_n^*(s,z) - A_{[ns]}K_n^*(1,z)$ is given by

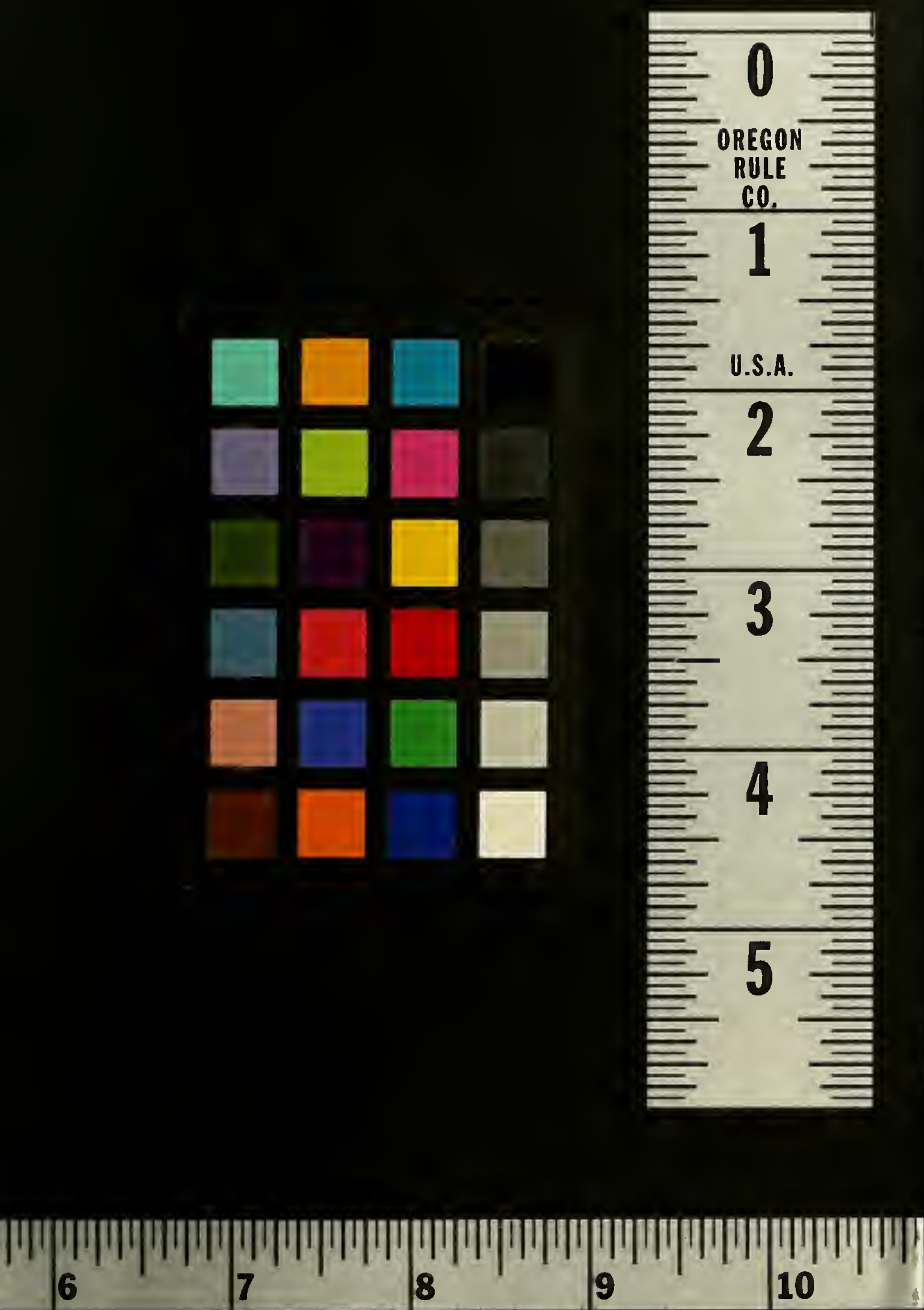
$$f(z)\Delta\left\{(\frac{C'C}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{[ns]} c_t z_t' g(t/n) - A_{[ns]}(\frac{C'C}{n})^{-1/2}\frac{1}{n}\sum_{t=1}^{n} c_t z_t' g(t/n)\right\}$$

plus an $o_p(1)$ term. Finally, if $c_t = 1$ then (18) follows; and if $c_t = x_t$, then (19) follows. □

Proof of Theorem 4. The proof is virtually identical to that of Theorem 2, except under (A.3′), (49)-(51) are replaced by

$$\frac{1}{n}\sum_{t=1}^{[ns]} x_t h(t/n) \xrightarrow{p} \int_0^s \frac{dQ(v)e}{dv} h(v)dv,$$
$$\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' g(t/n) \xrightarrow{p} \int_0^s \frac{dQ(v)}{dv} g(v)dv,$$
$$A_{[ns]}(X'X/n)^{-1/2} \xrightarrow{p} Q(1)^{-1/2}Q(s)Q(1)^{-1},$$

respectively, where $Q(v)e$ is the first column of $Q(v)$. The first convergence is a special case of the second due to the presence of a constant regressor. □


0
OREGON
RULE
CO.
1
U.S.A.
2
3
4
5
6
7
8
9
10