Working Paper 98-56
Statistics and Econometrics Series 25
July 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34)91 624-9849

A NONPARAMETRIC TEST FOR SERIAL INDEPENDENCE OF ERRORS IN LINEAR REGRESSION.

Miguel A. Delgado and Juan Mora.*

Abstract

A test for serial independence of regression errors, consistent in the direction of first order alternatives, is proposed. 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 tests.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Ph: 34-1-624.98.04; Fax: 34-1-624.98.49, e-mail: delgado@est-econ.uc3m.es; Mora, Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Apto. Correos 99, 03080 Alicante, España, e-mail: juan@merlin.fae.ua.es.

1. PRELIMINARIES AND STATEMENT OF THE PROBLEM.

Consider observations $\{Y_i,\ i = 1, \ldots, n+1\}$, which are related to a k-dimensional vector of design variables $\{\mathbf{x}_i,\ i = 1, \ldots, n+1\}$ according to the linear model $Y_i = \mathbf{x}_i'\beta_0 + U_i$, where β_0 a is k-dimensional vector of unknown parameters and the errors $\{U_i,\ i = 1, \ldots, n+1\}$ are unobservable real random variables. Under the maintained hypothesis of stationary $\{U_i,\ i\geq 1\}$, we are interested in testing serial independence of the unobserved series $\{U_i,\ i\geq 1\}$ consistently in the direction of general first order dependence alternatives. Formally, the null and alternative hypotheses can be written as,

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

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

where $\mathbf{u} = (u_1,u_2)'$, $S(\mathbf{u}) \equiv F(\mathbf{u}) - F_1(u_1)F_1(u_2)$, $F(.)$ is the joint distribution function of $(U_i,U_{i+1})'$ and $F_1(.)$ is the marginal distribution function of U_i.

For observable $\{U_i,\ i\geq 1\}$, Skaug and Tjøstheim (1993), Delgado (1996) and Hong (1998), among others, have proposed test statistics based on functionals of the Hoeffding-Blum-Kiefer-Rosenblatt (HBKR) empirical process

$$S_n(\mathbf{u}) = F_n(\mathbf{u}) - F_{1n}(u_1)F_{2n}(u_2),$$

where $F_n(\mathbf{u}) \equiv n^{-1}\sum_{i=1}^{n} I(U_i\leq u_1)I(U_{i+1}\leq u_2)$ estimates $F(\mathbf{u})$, $F_{1n}(u_1) \equiv n^{-1}\sum_{i=1}^{n} I(U_i\leq u_1)$ estimates $F_1(u_1)$, $F_{2n}(u_2) \equiv n^{-1}\sum_{i=2}^{n+1} I(U_i\leq u_2)$ estimates $F_1(u_2)$ and, hence, $S_n(\mathbf{u})$ estimates $S(\mathbf{u})$; $I(.)$ denotes the indicator function.

Functionals of $n^{1/2}S_n(\mathbf{u})$ form a basis for constructing test statistics of H_0 (see, e.g., Delgado 1998). A popular one is the Cramèr-von Mises statistic $C_n \equiv n^{-1}\sum_{i=1}^{n}[n^{1/2}S_n(U_i,U_{i+1})]^2$. Hoeffding (1948) and Blum *et al* (1961) proposed this type of statistic in the context of testing independence between two samples, and tabulated its limiting distribution under the null hypothesis. Skaug and Tjøstheim (1993) show that, if $F(.)$ is continuous, then C_n has the same limiting distribution as the statistic of Blum *et al* (1961). Other functionals of $n^{1/2}S_n(\mathbf{u})$ could be used, e.g., based on the Kolmogorov-Smirnov norm.

We propose to test H_0 using residuals $\hat{U}_{ni} \equiv Y_i - \mathbf{x}_i'\hat{\beta}_n$, where $\hat{\beta}_n$ is some reasonable estimate of β_0 (as usual, hereafter we suppress the subscript n and denote simply $\hat{U}_i$ and $\hat{\beta}$). Thus, $S(\mathbf{u})$ is estimated by the empirical process

$$\hat{S}_n(\mathbf{u}) \equiv \hat{F}_n(\mathbf{u}) - \hat{F}_{1n}(u_1)\hat{F}_{2n}(u_2),$$

where $\hat{F}_n(\mathbf{u})$, $\hat{F}_{1n}(u_1)$ and $\hat{F}_{2n}(u_2)$ are defined as $F_n(\mathbf{u})$, $F_{1n}(u_1)$ and $F_{2n}(u_2)$, respectively, but replacing errors U_i by residuals $\hat{U}_i$. Functionals of $n^{1/2}\hat{S}_n(\mathbf{u})$ can be used as test statistics, e.g., the Cramèr-von Mises statistic $\hat{C}_n \equiv n^{-1}\sum_{i=1}^{n}[n^{1/2}\hat{S}_n(\hat{U}_i,\hat{U}_{i+1})]^2$.

