ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Econometrica, Vol. 64, No. 3 (May, 1996), 597–622

TESTING FOR PARAMETER CONSTANCY IN LINEAR REGRESSIONS: AN EMPIRICAL DISTRIBUTION FUNCTION APPROACH

BY JUSHAN BAI[1]

This paper proposes some tests for parameter constancy in linear regressions. The tests use weighted empirical distribution functions of estimated residuals and are asymptotically distribution free. The local power analysis reveals that the proposed tests have nontrivial local power against a wide range of alternatives. In particular, the tests are capable of detecting error heterogeneity that is not necessarily manifested in the form of changing variances. The model allows for both dynamic and trending regressors. The residuals may be obtained based on any root-n consistent estimator (under the null) of regression parameters. As an intermediate result, some weak convergence for (stochastically) weighted sequential empirical processes is established.

KEYWORDS: Structural change, empirical distribution function, sequential empirical process, weak convergence, two-parameter Brownian bridge.

1. INTRODUCTION

MANY ECONOMIC FACTORS may 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 in the Phillips curve over time serve as one illustration (Alogoskoufis and Smith (1991)). As a result, model stability has always been an important concern in econometric modeling; see, for example, Chow (1960) and Quandt (1960) for earlier studies and Andrews (1993) and the references therein for more recent ones. The purpose of this paper is to provide additional tools for the diagnosis of parameter instability in linear regressions.

Two classes of tests are proposed, resembling the prototypical Kolmogorov-Smirnov two-sample test. The first class is based on nonweighted sequential empirical processes of residuals. This class has received considerable attention in the i.i.d. context, for example, Csörgő and Horváth (1987, 1988), Deshayes and Picard (1986), and Szyszkowicz (1994), among others. Carlstein (1988) and Dümbgen (1991) proposed to estimate a break point under the alternative

[1] This paper is developed from a chapter of my dissertation written at the University of California at Berkeley. I thank Professors Tom Rothenberg, James Stock, and Deborah Nolan for their advice. I also thank seminar participants at Yale and Harvard/MIT for very useful comments on an earlier version of this paper. Comments from three anonymous referees and a co-editor led to a substantially improved presentation. Finally, financial support from an Alfred Sloan Foundation dissertation fellowship is gratefully acknowledged. All remaining errors are my own responsibility.

hypothesis based on these test statistics. We extend this class of tests to apply to regression models with estimated parameters.

The first class of tests has limited applicability in time series regressions because the tests will no longer be asymptotically distribution free when trending regressors are included in the regression model. In this case, the second class of tests can be considered, obtained by constructing a weighted empirical process of residuals. It is interesting to note that we can construct asymptotically distribution free tests by choosing weighting vectors, in a natural way, in the construction of the empirical processes upon which our tests are based. This is in contrast to the well known result that goodness-of-fit tests based on empirical processes involving estimated parameters will generally depend upon both the estimated parameters and the underlying error-distribution function even in the limit (see Durbin (1973)).

Testing for parameter constancy in regressions is a much studied subject. Various test statistics are proposed in the literature, e.g., Brown, Durbin, and Evans (1975), Gombay and Horváth (1994), Hawkins (1988, 1989), Huskova (1991), Jandhyala (1993), Kim and Siegmund (1989), and Ploberger, Kramer, and Kontrus (1989) to name a few. In a time series regression that allows for integrated and co-integrated variables, tests are proposed by Hansen (1992), Perron and Vogelsang (1992), and Zivot and Andrews (1992). More recently, Andrews and Ploberger (1994) propose some optimal tests. However, tests based on sequential empirical processes for regression models are not as well studied in the literature. In the ARMA context, Bai (1991) considers a nonweighted test, in which stationarity and zero mean under the null are heavily used. We consider in this paper a regressor-weighted test. To derive its limiting distribution, we also establish some convergence results for stochastically weighted sequential empirical processes.

This paper is organized as follows. Section 2 specifies the models and describes the assumptions. Section 3 defines the test statistics. Section 4 examines the local power of the tests. Trending regressors are considered in Section 5. Section 6 concludes. Technical materials are collected in the Appendix.

2. MODELS AND ASSUMPTIONS

The null hypothesis specifies the regression model:

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

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 disturbances ε_t are i.i.d. with distribution function F.

The alternative hypothesis specifies the following model:

$$y_t = x_t' \beta_t + \varepsilon_t^* \qquad (t = 1, 2, \ldots, n), \tag{2}$$

where β_t may not be constant over time and/or the disturbances ε_t^* may not be identically distributed. When examining the power property for the proposed tests, we consider the local alternative described by (13) and a more general type of alternatives given in Section 4.2 below.

In what follows, the norm $\|\cdot\|$ represents the Euclidean norm, i.e. $\|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ for $x \in R^p$. For vectors x and y, we write $x \le y$ if the inequality holds true for each coordinate. Furthermore, $[\cdot]$ denotes the greatest integer function, $u \wedge v = \min\{u, v\}$, and $u \vee v = \max\{u, v\}$.

We make the following assumptions with implications discussed below:

(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 $f(z)$ and $zf(z)$ are assumed to be uniformly continuous on the real line. Furthermore, there exists a finite number L such that $|zf(z)| < L$ and $|f(z)| < L$ for all z. 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*

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

where Q is a $p \times p$ nonrandom positive definite matrix.

(A.4) $\max_{1 \le t \le n} n^{-1/2} \|x_t\| = o_p(1)$.

(A.5) *There exist a random variable Z_n and a constant κ $(1/2 > \kappa \ge 0)$ such that for all s and s_1 $(s, s_1 \in [0,1]$ and $s \ge s_1)$,*

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

In addition, for some $\rho > 2$ and $M < \infty$:

$$P(|Z_n| > C) < M/C^{\rho}. \tag{3}$$

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

$$\frac{1}{n} \sum_{i<t\le j} E(x_t' x_t)^{\gamma} \le K(v-u) \quad \textit{and} \quad E\left(\frac{1}{n} \sum_{i<t\le j} x_t' x_t\right)^{\gamma} \le K(v-u)^{\alpha}, \tag{4}$$

where $i=[nu]$, $j=[nv]$. *Because* $|v-u|\le 1$, *we can assume* $\alpha\le\gamma$. *We shall choose* α *such that, for* κ *in* (*A*.5),

$$(5)\qquad \frac{\gamma-1}{\alpha-1}>1+2\kappa,$$

which is possible by choosing α *close to* 1.

(A.7) *Let* $\hat{\beta}$ *be an estimator of* β. *Under the null hypothesis,*

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

where $X=(x_1,x_2,\dots,x_n)'$.

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

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

(A.9) *Finally,*

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

where $\bar{x}$ *is a* $p\times 1$ *constant vector.*

We make some comments pertaining to these assumptions. Assumption (A.1) is typical for residual empirical processes; see Boldin (1989), Koul (1984, 1992), and Kreiss (1991). Assumption (A.2) allows for dynamic variables. Assumptions (A.3) and (A.9) are needed to assure that the nonweighted test M_n (below) is asymptotically distribution free. These two assumptions, however, are redundant for the weighted test M_n^* (below). Moreover, (A.3) and (A.9) rule out trending regressors, which are discussed separately in Section 5. When a constant regressor is included, (A.9) is implied by (A.3). Assumption (A.4) is conventional for linear models and is used for obtaining normality. Assumptions (A.5) and (A.6) assure the tightness of sequential empirical processes (Theorem A.1 in the Appendix). In (A.5), $Z_n n^{\kappa}$ may be taken to be $\psi_n=\max_{1\le i<j\le n}(j-i)^{-1}\sum_{t=i}^{j}\|x_t\|$. When $E\|x_t\|^{\rho}<M$ for all t ($\rho>2$), it can be shown that (A.5) holds for $\kappa=1/\rho$. When $E(x_t'x_t)^2\le M$ for all t, then the first half of (A.6) 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-Schwarz inequality. Furthermore, by choosing $\alpha=3/2$, inequality (5) is also satisfied. When the disturbances are i.i.d. and have finite variance, Assumption (A.7) is fulfilled by the least squares estimator. For infinite variance models, robust estimation such as the LAD has to be used to assure (A.7). Finally, Assumption (A.8) is used to prove the tightness of sequential empirical processes based on estimated residuals. It can be shown that when the sixth moment of the regressors is uniformly bounded, (A.4)–(A.6) and (A.8) are all satisfied. In particular, these assumptions are satisfied when the regressors are themselves uniformly bounded.

