BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR–90/06

ASSESSING CONTEMPORANEOUS CORRELATION IN LONGITUDINAL DATA

by

Edward W. Frees
Visiting Principal Researcher

U.S. Bureau of the Census and

University of Wisconsin – Madison

Statistical Research Division
Bureau of the Census
Room 3000, FOB #4
Washington, DC 20233 U.S.A.

Recommended by: Nash J. Monsour

Report completed: March, 1990

Report issued: April 17, 1990

ASSESSING CONTEMPORANEOUS CORRELATION IN LONGITUDINAL DATA

by

Edward W. Frees

U. S. Bureau of the Census and
University of Wisconsin - Madison

School of Business and Department of Statistics
1155 Observatory Drive
Madison, Wisconsin 53706

March, 1990

Keywords and Phrases: Residuals, diagnostic tests, random effects, nonparametric estimation.

Running Head: ASSESSING CONTEMPORANEOUS CORRELATION

Author: Edward W. Frees
School of Business
and Department of Statistics
1155 Observatory Drive
Madison, Wisconsin 53706

ASSESSING CONTEMPORANEOUS CORRELATION IN LONGITUDINAL DATA

ABSTRACT

Consider drawing a sample of 'n' experimental units where each unit is observed over 'T' time periods. Are the draws independent? Small and large (in 'n') sample properties of a distribution-free statistic, built to assess the degree of correlation, are provided. Further, the large sample distribution is also investigated in the case where the statistic is evaluated using residuals from a complex model.

1. INTRODUCTION

A strong case can be made for the statement that the regression model is the pre-eminent statistical model, at least in terms of the number of applications. Because arguments using regression models have become widely accepted in scientific communities, researchers are becoming more willing to collect regression data over time. If the data is identified at the level of the experimental unit then longitudinal, or sometimes called pooled time series cross-sectional, models are appropriate for fitting. At least in the social sciences, even though this data often consists of many experimental units, experimenters are often unwilling or unable to consider the data over long periods of time because basic conditions appear unstable.

In the life sciences, often the parent population is sufficiently large and randomization techniques for observation selection are sufficiently sound so that experimental units may be modeled independently. In the social sciences, however, there may exist important correlations between experimental units especially in applications where economic entities are the experimental unit. I follow standard terminology in econometrics and call these "contemporaneous" correlations to distinguish them from autocorrelations, or correlations through time. If detected, there are several modeling techniques available. Generalized least squares techniques and, in particular, seemingly unrelated regression models of the variance structure, are prominent in applications. Further, contemporaneous correlations can be induced by omitted covariates that are possibly common to the experimental units. A subtler cause of contemporaneous correlation can arise from the random nature of one or several parameters in the model of the mean effect. The point here is that there are several type of models to handle contemporaneous correlation when detected. This article concerns "omnibus" methods for detecting contemporaneous correlations, particularly when the number of experimental units, n, is large compared to the number of observations per experimental unit, T. Hsiao (1986) provides a detailed discussion of the importance of considering the case of large n in addition to the usual approach in econometrics for considering large T.

Specifically, initially consider the model

$$Y_{it} = \mu_i + \sigma_i e_{it} \qquad i=1, \ldots, n \qquad t=1, \ldots, T \qquad (1.1)$$

where Y_{it}, $t=1, \ldots, T$ are observations from the i^{th} experimental unit. Assume $\{e_{it}\}$ are mean zero, unobservable random variables. Only the case of an equal number of observations per unit is explicitly discussed here as this is the rule when considering economic data (often violated, as with most rules). Many of the results presented in this article could be generalized to the unequal case with considerably more tedious algebra and at the expense of the interpretation of results. Both μ and σ may be considered to be either parameters to be estimated, covariate effects, or some combination. It is interesting to note that it turns out that μ_i may be either a fixed or random effect. It is well-known that location and scale parameters do not affect correlations and this also turns out to be the case here. Model extensions are considered in Section 4.