In the next section we discuss the asymptotic behaviour of the empirical process $n^{1/2}\hat{S}_n(\mathbf{u})$, derive the limiting distribution of $\hat{C}_n$ under the null hypothesis and show that the test based on $\hat{C}_n$ is consistent. Surprisingly, $n^{1/2}\hat{S}_n(\mathbf{u})$ converges to the same limiting process as $n^{1/2}S_n(\mathbf{u})$. This is not the case with other empirical process which depend on parameter estimates, as those used in goodness-of-fit tests (see, e.g., Durbin 1973). In Section 3 we report the results of a small Monte Carlo experiment. Proofs are confined to the Appendix.

2. ASYMPTOTIC PROPERTIES

The following assumptions will be used to derive asymptotic properties:

$$Y_i = \mathbf{x}_i'\beta_0 + U_i, \quad i \geq 1, \tag{1}$$

where $\{U_i, i \geq 1\}$ is a strictly stationary sequence of real random variables;

$\mathbf{X}_n \equiv [\mathbf{x}_1, ..., \mathbf{x}_n]'$ is a non-random full-rank matrix and

$$\max_{1 \leq i \leq n} \mathbf{x}_i'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_i = o(1); \tag{2}$$

The distribution function of $(U_i, U_{i+1})'$ has a density function with marginal density function $h_1(.)$ continuous and such that $h_1(x) > 0$, for all $x \in \mathbb{R}$. (3)

$$(\mathbf{X}_n'\mathbf{X}_n)^{1/2}(\hat{\beta} - \beta_0) = O_P(1). \tag{4}$$

Assumption (2) is typical when studying asymptotic properties of statistics in this context. Observe that this assumption does not rule out trending explanatory variables. Under assumption (3), which is necessary to ensure that empirical processes based on residuals behave properly (see Koul 1992), the marginal distribution function is strictly increasing. When (2) holds, assumption (4) is satisfied by most estimates, e.g. ordinary least squares and least absolute deviations.

Hereafter, the interval $[0,1]$ is denoted by I, the rectangle $[0,1]^2$ is denoted by I^2, $\mathbb{D}(I^2)$ denotes the set of all real functions on I^2 which are "continuous from above with limits from below" as in Definition 1.1. of Neuhaus (1971), and "$\Rightarrow$" denotes weak convergence of stochastic processes. As usual, to derive asymptotic results it is convenient to express $\hat{S}_n(\mathbf{u})$ as an empirical process in $\mathbb{D}(I^2)$. For $\mathbf{t} = (t_1,t_2)' \in I^2$, we denote $\hat{S}^*_n(\mathbf{t}) \equiv \hat{S}_n(F_1^{-1}(t_1),F_1^{-1}(t_2))$ and $S^*_n(\mathbf{t}) \equiv S_n(F_1^{-1}(t_1),F_1^{-1}(t_2))$, where $F_1(.)$ denotes the distribution function of U_i. The following theorem provides the first order asymptotic equivalence between the empirical processes $\hat{S}^*_n(\mathbf{t})$ and $S^*_n(\mathbf{t})$.