3. THE TEST STATISTICS

The test statistics are based on estimated residuals. The model is estimated under the null hypothesis. Let $\hat{\beta}$ be an estimator of β (e.g., least squares estimator) and $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$. Let us first introduce the nonweighted test. For each fixed k, define the empirical distribution function (e.d.f.) based on the first k residuals as

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

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

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

where $I(\cdot)$ is the indicator function. Further define

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

and the test statistic

$$M_n = \max_k \sup_z |T_n(k/n, z)|,$$

where the max is taken over $1 \le k \le n$ and the supremum with respect to z 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 weight $[(k/n)(1-k/n)]^{1/2}$. Thus the test M_n looks for the maximum value of weighted Kolmogorov-Smirnov statistics for all possible sample splits. This test was considered by Bai (1991) for testing changes in the innovations of an ARMA process, and is an extension of Csörgő and Horváth (1987, 1988) and Deshayes and Picard (1986) for i.i.d. settings. To obtain the asymptotic null distribution for M_n, we shall prove the weak convergence for $T_n(s, z)$ and then apply the continuous mapping theorem.

We have the following identities:

$$T_n\left(\frac{k}{n}, z\right) = n^{-1/2}\sum_{t=1}^{k} I(\hat{\varepsilon}_t \le z) - \frac{k}{n} n^{-1/2}\sum_{t=1}^{n} I(\hat{\varepsilon}_t \le z) \tag{6}$$

$$= n^{-1/2} \sum_{t=1}^{k} \{I(\hat{\varepsilon}_t \le z) - F(z)\} \tag{7}$$
$$- \frac{k}{n} n^{-1/2} \sum_{t=1}^{n} \{I(\hat{\varepsilon}_t \le z) - F(z)\}.$$

Equation (7) holds for an arbitrary function F, although F will be assumed to be the distribution function of ε_t. Writing in the form (7) will be convenient for studying the limiting distribution of T_n and hence of M_n.

As will be shown, the test M_n is asymptotically distribution free and has nontrivial local power against changes in the scale parameter of the disturbances. However, like the CUSUM test, when testing shifts in the regression parameters local power disappears if the mean regressor is zero; see Kramer, Ploberger, and Alt (1988, henceforth KPA). In addition, if a trending regressor exists, M_n will not be asymptotically distribution free. To circumvent these undesirable features, we introduce a new class of tests based on the regressor-weighted empirical distribution functions of residuals. Let $X_k = (x_1, \ldots, x_k)'$ and

$$A_k = (X'X)^{-1/2}(X_k'X_k)(X'X)^{-1/2}. \tag{8}$$

Analogous to (6), define the $p \times 1$ vector process T_n^*,

$$T_n^*\left(\frac{k}{n}, z\right) = (X'X)^{-1/2} \sum_{t=1}^{k} x_t I(\hat{\varepsilon}_t \le z) - A_k (X'X)^{-1/2} \sum_{t=1}^{n} x_t I(\hat{\varepsilon}_t \le z) \tag{9}$$

and the test statistic

$$M_n^* = \max_k \sup_z \left\| T_n^*\left(\frac{k}{n}, z\right) \right\|_\infty$$

where $\|y\|_\infty = \max\{|y_1|, \ldots, |y_p|\}$, the maximum norm. The process T_n^* and test M_n^* reduce to T_n and M_n, respectively, when $x_t = 1$ for all t. The process T_n^* only takes on n^2 different values; its maximum value gives rise to M_n^*. The actual computation of M_n^* is straightforward. A programmable formula for M_n^* is given by

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

where D_i is the location (index) of the ith order statistic $\hat{\varepsilon}_{(i)}$.

If there is a constant regressor (and we shall assume this), then the following identity holds:

$$(X'X)^{-1/2} \sum_{t=1}^{k} x_t - A_k (X'X)^{-1/2} \sum_{t=1}^{n} x_t = 0, \quad \forall k, \tag{11}$$

so that $T_n^*(k/n, z)$ can be written alternatively as

$$(12) \qquad (X'X)^{-1/2}\sum_{t=1}^{k} x_t'\{I(\hat{\varepsilon}_t \le z) - F(z)\} \\ - A_k(X'X)^{-1/2}\sum_{t=1}^{n} x_t'\{I(\hat{\varepsilon}_t \le z) - F(z)\},$$

for any F. Again, F will be chosen to be the distribution function of ε_t. Expression (12) is a weighted version of (7) and is useful for deriving the limiting process of T_n^*. The choice of matrix A_k turns out to be important. It plays two key roles. First, because of (11), we can express T_n^* in (12), which is a conditionally centered process if $\hat{\varepsilon}_t$ is replaced by ε_t. This is why (12) is useful for studying the limiting process of T_n^*. Second, the choice of A_k makes the test statistic asymptotically distribution free under the null.

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

$$E\{B(r,u)B(s,v)\} = (r \wedge s - rs)(u \wedge v - uv),$$

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 Skorohod J_1 topology; see Pollard (1984).

THEOREM 1: *Under model* (1) *and Assumptions* (A.1)–(A.9),

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

and

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

where $B^* = (B_1, B_2, \ldots, B_p)'$ *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. $\sup_{0 \le u \le 1}\sup_{0 \le v \le 1}|B(u,v)|$. We have the following from the continuous mapping theorem:

COROLLARY 1: *Under the assumptions of Theorem* 1,

$$\lim_{n\to\infty} P(M_n \le a) = G(a), \quad a > 0$$

and

$$\lim_{n\to\infty} P(M_n^* \le a) = [G(a)]^p, \quad a > 0.$$

Thus the tests are asymptotically distribution free despite parameter estimation. Some selected critical values are reported in Table I. These values are obtained via simulation with 100,000 repetitions and $n = 200$. In each repetition, a sequence of i.i.d. uniformly distributed random variables on $[0,1]$ is generated. The process $T_n(k/n, z)\,(0 \le z \le 1)$ is constructed using this sequence. The value of M_n is then obtained by maximizing $T_n(k/n, z)$ with respect to k and z.

OTHER TESTS: Besides the sup-type tests, the mean-type test can be used. Let

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

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 $B_1, \ldots, B_p$ are independent copies of $B(\cdot,\cdot)$. Many other tests can be constructed based on the weak convergence of Theorem 1.

4. LOCAL POWER ANALYSIS

Two types of alternatives will be considered. The first type is associated with changes in regression parameters and scales. The second is associated with

TABLE I

SELECTED ASYMPTOTIC QUANTILES OF THE TEST M_n^*

$\lim_{n\to\infty} P(M_n^* < x) = \alpha$

	α			
p	85%	90%	95%	99%
1	0.712	0.750	0.811	0.935
2	0.773	0.809	0.866	0.980
3	0.806	0.841	0.897	1.005
4	0.829	0.864	0.916	1.018
5	0.848	0.882	0.933	1.032
6	0.862	0.894	0.945	1.042
7	0.874	0.906	0.955	1.053
8	0.884	0.915	0.964	1.060
9	0.891	0.923	0.972	1.063
10	0.901	0.932	0.979	1.068
11	0.908	0.938	0.984	1.074
12	0.914	0.943	0.990	1.085
13	0.919	0.948	0.993	1.087
14	0.923	0.953	0.997	1.091
15	0.930	0.957	1.004	1.095
16	0.934	0.962	1.007	1.098
17	0.938	0.966	1.010	1.101
18	0.941	0.971	1.013	1.104
19	0.945	0.974	1.015	1.107
20	0.948	0.976	1.018	1.109

changes in error distribution functions. Examination of the second type shows that the proposed tests are able to detect changes that may occur beyond the second moment.