Classical statistics for assessing the independence of experimental observations can be found in multivariate analysis. Perhaps the most well-known is Bartlett's (1954) statistic, defined as

$$C_B = \det (B)^{T/2} / \prod_{i=1}^{n} b_{ii}^{T/2} \tag{1.2}$$

where B is a n x n matrix whose $(i,j)^{th}$ element is $b_{ij} = (T-1)^{-1} \sum_t (Y_{it} - \overline{Y}_i)(Y_{jt} - \overline{Y}_j)$. This statistic is the maximum likelihood statistic under normality. It is straightforward to check that rank (B) = minimum of {n, T-1} and thus det(B) = 0 for large n. The diagnostic statistics used in this paper are versions of a Lagrange Multiplier test statistic due to Breusch and Pagan (1980),

$$C^2_{AVE} = \binom{n}{2}^{-1} \sum_{i<j} c^2_{ij}, \tag{1.3}$$

where $c_{ij} = b_{ij} / (b_{ii}\, b_{jj})^{1/2}$ is the Pearson correlation coefficient between the i^{th} and j^{th} flows. Breusch and Pagan (1980) showed that as $T \to \infty$, the limiting distribution of $(T-1) \binom{n}{2} C^2_{AVE}$, is χ^2 with n(n-1)/2 degrees of freedom. This suggests the modified version $C_{STD} = n\,((T-1)\, C^2_{AVE} - 1)/2$ which has an asymptotic standard normal distribution (as $T \to \infty$ first, then as $n \to \infty$). However, when $n \to \infty$ and T remains fixed, it turns out that the finite, and even the limiting, distribution of C_{STD} is not distribution-free. For example, under the null hypothesis of i.i.d. errors in (1.1), note that $E\, C^2_{AVE} = E\, c^2_{12}$ depends on the distribution of errors. See Section 2 for further discussion of this point. This dependence on the parent distribution suggests using a nonparametric version of C^2_{AVE},

$$R^2_{AVE} = \binom{n}{2}^{-1} \sum_{i<j} r^2_{ij} \tag{1.4}$$

where r_{ij} is the Spearman rank correlation coefficient between the i^{th} and j^{th} flows. More specifically, define $r_{ij} = s_{ij} / (s_{ii}\ s_{jj})^{1/2}$, where $s_{ij} = (T-1)^{-1} \sum_t (R_{i,t} - (T+1)/2)(R_{j,t} - (T+1)/2)$ and $\{R_{i,1}, \ldots, R_{i,T}\}$ are the ranks of $\{Y_{i1}, \ldots, Y_{iT}\}$. This statistic is distribution-free and the limiting distribution is established in Section 3. In applications where either positive or negative contemporaneous correlations prevail, one could also consider the statistic

$$R_{AVE} = \binom{n}{2}^{-1} \sum_{i<j} r_{ij}. \tag{1.5}$$

The advantage of this statistic is that the limiting distribution is known, that is, $(T-1)\,((n-1)\, R_{AVE} + 1) = FR \to$

$\chi^2_{(T-1)}$ where FR is Friedman's statistic, cf., Hettmansperger (1984, pp. 196, 210). For completeness, also define

$$C_{AVE} = \binom{n}{2}^{-1} \Sigma_{i<j}\ c_{ij}. \qquad (1.6)$$

to be the version of R_{AVE} using Pearson correlation coefficients.

In Section 2, I discuss the finite sample properties of the statistics in (1.3)-(1.6). In the case of C^2_{AVE}, this supplements the known asymptotic (for large T) properties. The technique is to relate these statistics to a class of nonparametric unbiased estimators called U-statistics and use the well-known properties of this class. Because of the desirable properties of R^2_{AVE}, in Section 3 the large (in n) sample distribution of this statistic are established using classical techniques. Since the statistics are primarily for model diagnostics, they are most likely to be useful when evaluated using residuals from a preliminary model fit. Thus, in Section 4, I extend the discussion to statistics based on residuals. Sufficient conditions are provided so that the limiting distribution remains unchanged when using residuals in lieu of i.i.d. random variables. Further, the form of the distribution is identified when these conditions are not satisfied.

2. FINITE SAMPLE PROPERTIES

In this section, assume both n and T are fixed. Finite sample properties of the statistics (1.3)-(1.6) are developed using U-statistic theory, cf., Serfling (1980) for an introduction to these results.

Define $Y_i = (Y_{i1}, ..., Y_{It})'$ and $e_i = (e_{i1}, ..., e_{It})'$ to be random vectors in R^T, T-dimensional Euclidean space. Define four real-valued functions mapping R^{2T} into R, called kernels, corresponding to the four statistics, as follows. Define $h_1(Y_i, Y_j) = c_{ij}$, $h_2(Y_i, Y_j) = r_{ij}$, $h_3 = h_1^2$ and $h_4 = h_2^2$. For each kernel, the U-statistic is defined to be the average over all possible evaluations of the kernel, that is,

$$U_{n,k} = \binom{n}{2}^{-1} \Sigma_{i<j}\ h_k(Y_i, Y_j) \qquad k=1, ..., 4. \qquad (2.1)$$

Note that $U_{n,1} = C_{AVE}$, $U_{n,2} = R_{AVE}$, $U_{n,3} = C^2_{AVE}$ and $U_{n,4} = R^2_{AVE}$. The simplicity of the results of this article rests in the fact that the correlation statistics are location and scale invariant so that $h_k(Y_i, Y_j) = h_k(e_i, e_j)$, that is, h_k is a function of the i.i.d. random vectors $\{e_i\}$.

