A NONPARAMETRIC TEST FOR SERIAL INDEPENDENCE OF REGRESSION ERRORS

Miguel A. Delgado and Juan Mora

WP-AD 99-28

Correspondence to J. Mora: University of Alicante. Departamento de Fundamentos del Análisis Económcio. Ctra. San Vte. del Raspeig, s/n. 03071 Alicante, Spain. Tel.: 965903400 Ext: 3258. E-mail: juan@merlin.fae.ua.es

Editor: Instituto Valenciano de Investigaciones Económicas, s.a.

First Edition Diciembre 1999

Depósito Legal: V–5424-1999

---

* This research was partially supported by Spanish Dirección General de Enseñanza Superior and Instituto Valenciano de Investigaciones Económicas. We thank the referees and an associate editor for helpful comments.

** Miguel A. Delgado: University Carlos III (Madrid), Juan Mora: University of Alicante.

A NONPARAMETRIC TEST FOR SERIAL INDEPENDENCE OF REGRESSION ERRORS

Miguel A. Delgado and Juan Mora

A B S T R A C T

A test for serial independence of regression errors is proposed that is consistent in the direction of serial dependence alternatives of first order. The test statistic is a function of a Hoeffding-Blum-Kiefer-Rosenblatt type of empirical process, based on residuals. The resultant statistic converges, surprisingly, to the same limiting distribution as the corresponding statistic based on true errors.

KEYWORDS: Empirical process based on residuals; Hoeffding-Blum-Kiefer-Rosenblatt statistic; Serial independence test.

1. PRELIMINARIES AND STATEMENT OF THE PROBLEM

Consider a strictly stationary discrete time process $\{U_i, i \geq 1\}$. Let $F(\cdot)$ be the distribution function of $(U_i, U_{i+1})'$ and $F_1(\cdot)$ the marginal distribution function of U_i. Define $S(u) = F(u) - F_1(u_1)F_1(u_2)$, for $u = (u_1, u_2)' \in \mathbb{R}^2$. Given observations $\{U_i\}_{i=1}^{n+1}$, Skaug & Tjøstheim (1993), Delgado (1996) and Hong (1998), among others, have proposed to test

$$H_0 : \{U_i, i \geq 1\} \text{ are independently distributed,}$$

$$H_1 : \ S(u) \neq 0, \text{ for some } u \in \mathbb{R}^2,$$

using statistics which are functionals of $n^{1/2}S_n(\cdot)$, where $S_n(\cdot)$ is the Hoeffding-Blum-Kiefer-Rosenblatt process (Delgado, 1999), defined by

$$S_n(u) = F_n(u) - F_{n1}(u_1)F_{n1}(u_2),$$

where $F_n(u) = n^{-1}\sum_{i=1}^{n} 1(U_i \leq u_1)1(U_{i+1} \leq u_2)$, $1(\cdot)$ is the indicator function and $F_{1n}(\cdot)$ is the univariate empirical distribution function based on $\{U_i\}_{i=1}^{n+1}$. A popular test statistic for H_0 which is based on $n^{1/2}S_n(\cdot)$ is the Cramér-von Mises statistic

$$C_n = n^{-1}\sum_{i=1}^{n}\{n^{1/2}S_n(U_i, U_{i+1})\}^2.$$

Hoeffding (1948) and Blum, Kiefer & Rosenblatt (1961) proposed this type of statistic for testing independence between two samples, and tabulated its limiting distribution under the null hypothesis. Skaug & Tjøstheim (1993) showed that, if $F(\cdot)$ is continuous, C_n and the statistic of Blum et al. (1961) have the same limiting distribution. Delgado (1996) showed that this is not the case when higher-order dependence alternatives are considered. Other functionals of $n^{1/2}S_n(\cdot)$ could be used, e.g. based on the supremum distance, as in the case of Kolmogorov-Smirnov statistics.