4.1. *Changes in Regression Parameters and Scales*

We consider model (2) with the class of local alternatives studied by KPA:

$$\beta_t = \beta + \Delta_1 g(t/n)n^{-1/2} \quad \text{and} \quad \varepsilon_t^* = \varepsilon_t(1 + \Delta_2 h(t/n)n^{-1/2})^{-1} \tag{13}$$

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

$$\lambda_g(s) = \int_0^s g(v)\,dv - s\int_0^1 g(v)\,dv \tag{14}$$

and the function

$$\lambda_h(s) = \int_0^s h(v)\,dv - s\int_0^1 h(v)\,dv. \tag{15}$$

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

THEOREM 2: *Under Assumptions* (A.1)–(A.9) *and the local alternatives* (13), *we have*

$$M_n \xrightarrow{d} \sup_{0 \le s \le 1} \sup_{0 \le u \le 1} |B(s,u) + \Delta_1 p(u)\bar{x}'\lambda_g(s) + \Delta_2 q(u)\lambda_h(s)| \tag{16}$$

and

$$M_n^* \xrightarrow{d} \sup_{0 \le s \le 1} \sup_{0 \le u \le 1} \|B^*(s,u) + \Delta_1 p(u)Q^{1/2}\lambda_g(s) + \Delta_2 q(u)Q^{-1/2}\bar{x}\lambda_h(s)\|_\infty \tag{17}$$

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

Several observations are made here. First, when $\Delta_2 = 0$, Theorem 2 gives rise to the limiting distribution for changes in regression parameters only, and when $\Delta_1 = 0$, the theorem reduces to the case of changes in scale only. Second, 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 and Ploberger and Kramer (1990, 1992). The test M_n^*, however, does have local power irrespective of the relationship between $\bar{x}$ and g. Thus M_n^*

behaves like the fluctuation test of PKK. Third, when testing for a shift in the scale parameter, the local power of M_n^* vanishes if each component of the regressor mean, $\bar{x}$, is zero. Of course, if a constant regressor is included, then M_n^* will have nontrivial local power for testing changes in variance.

The classical statistical literature (e.g., Durbin (1976)) suggests that the Kolmogorov-Smirnov test is less efficient than the t test for a mean shift under normality assumption. It is reasonable to expect that the tests proposed in this paper are less efficient than the sup-F type tests under normality. However, the proposed tests are more efficient for heavy-tailed distributions. This is also confirmed by a small Monte Carlo simulation. The simulation considers a sequence of i.i.d. random variables ($n = 100$) with a mean shift in the middle ($n/2$) and with a magnitude of shift c. We compare the power of T_n and that of the sup-Wald test (e.g., Andrews (1993) with $\pi_0 = 0.05$) at the 5% significance level. For normal random variables, the number of rejections from 1,000 repetitions is 509 with $c = 0.5$ and 984 with $c = 1.0$ for the test T_n and 502 and 989, respectively, for the sup-Wald test. For double exponential random variables, the corresponding result is 527 and 971 for T_n and 260 and 842 for sup-Wald test. Thus the test T_n performs better than sup-Wald for heavy-tailed distributions. It also fares well even for normal random variables.

REMARK 1: When the regressor x_t contains endogenous variables, as in the case of a structural equation of a simultaneous equations system, the weighting vector should be replaced by a vector of instrumental variables. Similar tests can then be constructed. For the instrumental-variable weighted test to have non-trivial local power, the instrumental variables must be valid in the usual sense. That is, the instrumental variables are uncorrelated with the disturbances and correlated with the regressors. The details can be found in an earlier version of this paper.

REMARK 2: Upon the rejection of the null hypothesis, it is often of interest to estimate the shift point as well as the pre-shift and post-shift parameters if a one-time shift model is thought to be appropriate. The point $\hat{k}$ of $(\hat{k}, \hat{z})$ at which $T_n(k/n, z)$ is maximized may serve as a reasonable estimator for the shift point (Dümbgen (1991)). Once the shift point is obtained, it is straightforward to estimate the pre- and post-shift parameters. Another framework for estimating the changing regression parameters and changing variances is proposed by Robinson (1989, 1991). Robinson's approach is nonparametric and is suitable for parameter shifts in the form of (13).

4.2. *General Type of Alternatives*

We now consider a more general type of changes in the error distribution functions. Although changes in regression parameters can be treated as a special case of this general type, we shall assume there is no change in the regression parameters and instead focus on changes in error distributions. The tests developed in this paper are capable of detecting heteroskedasticity other than in the form of changing variances. Let

$$y_{nt} = x'_{nt}\beta + \varepsilon_{nt} \qquad (t = 1, 2, \ldots, n),$$

where ε_{nt} has a density function $F_{nt}(t = 1, 2, \ldots, n)$. Assume F_{nt} admits a density function f_{nt}. The null hypothesis is that $F_{nt} = F$ for all $t \le n$, where F is a density function not necessarily known. Let $\hat{\varepsilon}_{nt} = y_{nt} - x'_{nt}\hat{\beta}$ for some estimator $\hat{\beta}$. We shall assume that β can be consistently estimated as in (A.7) under the local alternatives. Even under nonlocal alternatives, this assumption is still reasonable as long as the means of ε_{nt} are zero and the variances are uniformly bounded (recall the least squares estimator can be root-n consistent under heteroskedasticity).

We consider the nonweighted test M_n only for simplicity and examine the behavior of the test under both fixed and local alternatives. The fixed alternative is specified as

$$H_1: \quad F_{nt} = F, \text{ for } t \le [n\tau] \text{ and } F_{nt} = G, \text{ for } t > [n\tau],$$

where $F \neq G$. The local alternative is

$$H_2: \quad F_{nt} = F, \text{ for } t \le [n\tau] \text{ and}$$
$$F_{nt} = (1 - \Delta n^{-1/2})F + \Delta n^{-1/2}H, \text{ for } t > [n\tau]$$

where $\Delta > 0$ and $\Delta n^{-1/2} < 1$, and $F \neq H$. Thus under the local alternative, the errors $\varepsilon_{nt}\,(t > [n\tau])$ have a mixture distribution.

Let K_F denote a Kiefer process on $[0,1] \times R$ with $K_F(0, \cdot) = 0$. A Kiefer process is a Gaussian process with mean zero and covariance function $E\{K_F(r, y)K_F(s, z)\} = (r \wedge s)\{F(y \wedge z) - F(y)F(z)\}$; see Bickel and Wichura (1971). Let K_G be another Kiefer process independent of K_F with $K_G(0, \cdot) = 0$. Define

$$\overline{K}(s, z) = K_F(s \wedge \tau, z) - sK_F(\tau, z) + K_G(s - s \wedge \tau, z) - sK_G(1 - \tau, z). \tag{18}$$

Then we have the following theorem.

THEOREM 3: *Assume F, G, and H are distribution functions satisfying* $(A.1)$. *Also assume* $(A.2)$–$(A.9)$ *hold. Then*:

(i) *Under the fixed alternative* H_1,

$$M_n = \sup_{s,z} |\overline{K}(s, z) + \sqrt{n}\,(s \wedge \tau)(1 - s \vee \tau)(F - G)| + O_p(1)$$

where $O_p(1)$ is uniform in s and z.