Theorem: Assume that (1), (2), (3) and (4) hold. Then:

a) Under H_0, $\sup_{\mathbf{t}\in I^2} |\hat{S}^*_n(\mathbf{t})-S^*_n(\mathbf{t})| = o_p(n^{-1/2})$.

b) Under H_1, if $\{U_i,\ i\geq 1\}$ is ergodic, then $\sup_{\mathbf{t}\in I^2} |\hat{S}^*_n(\mathbf{t})-S^*_n(\mathbf{t})| = o_p(1)$. ■

From this result it follows straightforwardly that, if (1)-(4) hold, then, under H_0, $n^{1/2}\hat{S}^*_n(\mathbf{t}) \Rightarrow S_\infty(\mathbf{t})$, where $S_\infty(\mathbf{t})$ is a Gaussian process in $\mathbb{D}(I^2)$ with zero mean and covariance structure: $\text{cov}(S_\infty(\mathbf{s}),S_\infty(\mathbf{t})) = [\min\{s_1,t_1\}-s_1t_1][\min\{s_2,t_2\}-s_2t_2]$; and, under H_1, $\hat{S}^*_n(\mathbf{t})$ converges in probability to $F(F_1^{-1}(t_1),F_1^{-1}(t_2)) - t_1t_2$; (see Proof of the Corollary in the appendix below). This results are exploited in the following corollary, which justifies asymptotic inferences based on $\hat{C}_n$.

Corollary: Assume that (1), (2), (3) and (4) hold. Then:

a) Under H_0, $\hat{C}_n$ converges in distribution to $C_\infty \equiv \int_{I^2} S_\infty(\mathbf{t})^2 d\mathbf{t}$.

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

The Corollary guarantees the implementation of the test using $\hat{C}_n$ and critical values from the distribution of the random variable C_∞, which has been tabulated by Blum *et al* (1961). This result may seem surprising at first sight because, when testing goodness of fit, the asymptotic distribution of the test statistic computed with observations is not the same as the asymptotic distribution of the test statistic computed with residuals (see, for example, Koul 1992). When testing goodness of fit, replacing β_0 by $\hat{\beta}$ introduces a random term in the empirical distribution function and this affects the distribution of the test statistic. When testing independence, replacing the true parameter value by an estimator introduces random terms in the joint empirical distribution function and in the marginal empirical distribution functions, but these random terms cancel out asymptotically when we consider the HBKR empirical process.

3. SIMULATIONS

In order to examine how the replacement of observations by residuals affects the finite-sample behaviour of the test statistic, we have carried out some Monte Carlo experiments, similar to those performed in Skaug and Tjøstheim (1993). All programs have been written in GAUSS. To study the size of the test, we have generated observations from a regression model satisfying (1), with $x_i' = (1,i)$, $\beta_0' = (1,1)$ and errors $\{U_i\}_{i=1}^{n+1}$ generated independently from a standard normal distribution. We compare the behaviour C_n, statistic based on errors, and $\hat{C}_n$, statistic based on least squares residuals. In Table 1 we report the percentage of rejections of the null hypothesis for different theoretical significance levels α and sample sizes n. Reported values are based on *5000* Monte Carlo replications. As critical values we used: *0.04694* for $\alpha=0.1$, *0.05840* for $\alpha=0.05$ and *0.08685* for $\alpha=0.01$; these values have been obtained from Table II in Blum *et al.* (1961).

TABLE 1 ABOUT HERE

In this table we observe that the results obtained when using residuals are similar to those obtained with errors. Moreover, the real level of the test is not far from the intended level, regardless of whether the statistic is computed with errors or residuals. To study the power of the test, we have performed Monte Carlo experiments with the same characteristics as those described in Skaug and Tjøstheim (1993), Section 4.4. The results of these experiments are not reported. When using errors, as expected, we obtain the same results as those reported by Skaug and Tjøstheim (1993) in Figure 1. When using residuals, all results are similar to those obtained with errors.

ACKNOWLEDGEMENTS

We are grateful to Spanish Dirección General de Investigación Científica y Técnica (DGICYT) for financial support (Reference number: PB95-0292).

APPENDIX: Proofs.

Hereafter, $\mathbf{t} = (t_1,t_2)'$ is a generic element in I^2, $j=1,2$ and $i=1,...,n$ unless otherwise stated; and $\mathbb{C}(I^2)$ is the set of all real continuous functions on I^2.