Suppose now that $\{U_i, i \geq 1\}$ are unobservable errors in the linear regression model $Y_i = X_i'\beta_0 + U_i$, where X_i are fixed regressors and β_0 is a k-dimensional vector of unknown parameters. In this case, we propose to test H_0 as before, replacing the unobservable errors U_i by residuals $\hat{U}_{ni} = Y_i - X_i'\hat{\beta}_n$, where $\hat{\beta}_n$ is a suitable estimate of β_0. Thus, $S(u)$ is estimated by

$$\hat{S}_n(u) = \hat{F}_n(u) - \hat{F}_{n1}(u_1)\hat{F}_{n1}(u_2),$$

where $\hat{F}_n(\cdot)$ and $\hat{F}_{n1}(\cdot)$ are defined as $F_n(\cdot)$ and $F_{n1}(\cdot)$, but replacing U_i by $\hat{U}_{ni}$. Functionals of $n^{1/2}\hat{S}_n(\cdot)$ can be used as test statistics, e.g. the Cramér-von Mises statistic

$$\hat{C}_n = n^{-1}\sum_{i=1}^{n}\{n^{1/2}\hat{S}_n(\hat{U}_{ni}, \hat{U}_{n,i+1})\}^2.$$

In view of the existing results on empirical processes depending on parameter estimates, see e.g. Durbin (1973) for a discussion of this problem in the context of goodness-of-fit tests, we would expect a different asymptotic behaviour for $n^{1/2}S_n(\cdot)$ and $n^{1/2}\hat{S}_n(\cdot)$. Surprisingly, we prove in § 2 that $n^{1/2}S_n(\cdot)$ and $n^{1/2}\hat{S}_n(\cdot)$ have the same limiting distribution, and hence $\hat{C}_n$ can be used to test H_0 in the same way as C_n. The results of a Monte Carlo experiment are reported in § 3. Proofs are confined to an Appendix.

2. ASYMPTOTIC PROPERTIES

The following assumptions must hold under both H_0 and H_1.

Assumption 1: $Y_i = X_i'\beta_0 + U_i$, and $\{U_i,\ i \geq 1\}$ is a strictly stationary discrete time process.

Assumption 2: $\sum_{i=1}^n X_i X_i'$ is a non-random and non-singular matrix such that

$$\max_{1\leq i\leq n} X_i' (\sum_{i=1}^n X_i X_i')^{-1} X_i = o\,(1)\,.$$

Assumption 3: The distribution function of $(U_i, U_{i+1})'$ has a density function with marginal density function $f_1(\cdot)$ uniformly continuous and such that $f_1(x) > 0$ for all $x \in \mathbb{R}$.

Assumption 4: $\hat{\beta}_n$ is an estimator of β_0 such that

$$(\sum_{i=1}^n X_i X_i')^{1/2}(\hat{\beta}_n - \beta_0) = O_p\,(1)\,.$$

Assumption 2 is typical when studying asymptotic properties of statistics in this context; this assumption does not rule out trending regressors. Under Assumption 3, which is necessary to ensure that empirical processes based on residuals behave properly (Koul, 1992, pp. 36-9), the marginal distribution function is strictly increasing. If Assumption 2 holds, Assumption 4 is satisfied by most estimates, such as ordinary least squares.

The following theorem establishes the asymptotic equivalence between $\hat{S}_n(\cdot)$ and $S_n(\cdot)$.

THEOREM 1: *If Assumptions 1-4 hold, then*

(a) *under* H_0, $\sup_{u\in\mathbb{R}^2}\left|\hat{S}_n(u) - S_n(u)\right| = o_p(n^{-1/2})$,

(b) *under* H_1, *if* $\{U_i, i \geq 1\}$ *is ergodic, then* $\sup_{u\in\mathbb{R}^2}\left|\hat{S}_n(u) - S_n(u)\right| = o_p(1)$.

It follows from Theorem 1, see the proof of the Corollary in the Appendix, that, under H_0, $n^{1/2}\hat{S}_n(\cdot)$ and $n^{1/2}S_n(\cdot)$ converge weakly to the same process, which is, as Skaug & Tjøstheim (1993) prove, a Gaussian process, $S_\infty(\cdot)$ say, with $E\{S_\infty(u)\} = 0$ and $\mathrm{cov}\{S_\infty(u), S_\infty(v)\} = \prod_{j=1}^{2}[\min\{F_1(u_j), F_1(v_j)\} - F_1(u_j)F_1(v_j)]$; and, under H_1, $\hat{S}_n(\cdot)$ and $S_n(\cdot)$ converge in probability to $S(\cdot)$. These results are exploited in the following corollary, which justifies asymptotic inferences based on $\hat{C}_n$.