(ii) *Under the local alternative* H_2,

$$M_n \xrightarrow{d} \sup_{s,z} |B(s, F(z)) + \Delta(s \wedge \tau)(1 - s \vee \tau)(F - H)|.$$

The Kiefer processes K_F and K_G are uniformly bounded in probability and consequently $\bar{K}$ is also uniformly bounded in probability. This together with $\sqrt{n}(s \wedge \tau)(1 - s \vee \tau)(F - G)| \to \infty$ (for some s and z if $F \neq G$) implies that the test M_n is consistent under H_1. Part (ii) implies that M_n has nontrivial power in testing local shifts in error distributions. Note that the assumption $F \neq G$ (or $F \neq H$) can be true even though the two distributions have the same mean and same variance.

5. TRENDING REGRESSORS

The computation of the test statistics is still the same with the presence of trending regressors. The limiting distributions of the tests, however, are different. We consider the following model:

$$y_t = z_t'\alpha + \gamma_0 + \gamma_1(t/n) + \cdots + \gamma_q(t/n)^q + \varepsilon_t \tag{19}$$

where z_t is a $r \times 1$ vector of stochastic regressors and $\{z_s; s \leq t-1\}$ are independent of ε_t. Let $x_t = (z_t', 1, t/n, \ldots, (t/n)^q)'$ 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 the fashion of (19) saves notations by eliminating the weighting matrix such as $\mathrm{diag}(n^{-1/2}, \ldots, n^{-(q+1)/2})$ that would otherwise be needed. We shall maintain all assumptions (A.1)–(A.8) of Section 2, except changing (A.3) to

$$\text{(A.3}'\text{)} \qquad \operatorname{plim} \frac{1}{n} \sum_{t=1}^{[ns]} x_t x_t' = \lim \frac{1}{n} E \sum_{t=1}^{[ns]} x_t x_t' = Q(s), \quad \text{uniformly in } s \in [0,1],$$

where $Q(s)$ is positive definite for $s > 0$ and $Q(0) = 0$. Assumption (A.3′) actually admits a much wider class of models than (19).

In the presence of trending regressors only the weighted version, M_n^*, is asymptotically distribution-free, as noted in the Appendix. We shall assume that there is a constant regressor. The process T_n^* and test statistic M_n^* are defined exactly the same as before. Note by (A.3′), we have, uniformly in s,

$$A_{[ns]} \xrightarrow{p} A(s) = Q(1)^{-1/2} Q(s) Q(1)^{-1/2}.$$

THEOREM 4: *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 2: *Under the assumptions of Theorem* 4,

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

The behavior of the test under the local alternative (13) can again be analyzed. Extending Lemma 4 of KPA, we can show that

$$\frac{1}{n} \sum_{t=1}^{[ns]} x_t x_t' g(t/n) \xrightarrow{p} \int_0^s Q^{(1)}(v) g(v)\, dv \tag{20}$$

and the convergence is uniform in s, where $Q^{(1)}(v) = dQ(v)/dv$. The above integral exists if g has bounded variation on $[0,1]$. Of course, we also assume the derivative of $Q(v)$ exists and is integrable. When $Q(v) = vQ(1)$, (20) reduces to the result of KPA. Let

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

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

where $e = (1,0,\ldots,0)'$.

THEOREM 5: *Under the local alternative* (13),

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

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

Again, the test possesses nontrivial local power. Other tests dealing with trending regressors include MacNeill (1978), Sen (1980), Chu and White (1992), among others.

6. CONCLUDING REMARKS

In this paper we propose a class of tests for parameter constancy in linear regressions. The proposed tests are based on regressor-weighted sequential empirical processes. We show that the proposed tests are able to detect changes in regression parameters as well as changes in variances. An important feature is that these tests can detect heteroskedasticity not necessarily manifested in the form of changing variances. In particular, the proposed tests are able to diagnose changes in higher moments or, more generally, changes in error distribution functions, whereas the conventional tests such as the sup-F test may not be suitable for this purpose. These tests are also less sensitive to departure from normality.

The assumption that the disturbances are independent is restrictive. This assumption may be weakened to linear processes on the lines of Boldin (1989) and Kreiss (1991). Another possible extension is to use recursive residuals in constructing the tests. To obtain the limiting distribution, some corresponding weak convergence result needs to be first established.

Dept. of Economics, E52-274B, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Manuscript received February, 1994; final revision received January, 1995.

APPENDIX A: PROOFS

Write

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

then, by (12),

$$(21) \qquad T_n^*\left(\frac{[ns]}{n}, z\right) = K_n^*(s,z) - A_{[ns]} K_n^*(1,z).$$

Thus to study T_n^*, it suffices to study K_n^*. Denote

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

Let $\mathscr{T} = [0,1] \times \mathscr{R}$ be the parameter set with metric $\rho(\{r,y\},\{s,z\}) = |s-r| + |F(z) - F(y)|$. Let $D[\mathscr{T}]$ be the set of functions defined on $\mathscr{T}$ that are right continuous and have left limits. We equip $D[\mathscr{T}]$ with the Skorohod metric (Pollard (1984)). The vector process H_n belongs to the Cartesian product space $D[\mathscr{T}]^p$, equipped with the corresponding product Skorohod topology. The weak convergence of H_n in the space $D[\mathscr{T}]^p$ is implied by the finite dimensional convergence together with stochastic equicontinuity.

THEOREM A.1: *Under Assumptions* (*A*.1), (*A*.2), (*A*.5), *and* (*A*.6), *the process H_n is stochastically equicontinuous on* ($\mathscr{T}$, ρ). *That is for any $\varepsilon > 0$, $\eta > 0$, there exists a $\delta > 0$ such that for large n*,

$$P\left(\sup_{[\delta]} \|H_n(r,y) - H_n(s,z)\| > \eta\right) < \varepsilon$$

where $[\delta] = \{(\tau_1, \tau_2);\ \tau_1 = (r,y), \tau_2 = (s,z), \rho(\tau_1,\tau_2) < \delta\}$ *with* $[\delta] \subset \mathscr{T} \times \mathscr{T}$.

When $x_t = 1$ for all t, the equicontinuity of H_n is proved by Bickel and Wichura (1971). This theorem states the stochastic equicontinuity holds for (randomly) weighted sequential process. Let $U_t = F(\varepsilon_t)$; then U_t are i.i.d. uniform on [0,1]. Define

$$(22) \qquad Y_n(s,u) = n^{-1/2} \sum_{t=1}^{[ns]} x_t\{I(U_t \le u) - u\};$$

then $H_n(s,z)=(X'X/n)^{-1/2}Y_n(s,F(z))$. By assumption, $(X'X/n)\overset{p}{\to}Q(1)$, a positive definite matrix, so Y_n and H_n are equivalent in terms of stochastic equicontinuity. Thus the proof focuses on Y_n.