The proof of Theorem and Corollary will be derived from Propositions 1 and 2 below. In these propositions, notation and assumptions are as follows: $\{(Y_{1i},\mathbf{x}'_{1i},Y_{2i},\mathbf{x}'_{2i})'\}_{i=1}^{n}$ are observations from an $\mathbb{R}\times\mathbb{R}^{p_1}\times\mathbb{R}\times\mathbb{R}^{p_2}$ variable such that the following linear regression models hold:

$$Y_{ji} = \mathbf{x}'_{ji}\beta_{j0} + U_{ji}, \tag{C1}$$

where $\{(U_{1i},U_{2i})',\ i\geq 1\}$ is a strictly stationary sequence of $\mathbb{R}\times\mathbb{R}$ random vectors, $H(.)$ is the distribution function of $(U_{1i},U_{2i})'$ and $H_1(.)$, $H_2(.)$ are its marginal distribution functions. We will also assume that:

$$\mathbf{X}_{jn} \equiv [\mathbf{x}_{j1}, \ldots, \mathbf{x}_{jn}]' \text{ are non-random full-rank matrices and } \max_{1\leq i\leq n} \mathbf{x}_{ji}'(\mathbf{X}'_{jn}\mathbf{X}_{jn})^{-1}\mathbf{x}_{ji} = o(1); \tag{C2}$$

$H(.)$ has a density function $h(.)$ whose marginal density functions $h_1(.)$, $h_2(.)$ are continuous and positive in $\mathbb{R}$; (C3)

$$(\mathbf{X}'_{jn}\mathbf{X}_{jn})^{1/2}(\hat{\beta}_j-\beta_{j0}) = O_P(1), \tag{C4}$$

where $\hat{\beta}_j$ is an estimator of β_{j0}. Other assumptions which will be required in some results are the following:

$\{(U_{1i},U_{2i})',\ i\geq 1\}$ is an ergodic sequence; (C5)

$\{(U_{1i},U_{2i})',\ i\geq 1\}$ is an m-dependent sequence, for $m \in \mathbb{N}\cup\{0\}$; (C6)

the concept of m-dependent sequences can be found, e.g., in Billingsley (1968), p. 167;

$$H(.) = H_1(.)H_2(.). \tag{C7}$$

With this notation we define $P_n(t) \equiv n^{1/2}\{n^{-1}\sum_{i=1}^{n} I(H_1(U_{1i})\leq t_1)I(H_2(U_{2i})\leq t_2) - n^{-2}\sum_{i=1}^{n} I(H_1(U_{1i})\leq t_1)\sum_{i=1}^{n} I(H_2(U_{2i})\leq t_2)\}$ and $\tilde{P}_n(t)$ in the same way as $P_n(t)$, but replacing errors U_{ji} by residuals $\hat{U}_{ji} \equiv Y_{ji} - \mathbf{x}'_{ji}\hat{\beta}_j$.

Proposition 1: Assume that (C1), (C2), (C3) and (C4) hold and define $G(\mathbf{t}) \equiv H(H_1^{-1}(t_1),H_2^{-1}(t_2))$, if $\mathbf{t} \in (0,1)\times(0,1)$ or t_1t_2 otherwise. Then:

a) If (C5) holds, then $P_n(\mathbf{t})$, $\tilde{P}_n(\mathbf{t})$ are processes in $\mathbb{D}(I^2)$ such that:

i) $\sup_{\mathbf{t}\in I^2} |\tilde{P}_n(\mathbf{t})-P_n(\mathbf{t})| = o_p(n^{1/2})$;

ii) $n^{-1/2}\tilde{P}_n(\mathbf{t})$ converges in probability to $L(\mathbf{t}) \equiv G(\mathbf{t}) - t_1t_2$.

b) If (C6) and (C7) hold then P_n, $\tilde{P}_n$ are processes in $\mathbb{D}(I^2)$ with:

i) $\sup_{\mathbf{t}\in I^2} |\tilde{P}_n(\mathbf{t})-P_n(\mathbf{t})| = o_p(1)$;

ii) $\tilde{P}_n(\mathbf{t}) \Rightarrow P^{(m)}(\mathbf{t})$, where $P^{(m)}(\mathbf{t})$ is a Gaussian process in $\mathbb{D}(I^2)$ centered at zero and with the following covariance structure: if $m>0$,

$$\text{cov}(P^{(m)}(\mathbf{s}),P^{(m)}(\mathbf{t})) = [\min\{s_1,t_1\}-s_1t_1][\min\{s_2,t_2\}-s_2t_2] +$$
$$\Sigma_{k=1}^{m} E[\Pi_{j=1}^{2}\{I(H_j(U_{j1})\le s_j)-s_j\}\{I(H_j(U_{j,k+1})\le t_j)-t_j\}] +$$
$$\Sigma_{k=1}^{m} E[\Pi_{j=1}^{2}\{I(H_j(U_{j,k+1})\le s_j)-s_j\}\{I(H_j(U_{j,1})\le t_j)-t_j\}],$$

and, if $m=0$, $\text{cov}(P^{(0)}(\mathbf{s}),P^{(0)}(\mathbf{t})) = [\min\{s_1,t_1\}-s_1t_1][\min\{s_2,t_2\}-s_2t_2]$.