COROLLARY: *If Assumptions 1-4 hold, then*

(a) *under* H_0, $\hat{C}_n$ *converges in distribution to* $C_\infty = \int_{\mathbb{R}^2} S_\infty(u)^2 dF(u)$,

(b) *under* H_1, *if* $\{U_i, i \geq 1\}$ *is ergodic, then, for all* $c < \infty$, $\lim_{n\to\infty}\mathrm{pr}\{\hat{C}_n > c\} = 1$.

The distribution of C_∞ does not depend on $F(\cdot)$ and has been tabulated by Blum et al. (1961). The Corollary states that, asymptotically, the test can be performed using $\hat{C}_n$ and critical values from the distribution of C_∞, i.e. in the same way as if we used C_n. This result may seem surprising at first sight because, in goodness-of-fit tests, the statistic computed with errors and the statistic computed with residuals have different asymptotic distributions; see e.g. Koul (1992, pp. 178-86). When testing goodness of fit, replacing the true parameter value by an estimator introduces a non-negligible random term in the empirical distribution function, and this affects the limiting distribution of the test statistic. When testing independence, replacing β_0 by $\hat{\beta}_n$ introduces random terms in the joint empirical distribution function and in the two marginal empirical distribution functions, but these random terms cancel out asymptotically when we consider the Hoeffding-Blum-Kiefer-Rosenblatt process.

In a nonlinear regression model $Y_i = m(X_i, \beta_0) + U_i$, where $m(\cdot)$ is a known function, continuously differentiable in a neighbourhood of β_0, the equivalence result we establish is also expected to hold if we assume, instead of Assumptions 2 and 4, that the estimator $\hat{\beta}_n$ is such that $\max_{1 \leq i \leq n}\{\dot{m}(X_i, \bar{\beta}_n)' R_n(\bar{\beta}_n)^{-1} \dot{m}(X_i, \bar{\beta}_n)\} = o_p(1)$ and $R_n(\bar{\beta}_n)^{1/2}(\hat{\beta}_n - \beta_0) = O_p(1)$, for any $\bar{\beta}_n$ such that $\|\bar{\beta}_n - \beta_0\| \leq \|\hat{\beta}_n - \beta_0\|$, where $\dot{m}(x, \beta) = \partial m(x, \beta)/\partial \beta$ and $R_n(\beta) = \sum_{i=1}^n \dot{m}(X_i, \beta)\dot{m}(X_i, \beta)'$. However, the reasoning which we use to prove Theorem 1 does not apply directly in the nonlinear case because it is based on results derived in Koul (1992, Ch.3), where only linear models are considered.

3. SIMULATIONS

In order to study how the replacement of errors by residuals affects the finite sample behaviour of the test statistic, we carried out some Monte Carlo experiments with programs written in GAUSS. We generated $n + 1$ observations from a linear regression model with $X_i = (1, i)'$, $\beta_0 = (1, 1)'$ and errors U_i satisfying a first-order autoregressive model $U_i = \rho U_{i-1} + \varepsilon_i$, where ε_i are independent identically distributed $N(0, 1)$ variables; hence H_0 is true if and only if $\rho = 0$. We used least squares residuals to compute the test statistic $\hat{C}_n$. In Table 1, we report the proportion of rejections of H_0 in 5000 Monte Carlo samples for different parameter values ρ, significance levels α and sample sizes n. The critical values we used, 0.04694 for $\alpha = 0.1$, 0.0584 for $\alpha = 0.05$ and 0.08685 for $\alpha = 0.01$, were obtained from Table II in Blum et al. (1961).