LEMMA A.1: *Assume the conditions of Theorem A.1 hold. Then there exists a $K<\infty$, 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 $\alpha\le\gamma$, because $|u_2-u_1|\le 1$ and $|s_2-s_1|\le 1$. Moreover, for a constant $\tau>0$, when

$$\tau n^{-(\gamma-1)/2(\alpha-1)}\le u_2-u_1 \quad\text{and}\quad \tau n^{-(\gamma-1)/2(\alpha-1)}\le s_2-s_1 \tag{23}$$

the Lemma implies

$$E\|Y_n(s_2,u_2)-Y_n(s_1,u_2)-Y_n(s_2,u_1)+Y_n(s_1,u_1)\|^{2\gamma}$$
$$\le K[1+\tau^{-2(\alpha-1)}](u_2-u_1)^{\alpha}(s_2-s_1)^{\alpha}. \tag{24}$$

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 $\{x_t\eta_t,\mathscr{F}_t\}$ is a sequence of (not necessarily stationary and bounded) vector martingale differences, where $\mathscr{F}_t$ is the σ-field generated by $\dots,x_t,x_{t+1};\dots,U_{t-1},U_t$. By the inequality of Rosenthal (Hall and Heyde (1980, p. 23)), there exists a constant $M<\infty$ only depending on γ and p such that

$$E\|Y_n^*\|^{2\gamma}=E\left\{\left(\frac{1}{n}\left[\sum_{i<t\le j}x_t\eta_t\right]'\sum_{i<h\le j}x_h\eta_h\right)^{\gamma}\right\}$$
$$\le ME\left(\frac{1}{n}\sum_{i<t\le j}E\{(x_t'x_t)\eta_t^2|\mathscr{F}_{t-1}\}\right)^{\gamma}+Mn^{-\gamma}\sum_{i<t\le j}E\{(x_t'x_t)^{\gamma}\eta_t^{2\gamma}\}. \tag{25}$$

Note that x_t is measurable with respect to $\mathscr{F}_{t-1}$ and η_t is independent of $\mathscr{F}_{t-1}$. 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 (25). 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

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

Renaming MK as K, the Lemma follows from $(u_2-u_1)^{\gamma}\le(u_2-u_1)^{\alpha}$, for $\gamma\ge\alpha$.

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

$$\|Y_n(s,u)-Y_n(s_1,u_1)\|\le\|Y_n(s_2,u_2)-Y_n(s_1,u_1)\|+Z_nn^{1/2+\kappa}[(u_2-u_1)+(s_2-s_1)]$$

where Z_n and κ are defined in (A.5).

PROOF: First notice that all of the components of x_t can be assumed to be nonnegative. Otherwise write $x_t = \Sigma_{i=1}^p x_t^+(i) - \Sigma_{i=1}^p x_t^-(i)$ where $x_t^+(i) = (0,\dots,0,x_{ti},0,\dots,0)'$ if $x_{ti} \geq 0$ and $x_t^-(i) = (0,\dots,0,-x_{ti},0,\dots,0)'$ 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 assumption (A.5) is satisfied for $x_t^+(i)$ and $x_t^-(i)$. It is thus enough to assume that the x_t are nonnegative. In what follows, for vectors a and b, we write $a \leq b$ if $a_i \leq b_i$ 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. This implies that

$$\begin{aligned}
&Y_n(s,u) - Y_n(s_1,u_1)\\
&\quad \leq Y_n(s_2,u_2) - Y_n(s_1,u_1)\\
&\qquad + n^{1/2}\left(\frac{1}{n}\sum_{t=1}^{[ns]} x_t\right)(u_2 - u) + n^{1/2}\left(\frac{1}{n}\sum_{t=[ns]}^{[ns_2]} x_t\{I(U_t \leq u_2) - u_2\}\right)\\
&\quad \leq Y_n(s_2,u_2) - Y_n(s_1,u_1) + n^{1/2}\left(\frac{1}{n}\sum_1^n \|x_t\|\right)(u_2 - u) + n^{1/2}\left(\frac{1}{n}\sum_{t=[ns_1]}^{[ns_2]} \|x_t\|\right)
\end{aligned}$$

and

$$\begin{aligned}
Y_n(s_1,u_1) - Y_n(s,u) &\leq n^{1/2}\left(\frac{1}{n}\sum_{t=1}^{[ns]} x_t\right)(u - u_1) + n^{1/2}\left(\frac{1}{n}\sum_{t=[ns_1]}^{[ns]} x_t\{I(U_t \leq u) - u_1\}\right)\\
&\leq n^{1/2}\left(\frac{1}{n}\sum_1^n \|x_t\|\right)(u - u_1) + n^{1/2}\left(\frac{1}{n}\sum_{t=[ns_1]}^{[ns_2]} \|x_t\|\right).
\end{aligned}$$

The lemma follows from (A.5).

PROOF OF THEOREM A.1: We shall evaluate directly the modulus of continuity. 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 each $\varepsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an integer n_0, such that

$$P(\omega_\delta(Y_n) > \varepsilon) < \eta, \quad n > n_0. \tag{26}$$

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

$$P\left(\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\varepsilon\right) < 2\delta^2\eta, \quad n > n_0, \tag{27}$$

for all $(s_1,u_1) \in [0,1]^2$, where $\langle\delta\rangle = \langle\delta, s_1, u_1\rangle = \{(s,u);\ s_1 \leq s \leq s_1 + \delta,\ u_1 \leq u \leq u_1 + \delta\} \cap [0,1]^2$ (see Billingsley (1968, p. 58) for processes indexed by a single parameter).

Because of (3), for a given $\delta > 0$ and $\eta > 0$ we can choose C (to be determined later) large enough such that

$$P(|Z_n| > C) < \delta^2\eta. \tag{28}$$

Next, Lemma A.2 implies (see (22.18) of Billingsley (1968, p. 199)), when $|Z_n| \leq C$,

$$\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| \leq 3 \max_{1 \leq i,j \leq m} \|Y_n(s_1 + i\varepsilon_n, u_1 + j\varepsilon_n) - Y_n(s_1,u_1)\| + 2\varepsilon \tag{29}$$

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

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

Then

$$(30) \qquad P\Big(\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\varepsilon\Big)$$
$$\le P(|Z_n| > C) + P\Big(|Z_n| \le C, \sup_{\langle\delta\rangle}\|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\varepsilon\Big)$$
$$\le \delta^2\eta + P\Big(\max_{1\le i,j\le m} \|X(i,j)\| > \varepsilon\Big).$$

Now for fixed i and $k\,(i \ge k)$ write $Z(j) = X(i,j) - X(k,j)$. We shall use (24) to bound $Z(j)$. But (24) requires condition (23). This condition is met here because

$$(\varepsilon/C)n^{-(\gamma-1)/2(\alpha-1)} \le (\varepsilon/C)n^{-(1/2)-\kappa} = \varepsilon_n \le j\varepsilon_n, \quad j \ge 1,$$

which follows from $n^{-(\gamma-1)/2(\alpha-1)} \le n^{-(1/2)-\kappa}$ in view of (5). By (23) and (24),

$$(31) \qquad E\|Z(j) - Z(l)\|^{2\gamma} \le KC_\varepsilon[(i-k)\varepsilon_n]^\alpha[(j-l)\varepsilon_n]^\alpha, \quad 1 \le l \le j \le m,$$

where, from (24) by letting $\tau = \varepsilon/C$,

$$(32) \qquad C_\varepsilon = [1 + (C/\varepsilon)^{2(\alpha-1)}] \le 2(C/\varepsilon)^{2(\alpha-1)} \quad \text{for small } \varepsilon.$$

Thus by Theorem 12.2 of Billingsley (1968, p. 94, applied with $S_j = Z(j)$), we have

$$(33) \qquad P\Big(\max_{1\le j\le m}\|Z(j)\| > \varepsilon\Big) \le \frac{K_1KC_\varepsilon}{\varepsilon^{2\gamma}}[(i-k)\varepsilon_n]^\alpha(m\varepsilon_n)^\alpha \le \frac{K_2C_\varepsilon}{\varepsilon^{2\gamma}}[(i-k)\varepsilon_n]^\alpha\delta^\alpha$$

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

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

if we let $V(i) = \max_j\|X(i,j)\|$, then (33) implies

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

Thus by Theorem 12.2 of Billingsley once again (applied with $S_i = V(i)$), we obtain

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

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)

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

By (32), the second term on the right-hand side above is bounded by

$$(34) \qquad \frac{K_3C_\varepsilon}{\varepsilon^{2\gamma}}\delta^{2\alpha} \le \delta^2\frac{2K_3}{\varepsilon^{2(\gamma+\alpha-1)}}(C\delta)^{2(\alpha-1)}.$$