Proof:

a-i) Define $\hat{W}_n(\mathbf{t}) \equiv n^{-1/2}\Sigma_{i=1}^{n}\{I(H_1(\hat{U}_{1i})\le t_1)I(H_2(\hat{U}_{2i})\le t_2) - G(\mathbf{t})\}$, and $\hat{W}_{jn}(t_j) \equiv n^{-1/2}\Sigma_{i=1}^{n}\{I(H_j(\hat{U}_{ji})\le t_j)-t_j\}$. With these definitions,

$$\tilde{P}_n(\mathbf{t}) = \hat{W}_n(\mathbf{t})-t_2\hat{W}_{1n}(t_1)-t_1\hat{W}_{2n}(t_2)-n^{-1/2}\hat{W}_{1n}(t_1)\hat{W}_{2n}(t_2)+n^{1/2}(G(\mathbf{t})-t_1t_2). \quad \text{(A1)}$$

In a similar way, when using errors instead of residuals

$$P_n(\mathbf{t}) = W_n(\mathbf{t})-t_2W_{1n}(t_1)-t_1W_{2n}(t_2)-n^{-1/2}W_{1n}(t_1)W_{2n}(t_2)+n^{1/2}(G(\mathbf{t})-t_1t_2), \quad \text{(A2)}$$

where $W_n(\mathbf{t})$ and $W_{jn}(t_j)$ are defined in the same way as $\hat{W}_n(\mathbf{t})$ and $\hat{W}_{jn}(t_j)$, respectively, but replacing $\hat{U}_{ji}$ by U_{ji}. Given $\mathbf{v}_j \in \mathbb{R}^{p_j}$, define

$g_j(t_j) \equiv h_j(H_j^{-1}(t_j))$ if $t_j \in (0,1)$ or 0 otherwise;

$$G_{jni}(t_j, \mathbf{v}_j) \equiv H_j(H_j^{-1}(t_j) + \mathbf{x}'_{ji}(\mathbf{X}'_{jn}\mathbf{X}_{jn})^{-1/2}\mathbf{v}_j) \text{ if } t_j \in (0,1) \text{ or } t_j \text{ otherwise;}$$

$$\hat{t}_{jni} \equiv G_{jni}(t_j, (\mathbf{X}'_{jn}\mathbf{X}_{jn})^{1/2}(\hat{\beta}_j - \beta_{j0}));$$

$$G_{ni}(\mathbf{t}, \mathbf{v}_1, \mathbf{v}_2) = H(H_1^{-1}(t_1) + \mathbf{x}'_{1i}(\mathbf{X}'_{1n}\mathbf{X}_{1n})^{-1/2}\mathbf{v}_1,\ H_2^{-1}(t_2) + \mathbf{x}'_{2i}(\mathbf{X}'_{2n}\mathbf{X}_{2n})^{-1/2}\mathbf{v}_2)$$
$$\text{if } \mathbf{t} \in (0,1) \times (0,1) \text{ or } t_1 t_2 \text{ otherwise;}$$

$$\hat{\mathbf{t}}_{ni} \equiv G_{ni}(\mathbf{t}, (\mathbf{X}'_{1n}\mathbf{X}_{1n})^{1/2}(\hat{\beta}_1 - \beta_{10}), (\mathbf{X}'_{2n}\mathbf{X}_{2n})^{1/2}(\hat{\beta}_2 - \beta_{20})).$$

Note that, as $H_j(.)$ is a one-to-one mapping, if $t_j \in (0,1)$ then $I(H_j(\hat{U}_{ji}) \le t_j) = I(U_{ji} \le H_j^{-1}(t_j) + \mathbf{x}'_{ji}(\hat{\beta}_j - \beta_{j0})) = I(H_j(U_{ji}) \le \hat{t}_{jni})$, and these equalities also hold if $t_j = 0$ or 1. Hence,