TABLE 1: Proportion of rejections of $H_0 : \rho = 0$ from sets of 5000 Monte Carlo samples, using the statistics C_n and $\hat{C}_n$.

n	ρ	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		C_n	$\hat{C}_n$	C_n	$\hat{C}_n$	C_n	$\hat{C}_n$
50	-0.6	0.975	0.978	0.954	0.961	0.882	0.897
	-0.4	0.753	0.771	0.650	0.676	0.424	0.454
	-0.2	0.289	0.317	0.189	0.213	0.070	0.079
	0	0.111	0.110	0.059	0.056	0.015	0.014
	0.2	0.397	0.332	0.278	0.234	0.116	0.093
	0.4	0.829	0.776	0.746	0.685	0.534	0.453
	0.6	0.981	0.964	0.966	0.943	0.908	0.860
		C_n	$\hat{C}_n$	C_n	$\hat{C}_n$	C_n	$\hat{C}_n$
250	-0.6	1.000	1.000	1.000	1.000	1.000	1.000
	-0.4	1.000	1.000	1.000	1.000	0.999	0.999
	-0.2	0.868	0.878	0.786	0.799	0.581	0.597
	0	0.105	0.105	0.057	0.057	0.011	0.010
	0.2	0.893	0.880	0.829	0.811	0.637	0.610
	0.4	1.000	1.000	1.000	1.000	0.999	0.999
	0.6	1.000	1.000	1.000	1.000	1.000	1.000

We observe that C_n and $\hat{C}_n$ yield very similar results. Moreover, the empirical level of the test is fairly close to the theoretical level and the power is reasonably high. To study the power of the test in other contexts, we performed some other Monte Carlo experiments with the same characteristics as those described in Skaug & Tjøstheim (1993, § 4.4). The results of these experiments are not reported here; we obtained the same results as Skaug & Tjøstheim (1993), both when using errors and when using residuals.

APPENDIX

Proofs

Detailed proofs are available from the authors on request. Hereafter, the interval $[0,1]$ is denoted by I, $I^2 \equiv I \times I$, $\mathbb{D}(I^2)$ is the set of all real functions on I^2 which are 'continuous from above with limits from below' as in Neuhaus (1971), $\mathbb{C}(I^2)$ is the set of all real continuous functions on I^2, '$\Rightarrow$' denotes weak convergence, $t = (t_1, t_2)'$ is a generic element in I^2, $j = 1, 2$ and $i = 1, ..., n$, unless otherwise stated. The proofs of Theorem 1 and the Corollary will be derived from the following proposition.

PROPOSITION A1: *Let $\{(Y_{1i}, X_{1i}', Y_{2i}, X_{2i}')'\}_{i=1}^n$ be observations from an $\mathbb{R} \times \mathbb{R}^{p_1} \times \mathbb{R} \times \mathbb{R}^{p_2}$-valued variable such that the following linear regression models hold: $Y_{ji} = X_{ji}'\beta_{j0} + U_{ji}$, where $\{(U_{1i}, U_{2i})', i \geq 1\}$ is a strictly stationary sequence of random vectors. We assume that both regression models satisfy Assumption 2, that we have estimators $\hat{\beta}_{nj}$ satisfying Assumption 4 and that the distribution function of $(U_{1i}, U_{2i})'$ has a density function with marginal density functions uniformly continuous and positive in $\mathbb{R}$. Let $H(\cdot)$ be the distribution function of $(U_{1i}, U_{2i})'$ and $H_j(\cdot)$ its marginal distribution functions. Define $P_n(t) = n^{1/2}(n^{-1}\sum_{i=1}^n[\prod_{j=1}^2 1\{H_i(U_{ji}) \leq t_j\}] - n^{-2}\prod_{j=1}^2[\sum_{i=1}^n 1\{H_i(U_{ji}) \leq t_j\}])$ and $\hat{P}_n(t)$ in the same way as $P_n(t)$, but replacing errors U_{ji} by residuals $\hat{U}_{nji} = Y_{ji} - X_{ji}'\hat{\beta}_{nj}$.*

(a) *If $\{(U_{1i}, U_{2i})', i \geq 1\}$ is an ergodic sequence, then* $\sup_{t \in I^2}\left|\hat{P}_n(t) - P_n(t)\right| =$ *$o_p(n^{1/2})$. Moreover, $n^{-1/2}\hat{P}_n(\cdot)$ converges in probability to $L(t) = G(t) - t_1 t_2$, where $G(t) = H\{H_1^{-1}(t_1), H_2^{-1}(t_2)\}$.*