By (3), one can choose $C=(\eta/M)^{\rho}\delta^{-2/\rho}$ to assure (28) so that the right-hand side of (34) becomes $\delta^2 K(\varepsilon,\eta)\delta^a$, where $K(\varepsilon,\eta)$ is a constant and $a=(\rho-2)(\alpha-1)/\rho>0$. By choosing δ such that $K(\varepsilon,\eta)\delta^a \le \eta$, (27) follows. The proof of Theorem A.1 is completed. *Q.E.D.*

COROLLARY A.1: *Under Assumptions* (*A*.1)–(*A*.6) *with* (*A*.3) *replaced by* (*A*.3′), *the process* H_n *converges weakly to a Gaussian process H with zero mean and covariance matrix*

$$E\{H(r,y)H(s,z)'\}=Q(1)^{-1/2}Q(r\wedge s)Q(1)^{-1/2}[F(z\wedge y)-F(z)F(y)]. \tag{35}$$

PROOF: The finite dimensional convergence to a normal distribution follows from the central limit theorem for martingale differences. This, together with Theorem A.1, implies that H_n converges weakly to some Gaussian process H. To verify (35), we consider the covariance matrix function of $Y_n=n^{-1/2}(X'X)^{1/2}H_n$. For $r<s$ and $u=F(z)<v=F(y)$, using double expectation and the martingale property, we obtain

$$E\{Y_n(r,u)Y_n'(s,v)\}=\frac{1}{n}E\left(\sum_{t=1}^{[nr]}x_t x_t'\right)(u-uv) \tag{36}$$

which tends to $Q(r)(u-uv)$. From $(X'X/n)^{-1/2}\xrightarrow{p}Q(1)^{-1}$, we arrive at (35). *Q.E.D.*

COROLLARY A.2: *Under the assumptions of Corollary A*.1, *the process* V_n *defined as*

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

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

$$E\{V(r,y)V(s,z)'\}=\{A(r\wedge s)-A(r)A(s)\}\{F(y\wedge z)-F(y)F(z)\}. \tag{37}$$

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

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

Now (37) follows easily from (35). *Q.E.D.*

Note that (A.3) is a special case of (A.3′). When $Q(s)=sQ$ for some $Q>0$, the covariance matrix of V becomes $(r\wedge s-rs)\{F(z\wedge y)-F(z)F(y)\}I$, where I is the $p\times p$ identity matrix. Thus $V(\cdot,\cdot)$ has the same distribution as $B^*(\cdot,F(\cdot))$, where B^* is a vector of p independent Brownian bridges on $[0,1]^2$.

We next examine the asymptotic behavior of the sequential empirical process constructed using regression residuals. Under model (1), $\hat{\varepsilon}_t\le z$ if and only if $\varepsilon_t\le z+x_t'(\hat{\beta}-\beta)$, thus K_n^* under H_0 is given by

$$K_n^*(s,z)=(X'X)^{-1/2}\sum_{t=1}^{[ns]}x_t\{I(\varepsilon_t\le z+x_t'(\hat{\beta}-\beta))-F(z)\}. \tag{38}$$

Under the local alternative of (13), $\hat{\varepsilon}_t\le z$ if and only if

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

Thus K_n^* becomes, under H_1,

$$K_n^*(s,z)=(X'X)^{-1/2}\sum_{t=1}^{[ns]}x_t\{I(\varepsilon_t\le z(1+a_t n^{-1/2})+b_t n^{-1/2})-F(z)\} \tag{39}$$

where

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

Choosing the weights $x_t = 1$ in (39), K_n^* becomes the nonweighted sequential empirical process of residuals.

It is observed that K_n^* possesses a common form under the null and alternative hypotheses, only with different $a_t's$ and $b_t's$. This suggests the need to study a general sequential empirical process that can be specialized to various cases.

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

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

For $c_t = x_t, a_t = 0$, and $b_t = x_t' n^{1/2}(\hat{\beta}-\beta)$, $K_n(s,z,a,b)$ becomes (38). For a_t and b_t in (40), $K_n(s,z,a,b)$ becomes (39).

We impose the following conditions:

(B.1) *Assumption A.1 holds.*

(B.2) *The variable ε_t is independent of $\mathscr{F}_{t-1}$, where*

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

(B.3) *For some positive definite matrix Q,*

$$\frac{1}{n} C'C = \frac{1}{n} \sum_{t=1}^{n} c_t c_t' \overset{p}{\to} Q.$$

(B.4) $n^{-1/2} \max_{1 \leq i \leq n} |\eta_i| = o_p(1)$, *for* $\eta_i = a_i, b_i, c_i$.

(B.5) *Assumption A.5 holds for $x_t = c_t$.*

(B.6) *Assumption A.6 holds for $x_t = c_t$.*

(B.7) *There exist $\gamma > 1$ and $A < \infty$ such that for all n*

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

Obviously, (B.2) is not satisfied if $b_t = x_t' n^{1/2}(\hat{\beta}-\beta)$ because b_t depends on the entire date set through $\hat{\beta}$. It suffices to consider $b_t = x_t'\alpha$ with α varying in an arbitrary compact set. We then need some uniformity result with respect to α. This will be explained further below.

THEOREM A.2: *Under Assumptions of (B.1)–(B.7), we have:* (i)

(41) $$K_n(s,z,a,b) = K_n(s,z,0,0)$$

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

where $o_p(1)$ is uniform in s and z.

(ii) *Let* $b_t = x_t'\alpha$, *where* $\alpha \in D$, *with* D *an arbitrary compact set of* R^p. *Then* (*i*) *is still valid. Moreover, the* $o_p(1)$ *is also uniform in* $\alpha \in D$. *In particular,* (*i*) *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: Introduce an auxiliary process:

$$Z_n(s,z,a,b) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} c_t \{ I(\varepsilon_t \le z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) - F(z(1 + a_t n^{-1/2}) + b_t n^{-1/2}) \}. \tag{43}$$

The summands in this process are conditionally centered, conditional on $\mathscr{F}_{t-1}$ in view of (B.2). By adding and subtracting terms, we have

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

That the second term on the right-hand side is $o_p(1)$ follows from Theorem A.3(i) and (ii) below and Assumption (B.3). Taylor expansion of the last term gives rise to (42). *Q.E.D.*

THEOREM A.3: *Under the assumptions of Theorem A.2, we have*

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

(ii) *Let* $b_t = x_t'\alpha$ *for* α *in an arbitrary compact set of* D *of* R^P *and denote* $b(\alpha) = (x_1'\alpha, \dots, x_n'\alpha)$. *Then*

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

(iii) *Let* $a_t = r_t'\tau$; $r_t, \tau \in R^l$ *for some* $l \ge 1$; $\tau \in S$, *a compact set of* R^l. *Thus the scale parameter* a_t *is a linear function of the random vector* r_t, *a special form of heteroskedasticity. Denote* $a(\tau) = (r_1'\tau, \dots, r_n'\tau)$ *and assume* (*B*.1)–(*B*.7) *hold true when* $|a_t|$ *is replaced by* $\|r_t\|$. *Then,*

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

Part (ii) allows b_t to depend, in some way, on the entire data set, e.g., $b_t = x_t'\sqrt{n}(\hat{\beta} - \beta)$ provided $\sqrt{n}(\hat{\beta} - \beta) = O_p(1)$. This is because for any $\eta > 0$, there exists a compact set D such that $P(\sqrt{n}(\hat{\beta} - \beta) \notin D) < \eta$. Similarly, (iii) allows a scale parameter to be estimated. In our application, (i) and (ii) are sufficient. Part (iii) is not needed because no scale parameter is estimated in the model; its presence is solely for the purpose of completeness.