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

$$\hat{W}_n(\mathbf{t}) = E_n(\mathbf{t}) + Z_n(\mathbf{t}) + t_1 B_{2n}(t_2) + t_2 B_{1n}(t_1) + W_n(\mathbf{t}), \tag{A4}$$

where we define,

$$E_{jn}(t_j) \equiv n^{-1/2}\Sigma_{i=1}^n \{I(H_j(U_{ji}) \le \hat{t}_{jni}) - \hat{t}_{jni} - I(H_j(U_{ji}) \le t_j) + t_j\};$$

$$Z_{jn}(t_j) \equiv n^{-1/2}\Sigma_{i=1}^n \{\hat{t}_{jni} - t_j\} - n^{-1/2} g_j(t_j)\Sigma_{i=1}^n \mathbf{x}'_{ji}(\hat{\beta}_j - \beta_{j0});$$

$$B_{jn}(t_j) \equiv n^{-1/2} g_j(t_j)\Sigma_{i=1}^n \mathbf{x}'_{ji}(\hat{\beta}_j - \beta_{j0});$$

$$E_n(\mathbf{t}) \equiv n^{-1/2}\Sigma_{i=1}^n \{I(H_1(U_{1i}) \le \hat{t}_{1ni}) I(H_2(U_{2i}) \le \hat{t}_{2ni}) - \hat{\mathbf{t}}_{ni} -$$
$$I(H_1(U_{1i}) \le t_1) I(H_2(U_{2i}) \le t_2) + G(\mathbf{t})\};$$

$$Z_n(\mathbf{t}) \equiv n^{-1/2}\Sigma_{i=1}^n \{\hat{\mathbf{t}}_{ni} - G(\mathbf{t})\} - t_2 B_{1n}(t_1) - t_1 B_{2n}(t_2).$$

Under (C1)-(C5), it holds that $\sup_{t \in I} |Z_{jn}(t)| = o_P(1)$, $\sup_{\mathbf{t} \in I^2} |n^{-1/2} Z_n(\mathbf{t})| = o_P(1)$, $\sup_{t \in I} |n^{-1/2} E_{jn}(t)| = o_P(1)$, $\sup_{\mathbf{t} \in I^2} |n^{-1/2} E_n(\mathbf{t})| = o_P(1)$, $\sup_{t \in I} |B_{jn}(t)| = O_P(1)$, $\sup_{t \in I} |n^{-1/2} W_{jn}(t)| = o_P(1)$. These results may be proved using similar arguments as in Koul (1992) and a generalization of Theorem 2.4.3 in Koul (1992) which allows to use m-dependent sequences; detailed proofs of these results are available from the authors on request. Using these results and (A1), (A2), (A3), (A4), it follows that $\sup_{t\in I^2} n^{-1/2}|\tilde{P}_n(t)-P_n(t)| = o_P(1)$.

a-ii: Note that $n^{-1/2}P_n(t) - L(t) = n^{-1}\sum_{i=1}^n\{I(H_1(U_{1i})\leq t_1)I(H_2(U_{2i})\leq t_2) - G(t)\} - n^{-1/2}\{t_2W_{1n}(t_1) + t_1W_{2n}(t_2) + n^{-1/2}W_{1n}(t_1)W_{2n}(t_2)\}$. Using that $\sup_{t\in I}|n^{-1/2}W_{jn}(t)| = o_P(1)$ and the Glivenko-Cantelli Theorem in Stute and Schumann (1980), it follows that $\sup_{t\in I^2}|n^{-1/2}P_n(t)-L(t)| = o_P(1)$. Using Theorem 4.1 in Billingsley (1968), it follows that $n^{-1/2}\tilde{P}_n(t)$ converges in probability to $L(t)$.

b-i: If (C1)-(C4), (C6) and (C7) hold, it is possible to prove that $\sup_{t\in I}|Z_{jn}(t)|= o_P(1)$, $\sup_{t\in I^2}|Z_n(t)|= o_P(1)$, $\sup_{t\in I}|E_{jn}(t)|= o_P(1)$, $\sup_{t\in I^2}|E_n(t)|= o_P(1)$, $\sup_{t\in I}|B_{jn}(t)|= O_P(1)$, $\sup_{t\in I}|W_{jn}(t)|= O_P(1)$. Using these results, (A1), (A2), (A3) and (A4), it follows that $\sup_{t\in I^2}|\tilde{P}_n(t)-P_n(t)|= o_P(1)$.