(b) *If $\{(U_{1i}, U_{2i})', i \geq 1\}$ is an $m-$dependent sequence for $m \in \mathbb{N} \cup \{0\}$*

(Billingsley 1968, p. 167), and $H(u) = H_1(u_1)H_2(u_2)$ for all $u = (u_1, u_2)' \in \mathbb{R}^2$, *then* $\sup_{t \in I^2} \left| \hat{P}_n(t) - P_n(t) \right| = o_p(1)$. *Moreover,* $\hat{P}_n(\cdot) \Rightarrow P^{(m)}(\cdot)$, *where* $P^{(m)}(\cdot)$ *is a Gaussian process in* $\mathbb{D}(I^2)$ *with zero mean and*

$$\operatorname{cov}\{P^{(m)}(s), P^{(m)}(t)\} = \prod_{j=1}^{2}\{\min(s_j, t_j) - s_j t_j\} +$$

$$\sum_{k=1}^{m} E(\prod_{j=1}^{2}[1\{H_j(U_{j1}) \leq s_j\} - s_j][1\{H_j(U_{j,k+1}) \leq t_j\} - t_j]) +$$

$$\sum_{k=1}^{m} E(\prod_{j=1}^{2}[1\{H_j(U_{j,k+1}) \leq s_j\} - s_j][1\{H_j(U_{j1}) \leq t_j\} - t_j]),$$

where the last two terms on the right-hand side appear only if $m > 0$.

(c) *Let* $D : \mathbb{R} \to \mathbb{R}$ *be a continuous function and* $Q_n(\cdot)$, $Q(\cdot)$ *processes in* $\mathbb{D}(I^2)$ *such that* $\operatorname{pr}\{Q(\cdot) \in \mathbb{C}(I^2)\} = 1$. *If* $\{(U_{1i}, U_{2i})', i \geq 1\}$ *is an ergodic sequence, then the random variable* $n^{-1}\sum_{i=1}^{n} D[Q_n\{H_1(\hat{U}_{n1i}), H_2(\hat{U}_{n2i})\}]$ *converges in distribution to* $\int_{I^2} D\{Q(t)\}dG(t)$.

Proof:

(a) Define $W_n(t) = n^{-1/2}\sum_{i=1}^{n}[1\{H_1(U_{1i}) \leq t_1\}1\{H_2(U_{2i}) \leq t_2\} - G(t)]$, $W_{jn}(t_j) = n^{-1/2}\sum_{i=1}^{n}[1\{H_j(U_{ji}) \leq t_j\} - t_j]$ and $\hat{W}_n(t)$, $\hat{W}_{jn}(t)$ in the same way as $W_n(t)$, $W_{jn}(t)$, but replacing U_{ji} by $\hat{U}_{nji}$. Then

$$\hat{P}_n(t) = \hat{W}_n(t) - t_2\hat{W}_{1n}(t_1) - t_1\hat{W}_{2n}(t_2) - n^{-1/2}\hat{W}_{1n}(t_1)\hat{W}_{2n}(t_2) + n^{1/2}L(t), \quad \text{(A1)}$$

$$P_n(t) = W_n(t) - t_2W_{1n}(t_1) - t_1W_{2n}(t_2) - n^{-1/2}W_{1n}(t_1)W_{2n}(t_2) + n^{1/2}L(t). \quad \text{(A2)}$$

Define $g_j(t_j) = h_j\{H_j^{-1}(t_j)\}$, $\hat{t}_{jni} = H_j\{H_j^{-1}(t_j) + X'_{ji}(\hat{\beta}_{nj} - \beta_{j0})\}$ and $\hat{t}_{ni} = H\{H_1^{-1}(t_1) + X'_{1i}(\hat{\beta}_{n1} - \beta_{10}), H_2^{-1}(t_2) + X'_{2i}(\hat{\beta}_{n2} - \beta_{20})\}$. As $H_j(\cdot)$ is a one-to-one mapping, $1\{H_j(\hat{U}_{nji}) \leq t_j\} = 1\{H_j(U_{ji}) \leq \hat{t}_{jni}\}$. Hence, if we define