To prove Theorem A.3, we need the following lemma.

LEMMA A.3: *Under Assumptions* (*B*.1)–(*B*.4) *and* (*B*.7), *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 pairs of* (y,z) *such that* $|F(y) - F(z)| \le n^{-(1/2)-d}$.

PROOF: Follows from the mean value theorem; see also Lemma 1 of Bai (1991).

PROOF OF THEOREM A.3 (i): Note that (B.3) is required for Theorem A.2 but not for Theorem A.3, because Theorem A.3 obviously holds when $c_t = 0$ for all t. Further, we can assume $c_t \geq 0$ for all t as argued in the proof of Lemma A.2.

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 $F(z_i) = iN(n)^{-1}$. Because $c_t I(\varepsilon_t \leq z)$ and $c_t F(z)$ are nondecreasing in z, we have for $z_r < z < z_{r+1}$,

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

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

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

Because $\|\Sigma_{t=1}^{[ns]} \cdot\| \leq \Sigma_{t=1}^{n} \|\cdot\|$ 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 to last term is $o_p(1)$ because of Theorem A.1 (applied with $x_t = c_t$) and $N(n)^{-1} \to 0$ (the conditions of Theorem A.1 are satisfied under (B.1)–(B.7)). It remains to show

$$\max_{0 \leq r \leq N(n)} \max_{1 \leq j \leq n} \|\tilde{Z}_n(j/n, z_r)\| = o_p(1) \tag{44}$$

where $\tilde{Z}_n(j/n, z_r) := Z_n(s, z_r, a, b) - Z_n(s, z_r, 0, 0)$. But

$$P\left(\max_{0 \leq r \leq N(n)} \max_{1 \leq j \leq n} \|\tilde{Z}_n(j/n, z_r)\| > \varepsilon \right) \leq N(n) \max_r P\left(\max_j \|\tilde{Z}_n(j/n, z_r)\| > \varepsilon \right).$$

The remaining task is to bound the above probability. Let

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

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

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

By the Doob inequality,

$$P\left(\max_j \left\| n^{-1/2} \sum_{t=1}^{j} \xi_t \right\| > \varepsilon \right) \le \varepsilon^{-2\gamma} M_1 E \left\| n^{-1/2} \sum_{t=1}^{n} \xi_t \right\|^{2\gamma}, \tag{45}$$

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

$$E\left(\left\| \sum_{t=1}^{n} \xi_t \right\| \right)^{2\gamma} \le M_2 E\left\{ \sum_{t=1}^{n} E(\|\xi_t\|^2 | \mathscr{F}_{t-1}) \right\}^{\gamma} + M_2 \sum_{t=1}^{n} E\|\xi_t\|^{2\gamma} \tag{46}$$

for all n. Because (a_i, b_t, c_t) is measurable with respect to $\mathscr{F}_{i-1}$ and ε_i is independent of $\mathscr{F}_{i-1}$ by (B.2),

$$\begin{aligned} & E(\|\xi_i\|^2 | \mathscr{F}_{i-1}) \\ & \quad \le \|c_i\|^2 \{F(z_r(1 + a_i n^{-1/2}) + b_t n^{-1/2}) - F(z_r)\} \\ & \quad \le \frac{1}{\sqrt{n}} \|c_t\|^2 L(|a_i| + |b_t|) \end{aligned} \tag{47}$$

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

$$E\|\xi_i\|^{2\gamma} \le n^{-\gamma/2} L^{\gamma} E\{\|c_i\|^2 (|a_i| + |b_i|)\}^{\gamma}. \tag{48}$$

Combining (46), (47), and (48), we have for $M_3 = M_2 L^{\gamma}$,

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

The last inequality follows from Assumption (B.7). The above bound does not depend on z_r. Thus for $M_4 = 2 M_1 M_3 A$,

$$P\left(\max_r \max_j |\tilde{Z}_n(j/n, z_r)| > \varepsilon \right) \le \varepsilon^{-2\gamma} M_4 N(n) n^{-\gamma/2} = \varepsilon^{-2\gamma} M_4 n^{-[(\gamma - 1)/2] + d}$$

because $N(n) = n^{(1/2)+d}$. The above converges to zero if we choose $d \in (0, (\gamma - 1)/2)$ in Lemma A.3. The proof of (i) is completed. *Q.E.D.*

PROOF OF THEOREM A.3 (ii): This really follows from the compactness of D. We use similar arguments as in Koul (1991). Since D is compact, for any $\delta > 0$, the set D can be partitioned into a finite number of subsets such that the diameter of each subset is not greater than δ. Denote these subsets by $D_1, D_2, \ldots, D_{m(\delta)}$. Fix k and consider D_k. Pick $\alpha_k \in D_k$. For all $\alpha \in D_k$

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

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

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

and a reversed inequality holds when δ is replaced by $-\delta$. Using the mean value theorem and Assumption (B.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

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

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). *Q.E.D.*

PROOF OF THEOREM A.3 (iii): This 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 completed. *Q.E.D.*

We are now in the position to prove Theorems 1 through 5.

PROOF OF THEOREM 1 AND THEOREM 4: Under the null hypothesis, $\hat{\varepsilon}_t = \varepsilon_t - x_t'(\hat{\beta} - \beta)$ so $\hat{\varepsilon}_t \leq z$ if and only if $\varepsilon_t \leq z + x_t'(\hat{\beta} - \beta)$. By applying Theorem A.2 with $a_t = 0$, $b_t = x_t'\sqrt{n}(\hat{\beta} - \beta)$, and $c_t = x_t$ (under these choices, $K_n(s,z,a,b)$ becomes $K_n^*(s,z)$ given in (38)), we have

$$(49) \qquad K_n^*(s,z) - A_{[ns]}K_n^*(1,z)$$
$$= H_n(s,z) - A_{[ns]}H_n(1,z)$$
$$(50) \qquad + f(z)(X'X/n)^{-1/2}\frac{1}{n}\sum_{t=1}^{[ns]} x_t b_t - f(z)A_{[ns]}(X'X/n)^{-1/2}\frac{1}{n}\sum_{t=1}^{n} x_t b_t$$
$$(51) \qquad + o_p(1).$$

Expression (50) 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 4 now follows from Corollary A.2. Theorem 1 (ii) follows as a special case because $Q(s) = sQ$ and $A(s) = sI$ in the absence of trending regressors. To prove Theorem 1(i), take $x_t = 1$ and $A_{[ns]} = [ns]/n$ in the above proof, reducing (50) to

$$(52) \qquad f(z)\frac{1}{n}\sum_{t=1}^{[ns]} b_t - f(z)\frac{[ns]}{n}\sum_{t=1}^{n} b_t = f(z)\left(\frac{1}{n}\sum_{t=1}^{[ns]} x_t - \frac{[ns]}{n}\sum_{t=1}^{n} x_t\right)\sqrt{n}(\hat{\beta} - \beta),$$

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. *Q.E.D.*

We note that if $x_t = t/n$, a linear time trend, then (52) cannot be $o_p(1)$ because the two terms in the larger parentheses do not have the same limit. This implies the test M_n will not be asymptotically distribution free with the presence of a trending regressor.