b-ii: Using Theorem 4.1 in Billingsley (1968), it suffices to prove that $P_n(t) \Rightarrow P^{(m)}(t)$. If we denote $V_n(t) \equiv W_n(t) - t_2W_{1n}(t_1) - t_1W_{2n}(t_2)$, from (A2) it follows that $P_n(t) = V_n(t) - n^{-1/2}W_{1n}(t_1)W_{2n}(t_2)$, because now $G(t) = t_1t_2$. As $\sup_{t\in I}|W_{jn}(t)| = O_P(1)$, it suffices to prove that $V_n(t) \Rightarrow P^{(m)}(t)$. The convergence of finite-dimensional distributions follows using Cramer-Wold device and Theorem 27.4 in Billingsley (1995); and using Theorem 4 in Csörgö (1979), it follows that $\lim_{\delta\to 0} [\limsup_{n\to\infty} \mathrm{pr}\{\sup_{\|t-s\|<\delta}|V_n(t) - V_n(s)|\geq\varepsilon\}] = 0$. So, from the results in Neuhaus (1971) or Straf (1971), $V_n(t) \Rightarrow P^{(m)}(t)$. ∎

Proposition 2: Let $D:\mathbb{R} \longrightarrow \mathbb{R}$ be a continuous function and $Q_n(t)$, $Q(t)$ processes in $\mathbb{D}(I^2)$ such that $\mathrm{pr}\{Q(t) \in \mathbb{C}(I^2)\} = 1$ and $Q_n(t) \Rightarrow Q(t)$. If (C1), (C2), (C3), (C4) and (C5) hold, then $n^{-1}\sum_{i=1}^n D(Q_n(H_1(\hat{U}_{1i}),H_2(\hat{U}_{2i})))$ converges in distribution to $\int_{I^2} D(Q(t))dG(t)$, where $G(.)$ is as defined in Proposition 1.

Proof: Denote $\hat{G}_n(t) \equiv n^{-1}\sum_{i=1}^{n} I(H_1(\hat{U}_{1i})\leq t_1)I(H_2(\hat{U}_{2i})\leq t_2)$, and $G_n(t)$ 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 $\hat{G}_n(t) - G_n(t) = n^{-1/2}[\hat{W}_n(t) - W_n(t)] = n^{-1/2}[E_n(t) + Z_n(t) + t_1B_{2n}(t_2) + t_2B_{1n}(t_1)]$. As $\sup_{t\in I^2}|n^{-1/2}Z_n(t)| = o_p(1)$, $\sup_{t\in I^2}|n^{-1/2}E_n(t)| = o_p(1)$, $\sup_{t\in I}|B_{jn}(t)| = O_p(1)$, then $\sup_{t\in I^2}|\hat{G}_n(t)-G_n(t)| = o_p(1)$. Using the Glivenko-Cantelli Theorem in Stute and Schumann (1980), it follows that $\hat{G}_n$ converges in prob. to $G(.)$. Hence $(Q_n,\hat{G}_n)$ converges in distribution (in $\mathbb{D}(I^2) \times \mathbb{D}(I^2)$) to (Q,G) and, by Skorohod embedding theorem, we can find Q_n^*, $\hat{G}_n^*$, Q^*, random elements from a certain probability space to $\mathbb{D}(I^2) \times \mathbb{D}(I^2)$, with the same distribution as Q_n, $\hat{G}_n$, Q, and such that $(Q_n^*,\hat{G}_n^*)$ converges almost surely to (Q^*,G). So (A5) will follow if we prove

$$\int_{I^2} D(Q_n^*(t))d\hat{G}_n^*(t) - \int_{I^2} D(Q^*(t))dG(t) = o_p(1). \tag{A6}$$

If $D(.)$ is bounded and uniformly continuous then (A6) holds almost surely. Using this result it follows easily that (A6) holds for any continuous $D(.)$. ∎