Throughout this paper, I assume

A1. The random variables $\{e_{it}\}$ are i.i.d.

This can be weakened at certain points to exchangeability and/or conditionally uncorrelated ($E\,(h(e_1, e_2) \mid e_1) = 0$), but for ease of interpretation I use the stronger assumption A1. It is straightforward to establish that $E\, U_{n,k} = E\, h_k$ for k=1, ..., 4. One can also check that $E\, R_{AVE} = E\, C_{AVE} = 0$ under A1. The calculation of the variance is more complex. I summarize that calculation in the following

THEOREM 1. Assume A1. Then

$$\mathrm{Var}\, U_{n,k} = \binom{n}{2}^{-1} \mathrm{Var}\, h_k(e_1, e_2) \qquad k=1, ..., 4.$$

Before proving this result, it is helpful to note some important special cases.

COROLLARY 1. Assume A1 and that the random variables $\{e_{it}\}$ have a continuous distribution. Then $E\, R^2_{AVE} = (T-1)^{-1}$, $\mathrm{Var}\, R_{AVE} = \binom{n}{2}^{-1} (T-1)^{-1}$ and

$$\mathrm{Var}\, R^2_{AVE} = \binom{n}{2}^{-1} \{(T-2)(25T^2 -7T-54) / (18\,(T-1)^2(T^3-T))\}.$$

COROLLARY 2. Assume A1 and that the random variables $\{e_{it}\}$ are normally distributed. Then $E\, C^2_{AVE} = T^{-1}$, $\mathrm{Var}\, C_{AVE} = \binom{n}{2}^{-1} T^{-1}$ and

$$\mathrm{Var}\, C^2_{AVE} = \binom{n}{2}^{-1} 2\,(T-1) / \{T^2\,(T+2)\}.$$

These corollaries provide explicit moments for <u>fixed</u> values of n and T. The assumption of a continuous distribution is the usual one in rank statistics, made to prevent possible ties in the random variables. The statistic $h_1^2(Y_1, Y_2) = h_3(Y_1, Y_2)$ is the squared correlation or coefficient of determination between Y_1 and Y_2. This is well-known to be depend on the distribution of $\{Y_{it}\}$, at least for finite T. In Corollary 2 is an example of the calculation in an important special case.

To prove Theorem 1, first recall some terminology from U-statistics theory. In this article, I consider only kernels that are symmetric in their arguments and are square integrable, that is, $h(y_1,y_2) = h(y_2,y_1)$ and $E\, h^2(Y_1,Y_2) < \infty$. Such a kernel is said to be degenerate if $E\,(h(Y_1,Y_2) \mid Y_1=y_1) = E\, h(Y_1,Y_2)$ for all y_1. The main work in establishing Theorem 1 is in

LEMMA 1. Assume A1. Then h_k is degenerate, for each k=1, ..., 4.

Proof of Lemma 1: It is straightforward to check that each h_k is symmetric and square integrable. I only establish that h_4 is degenerate as the cases h_1, h_2 and h_3 are similar. For h_4, replace Y by e in the definition of b_{ij} so that $b_{ij} = (T-1)^{-1} \Sigma_t (e_{it} - \overline{e}_i)(e_{jt} - \overline{e}_j)$. Thus

$$E\,(h_4(e_1, e_2) \mid e_2) = E\,(((T-1)^{-1} \Sigma_t e_{1t} (e_{2t} - \overline{e}_2))^2 / (b_{11} b_{22}) \mid e_2)$$

$$= (T-1)^{-2} \Sigma_t \Sigma_u (e_{2t} - \overline{e}_2)(e_{2u} - \overline{e}_2) / b_{22}\ E\,(e_{1t} e_{1u} / b_{11})$$