PROOF OF THEOREM 2: Under local alternatives (13), $K_n(s, z, a, b)$ becomes (39) with a_t and b_t given by (40). Under these local alternatives, β is still estimable with root-n consistency. 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. We can thus assume $b_t = \Delta_1 x_t' g(t/n)$. Now by Theorem A.2, for $a_t = \Delta_2 h(t/n)$ and $c_t = x_t$,

$$(53)\qquad K_n^*(s, z) - A_{[ns]}K_n^*(1, z)$$

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

$$(54)\qquad + f(z) z \Delta_2 \left\{ \left(\frac{X'X}{n} \right)^{-1/2} \frac{1}{n} \sum_{t=1}^{[ns]} x_t h(t/n) - A_{[ns]} \left(\frac{X'X}{n} \right)^{-1/2} \frac{1}{n} \sum_{t=1}^{n} x_t h(t/n) \right\}$$

$$(55)\qquad + f(z) \Delta_1 \left\{ \left(\frac{X'X}{n} \right)^{-1/2} \frac{1}{n} \sum_{t=1}^{[ns]} x_t x_t' g(t/n) - A_{[ns]} \left(\frac{X'X}{n} \right)^{-1/2} \frac{1}{n} \sum_{t=1}^{n} x_t x_t' g(t/n) \right\}$$

$$+ o_p(1).$$

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

$$(56)\qquad \frac{1}{n} \sum_{t=1}^{[ns]} x_t h(t/n) \xrightarrow{p} \bar{x} \int_0^s h(v)\, dv,$$

$$(57)\qquad \frac{1}{n} \sum_{t=1}^{[ns]} x_t x_t' g(t/n) \xrightarrow{p} Q \int_0^s g(v)\, dv.$$

Furthermore, under Assumption (A.3),

$$(58)\qquad A_{[ns]}(X'X/n)^{-1/2} \xrightarrow{p} sQ^{-1/2}.$$

From these results, (54) converges to $f(z) z \Delta_2 Q^{-1/2} \bar{x} \lambda_h(s)$ where λ_h is given by (15); and similarly, (55) converges to $f(z) \Delta_1 Q^{1/2} \lambda_g(s)$ where λ_g is given by (14). Thus (17) is obtained and (16) is obtained similarly by choosing $c_t = 1$. The proof is completed. *Q.E.D.*

PROOF OF THEOREM 3: The proof makes use of Theorem A.3. The details are omitted here and the complete proof is available upon request from the author.

PROOF OF THEOREM 5: The proof is virtually identical to that of Theorem 2, except under (A.3′), (56)–(58) are replaced by

$$\frac{1}{n} \sum_{t=1}^{[ns]} x_t h(t/n) \xrightarrow{p} \int_0^s Q^{(1)}(v) e h(v)\, dv,$$

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

$$A_{[ns]}(X'X/n)^{-1/2} \xrightarrow{p} Q(1)^{-1/2} Q(s) Q(1)^{-1},$$

respectively, where $Q^{(1)}(v)$ is the derivative of $Q(v)$ and $e = (1, 0, \ldots, 0)'$. The first convergence is a special case of the second due to the presence of a constant regressor. *Q.E.D.*

REFERENCES

Alogoskoufis, G., and R. Smith (1991): "The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange Rate Regimes," *American Economic Review*, 81, 1254–1275.

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 W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383–1414.

Bai, J. (1991): "Weak Convergence of Sequential Empirical Processes of ARMA Residuals," *Annals of Statistics*, 22, 2051–2061.

Bickel, P. J., and M. J. Wichura (1971): "Convergences for Multiparameter Stochastic Processes and Some Applications," *Annals of Mathematical Statistics*, 42, 1656–1670.

Billingsley, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

Boldin, M. V. (1982): "Estimation of the Distribution of Noise in an Autoregression Scheme," *Theory of Probability and its Applications*, 27, 866–871.

——— (1989): "On Testing Hypotheses in Sliding Average Scheme by the Kolmogorov-Smirnov and ω^2 Tests," *Theory of Probability and its Applications*, 34, 699–704.

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.

Carlstein, E. (1988): "Nonparametric Change Point Estimation," *Annals of Statistics*, 16, 188–197.

Chow, G. C. (1960): "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605.

Chu, C-S. J., and H. White (1992): "A Direct Test for Changing Trend," *Journal of Business and Economic Statistics*, 10, 289–300.

Csörgő, M., and L. Horváth (1987): "Nonparametric Tests for the Change-Point Problem," *Journal of Statistical Planning and Inference*, 17, 1–9.

——— (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, pp. 403–425.

Deshayes, J., and D. Picard (1986): "Off-line Statistical Analysis of Change Point Models using Non Parametric and Likelihood Methods," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, Lecture Notes in Control and Information Sciences, No. 77, ed. by M. Basseville and A. Benveniste. Berlin: Springer-Verlag, pp. 103–168.

Dümbgen, L. (1991): "The Asymptotic Behavior of Some Nonparametric Change Point Estimators," *Annals of Statistics*, 19, 1471–1495.

Durbin, J. (1973): "Weak Convergence of the Sample Distribution Function when Parameters are Estimated," *Annals of Statistics*, 1, 279–290.

Durbin, J. (1976): "Kolmogorov-Smirnov Tests when Parameters are Estimated," in *Empirical Distributions and Processes*, Lecture Notes in Mathematics. Berlin: Springer-Verlag.

Gombay, E., and L. Horváth (1994): "Limit Theorem for Change in Linear Regression," *Journal of Multivariate Analysis*, 48, 43–69.

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

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

Hawkins, D. L. (1988): "Retrospective and Sequential Tests for Shifts in Linear Models with Time Series Regressors and Errors," *ASA Proceedings of the Business and Economic Statistics Section*, pp. 516–521.

——— (1989): "A U-L Approach to Retrospective Testing for Shifting Parameters in a Linear Model," *Communications in Statistics: Theory and Methods*, 18, 3117–3134.

Huskova, M. (1991): "Recursive M-tests for the Change-Point Problem," in *Economic Structural Change: Analysis and Forecasting*, ed. by P. Hackel and A. H. Westlund. New York: Springer-Verlag, pp. 13–33.

JANDHYALA, V. K. (1993): "A Property of Partial Sums of Regression Least Squares Residuals and its Applications," *Journal of Statistical Planning and Inference*, 37, 317.

KIM, H. J., AND D. SIEGMUND (1989): "The Likelihood Ratio Test for a Change Point in Simple Linear Regression," *Biometrika*, 76, 409–423.

KOUL, H. L. (1984): "Test of Goodness-of-fit in Linear Regression," *Colloaquia Mathematica Societatis Janos Bolyai.*, 45, *Goodness of Fit*. Debrecen, Hungary, pp. 279–315.

——— (1991): "A Weak Convergence Result useful in Robust Autoregression," *Journal of Statistical Planning and Inference*, 29, 1291–1308.

——— (1992): *Weighted Empiricals and Linear Models*, IMS Lecture Notes-Monograph Series. Hayward, CA: IMS.

KRAMER, W., W. PLOBERGER, AND R. ALT (1988): "Testing for Structural Changes in Dynamic Models," *Econometrica*, 56, 1355–1370.

KREISS, P. (1991): "Estimation of the Distribution of Noise in Stationary Processes," *Metrika*, 38, 285–297.

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.

PERRON, P., AND T. S. VOGELSANG (1992): "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10, 301–302.

PLOBERGER, W., AND W. KRAMER (1990): "The Local Power of the CUSUM and CUSUM of Squares Tests," *Econometric Theory*, 6, 335–347.

——— (1992): "The CUSUM Test with OLS Residuals," *Econometrica*, 60, 271–285.

PLOBERGER, W., W. KRAMER, AND K. KONTRUS (1988): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

QUANDT, R. E. (1960): "Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330.

ROBINSON, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl. Berlin: Springer-Verlag, pp. 253–264.

——— (1991): "Time Varying Nonlinear Regression, " in *Economic Structural Change: Analysis and Forecasting*, ed. by P. Hackel and A. H. Westlund. New York: Springer-Verlag, pp. 179–190.

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.

SZYSZKOWICZ, B. (1994): "Weak Convergence of Weighted Empirical Type Processes under Contagious and Changepoint Alternatives," *Stochastic Processes, Appl.*, 50, 281–313.

ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–271.