$E_{jn}(t_j) = n^{-1/2}\sum_{i=1}^{n}[1\{H_j(U_{ji}) \le \hat{t}_{jni}\} - \hat{t}_{jni} - 1\{H_j(U_{ji}) \le t_j\} + t_j]$,

$Z_{jn}(t_j) = n^{-1/2}\sum_{i=1}^{n}(\hat{t}_{jni} - t_j) - n^{-1/2}g_j(t_j)\sum_{i=1}^{n} X'_{ji}(\hat{\beta}_{nj} - \beta_{j0})$,

$B_{jn}(t_j) = n^{-1/2}g_j(t_j)\sum_{i=1}^{n} X'_{ji}(\hat{\beta}_{nj} - \beta_{j0})$,

$E_n(t) = n^{-1/2}\sum_{i=1}^{n}(\prod_{j=1}^{2}[1\{H_j(U_{ji}) \le \hat{t}_{jni}\}] - \hat{t}_{ni} - \prod_{j=1}^{2}[1\{H_j(U_{ji}) \le t_j\}] + G(t))$,

$Z_n(t) = n^{-1/2}\sum_{i=1}^{n}\{\hat{t}_{ni} - G(t)\} - t_2B_{1n}(t_1) - t_1B_{2n}(t_2)$,

then it is easily proved that

$$\hat{W}_{jn}(t_j) = E_{jn}(t_j) + Z_{jn}(t_j) + B_{jn}(t_j) + W_{jn}(t_j), \tag{A3}$$

$$\hat{W}_n(t) = E_n(t) + Z_n(t) + t_1B_{2n}(t_2) + t_2B_{1n}(t_1) + W_n(t). \tag{A4}$$

With our assumptions, and using similar arguments as in Koul (1992, pp. 28-39), it may be proved that $\sup_{t\in I}|Z_{jn}(t)| = o_p(1)$, $\sup_{t\in I^2}\left|n^{-1/2}Z_n(t)\right| = o_p(1)$, $\sup_{t\in I}\left|n^{-1/2}E_{jn}(t)\right| = o_p(1)$, $\sup_{t\in I^2}\left|n^{-1/2}E_n(t)\right| = o_p(1)$, $\sup_{t\in I}|B_{jn}(t)| = O_p(1)$, $\sup_{t\in I}\left|n^{-1/2}W_{jn}(t)\right| = o_p(1)$. In view of (A1)-(A4), all these results imply that $\sup_{t\in I^2} n^{-1/2}\left|\hat{P}_n(t) - P_n(t)\right| = o_p(1)$. On the other hand, $n^{-1/2}P_n(t) - L(t) = n^{-1}\sum_{i=1}^{n}[\prod_{j=1}^{2} 1\{H_j(U_{ji}) \le t_j\} - G(t)] - n^{-1/2}\{t_2W_{1n}(t_1) + t_1W_{2n}(t_2) + n^{-1/2}W_{1n}(t_1)W_{2n}(t_2)\}$. If we use the Glivenko-Cantelli Theorem in Stute & Schumann (1980) and Theorem 4.1 in Billingsley (1968, p. 25), it follows that $n^{-1/2}\hat{P}_n(t)$ converges in probability to $L(t)$.

(b) With these assumptions, $\sup_{t\in I}|Z_{jn}(t)| = o_p(1)$, $\sup_{t\in I^2}|Z_n(t)| = o_p(1)$, $\sup_{t\in I}|E_{jn}(t)| = o_p(1)$, $\sup_{t\in I^2}|E_n(t)| = o_p(1)$, $\sup_{t\in I}|B_{jn}(t)| = O_p(1)$, $\sup_{t\in I}|W_{jn}(t)| = O_p(1)$. Thus from (A1)-(A4) it follows that $\sup_{t\in I^2}\left|\hat{P}_n(t) - P_n(t)\right| = o_p(1)$. Moreover, write $V_n(t) = W_n(t) - t_2W_{1n}(t_1) - t_1W_{2n}(t_2)$. From (A2) it follows that $P_n(t) = V_n(t) - n^{-1/2}W_{1n}(t_1)W_{2n}(t_2)$; if we use Theorem 4 in Csörgö (1979), $V_n(\cdot) \Rightarrow P^{(m)}(\cdot)$ and hence $P_n(\cdot) \Rightarrow P^{(m)}(\cdot)$.