Proof of Theorem:

a) Apply Proposition 1 with $A_{1i} \equiv A_i$, $A_{2i} \equiv A_{i+1}$ for $A = Y$, x, U. All conditions in Proposition 1.b. hold with $m=1$; and $\tilde{P}_n(t)$, $P_n(t)$, $H(.)$, $H_1(.)$, $H_2(.)$ become, respectively, $n^{1/2}\hat{S}_n^*(t)$, $n^{1/2}S_n^*(t)$, $F(.)$, $F_1(.)$, $F_1(.)$; hence, from Proposition 1.b.i, $\sup_{t\in I^2}|\hat{S}_n^*(t)-S_n^*(t)| = o_p(n^{-1/2})$.

b) Apply Proposition 1 as before. All conditions in Proposition 1.a hold; hence, from Proposition 1.a.i, $\sup_{t\in I^2}|\hat{S}_n^*(t)-S_n^*(t)| = o_p(1)$. ∎

Proof of Corollary:

a) Applying Proposition 1.b as in the Theorem, if follows from Proposition 1.b.ii that $n^{1/2}\hat{S}^*_n(t) \Rightarrow S_\infty(t)$, because the limiting process has the same covariance structure as $S_\infty(t)$ (all additional terms turn out to be zero). As $\hat{C}_n = n^{-1}\Sigma_{i=1}^n[n^{1/2}\hat{S}^*_n(F_1(\hat{U}_i),F_1(\hat{U}_{i+1}))]^2$, this part of the corollary follows applying now Proposition 2 with the same notation as in the previous Theorem and $D(x) = x^2$, $Q_n(t) = n^{1/2}\hat{S}^*_n(t)$, $Q(t) = S_\infty(t)$.

b) Applying Proposition 1.a as in the Theorem, if follows from Proposition 1.a.ii that $\hat{S}^*_n(t)$ converges in probability to $G(t) - t_1t_2$, where now $G(t) = F(F_1^{-1}(t_1),F_1^{-1}(t_2))$. Applying Proposition 2 with $D(x) = x^2$, $Q_n(t) = \hat{S}^*_n(t)$, $Q(t) = G(t) - t_1t_2$, it follows that $n^{-1}\hat{C}_n$ converges in probability to $\int_{I^2} \{G(t)-t_1t_2\}^2dG(t) = \int_{\mathbb{R}^2} \{F(x_1,x_2)-F_1(x_1)F_1(x_2)\}^2dF(x_1,x_2) \equiv \Delta$. As H_1 is true and $F(.)$ is continuous then $\Delta>0$ (see Blum *et al.* 1961, p.490), and this part of the Corollary follows from this. ∎

REFERENCES

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

Billingsley, P. (1995). *Probability and Measure*. New York: Wiley.

Blum, J.R., Kiefer, J. & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics* **32**, 485-98.

Csörgö, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt multivariate empirical process. *Journal of Multivariate Analysis* **9**, 84-100.

Delgado, M.A. (1996). Testing serial independence using the sample distribution function. *Journal of Time Series Analysis* **17**, 271-86.

Delgado, M.A. (1998). Hoeffding-Blum-Kiefer-Rosenblatt process. In *Encyclopedia of Statistical Sciences Update*, Ed. Read, Banks and Kotz. New York: Wiley.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics* **1**, 279-90.

Hoeffding, W. (1948). A nonparametric test of independence. *Annals of Mathematical Statistics* **26**, 189-211.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *Journal of the Royal Statistical Society B* **60**, 429-53.

Koul, H. (1992). *Weighted Empiricals and Linear Models*. IMS Lecture Notes, 21.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics* **42**, 1285-95.

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

Straf, M.L. (1971). Weak convergence of stochastic processes with several parameters. *Proceedings Sixth Berkeley Symposium Mathematical Statistics and Probability* **2**, 187-221. Berkeley: University of California Press.

Stute, W. & Schumann, G. (1980). A general Glivenko-Cantelli theorem for stationary sequences of random observations. *Scandinavian Journal of Statistics* **7**, 102-4.

TABLE 1. Empirical Size

n	α=.10		α=.05		α=.01	
	C_n	$\hat{C}_n$	C_n	$\hat{C}_n$	C_n	$\hat{C}_n$
50	.1100	.1090	.0548	.0522	.0142	.0114
100	.1060	.1054	.0504	.0498	.0090	.0074
250	.1046	.0972	.0518	.0516	.0106	.0098