(c) Write $\hat{G}_n(t) = n^{-1}\sum_{i=1}^{n}\prod_{j=1}^{2}[1\{H_j(\hat{U}_{nji}) \leq t_j\}]$, and define $G_n(t)$ in the same way as $\hat{G}_n(t)$ but replacing residuals by errors. We must prove that

$$\int_{I^2} D\{Q_n(t)\}d\hat{G}_n(t) - \int_{I^2} D\{Q(t)\}dG(t) = o_p(1). \tag{A5}$$

From (A4) we obtain that $\hat{G}_n(t) - G_n(t) = n^{-1/2}\{\hat{W}_n(t) - W_n(t)\} = n^{-1/2}\{E_n(t) + Z_n(t) + t_1 B_{2n}(t_2) + t_2 B_{1n}(t_1)\}$. Hence, $\sup_{t\in I^2}\left|\hat{G}_n(t) - G_n(t)\right| = o_p(1)$, and (A5) may be proved from this result using the Skorohod embedding theorem.

Proof of Theorem 1: Apply Proposition A1 with $A_{1i} = A_i$, $A_{2i} = A_{i+1}$ for $A = Y, X, U$. Under H_0, all conditions in part (b) of Proposition A1 hold with $m = 1$, and, except for terms which are uniformly $o_p(1)$, $\hat{P}_n(\cdot)$, $P_n(\cdot)$, $H(\cdot)$, $H_1(\cdot)$, $H_2(\cdot)$ become, respectively, $n^{1/2}\hat{S}_n^*(\cdot)$, $n^{1/2}S_n^*(\cdot)$, $F(\cdot)$, $F_1(\cdot)$, $F_1(\cdot)$, where $\hat{S}_n^*(t) = \hat{S}_n\{F_1^{-1}(t_1), F_1^{-1}(t_2)\}$ and $S_n^*(t) = S_n\{F_1^{-1}(t_1), F_1^{-1}(t_2)\}$.

Proof of the Corollary: Under H_0, apply part (b) of Proposition A1 to deduce that $n^{1/2}\hat{S}_n^*(\cdot) \Rightarrow S_\infty^*(\cdot)$, where $S_\infty^*(t) = S_\infty\{F_1^{-1}(t_1), F_1^{-1}(t_2)\}$; then use part (c) of Proposition A1. Under H_1, apply part (a) of Proposition A1 and then use part (c) to derive that $n^{-1}\hat{C}_n$ converges in probability to $\Delta = \int_{\mathbb{R}^2}\{F(u_1, u_2) - F_1(u_1)F_2(u_2)\}^2 dF(u_1, u_2)$. As H_1 is true and $F(\cdot)$ is continuous, then $\Delta > 0$ (Blum et al. 1961, p. 490).

REFERENCES

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

BLUM, J. R., KIEFER, J. & ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485-98.

CSÖRGÖ, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process. *J. Mult. Anal.* **9**, 84-100.

DELGADO, M. A. (1996). Testing serial independence using the sample distribution function. *J. Time Ser. Anal.* **17**, 271-86.

DELGADO, M. A. (1999). Hoeffding-Blum-Kiefer-Rosenblatt process. In *Encyclopedia of Statistical Sciences Update, Vol.3*, Eds. S. Kotz, C. B. Read & D.L. Banks, pp. 326-8. New York: Wiley.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279-90.

HOEFFDING, W. (1948). A nonparametric test of independence. *Ann. Math. Statist.* **26**, 189-211.

HONG, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *J. R. Statist. Soc.* B **60**, 429-53.

KOUL, H. (1992). *Weighted Empiricals and Linear Models.* IMS Lecture Notes 21. Hayward, CA: Institute of Mathematical Statistics.

NEUHAUS, G. (1971). On weak convergence of stochastic processes with multi-dimensional time parameter. *Ann. Math. Statist.* **42**, 1985-95.

SKAUG, H. J. & TJØSTHEIM, D. (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591-602.

STUTE, W. & SCHUMANN, G. (1980). A general Glivenko-Cantelli theorem for stationary sequences of random observations. *Scand. J. Statist.* **7**, 102-4.