$$= (T-1)^{-2}\ E\,(e_{11}^2 / b_{11})\ \Sigma_t (e_{2t} - \overline{e}_2)^2 / b_{22} + E\,(e_{11} e_{12} / b_{11})\ \Sigma_{t \neq u} (e_{2t} - \overline{e}_2)(e_{2u} - \overline{e}_2) / b_{22}$$

$$= (T-1)^{-2}\ \{E\,(e_{11}^2 / b_{11}) - E\,(e_{11} e_{12} / b_{11})\}\ \Sigma_t (e_{2t} - \overline{e}_2)^2 / b_{22}$$

$$= (T-1)^{-1}\ \{E\,(e_{11}^2 / b_{11}) - E\,(e_{11} e_{12} / b_{11})\}$$

which is a constant. Since $E\,E\,(h_4(e_1, e_2) \mid e_1) = E\,h_4(e_1, e_2)$, this constant must equal $E\,h_4$. This is sufficient for the result. Q. E. D.

The proof of Theorem 1 is an immediate consequence of Lemma 1 and the usual variance decomposition result for U-statistics (cf., Serfling, 1980, Lemma A, page 183). To establish Corollary 2, from standard linear model results, under normality we have $c_{12}^2 = F / (F + T - 1)$, where F has an F-distribution with 1 and T-1 degrees of freedom, respectively. Calculating the expected values of c_{12}^2 and c_{12}^4 is a pleasant exercise using moments of the χ^2 - distribution and is omitted. Calculating moments of r_{12}^2 is also straightforward but, as some of these results are also used in Section 3, the details are sketched below.

Proof of Corollary 1: First note that, under the assumption of no ties, we have s_{ii} is degenerate and equal to $(T^2 + T)/12$ (cf., Hettmansperger, 1984, p. 20). Now, without loss of generality, the ranks of one vector may be arranged in ascending order to get $h_4(e_1, e_2) =_D 144/(T^3 - T)^2 (\Sigma_t (t-m)(R_t - m))^2$ where $m=(T+1)/2$ and R_t is the rank of the t^{th} element of the second vector. Basic calculations yield $E\,(R_1 - m)^2 = (T^2-1)/12$ and $E\,(R_1-m)(R_2-m) = -(T+1)/12$. This is sufficient to check that $E\,h_4(e_1, e_2) = 1/(T-1)$.

To calculate $E\, h_4^2$, the following quantities are useful: $S_1 = (T^3 - T)/12$ and $S_2 = (3T^2 - 7)/20$. Some basic equalities are $E\,(R_1\text{-}m)^4 = S_1\, S_2$, $E\,(R_1\text{-}m)^3\,(R_2\text{-}m) = -S_1\, S_2 \,/\, (T(T-1))$, $E\,(R_1\text{-}m)^2(R_2\text{-}m)(R_3\text{-}m) = S_1\,(2S_2 - S_1) \,/\, (T(T-1)(T-2))$, $E\,(R_1\text{-}m)^2(R_2\text{-}m)^2 = S_1\,(S_1 - S_2) \,/\, (T(T-1))$ and $E\,(R_1\text{-}m)\,(R_2\text{-}m)\,(R_3\text{-}m)\,(R_4\text{-}m) = 3\,S_1\,(S_1 - 2S_2) \,/\, (T(T-1)(T-2)(T-3))$. After checking these basic equalities, the next major step is to verify that

$$E\, h_4^2(e_1, e_2) = T\,(E\,(R_1\text{-}m)^4)^2 + 4\,T\,(T\text{-}1)\,(E\,(R_1\text{-}m)^3\,(R_2\text{-}m))^2 + 6\,T\,(T\text{-}1)\,(T\text{-}2)\,(E\,(R_1\text{-}m)^2(R_2\text{-}m)(R_3\text{-}m))^2 + 3\,T\,(T\text{-}1)\,(E\,(R_1\text{-}m)^2(R_2\text{-}m)^2)^2 + T\,(T\text{-}1)\,(T\text{-}2)\,(T\text{-}3)\,(E\,(R_1\text{-}m)\,(R_2\text{-}m)\,(R_3\text{-}m)\,(R_4\text{-}m))^2.$$

Putting the basic equalities into this equation, and some tedious algebra, are sufficient for the result. Q. E. D.

3. LARGE SAMPLE PROPERTIES

I first provide a useful computational version of R_{AVE}^2 similar to one available for R_{AVE}, cf., Hettmansperger (1984, p.210). Define $Z_{i,t,u} = 12\,(R_{it} - m)(R_{iu} - m)/(T^3 - T)$ where m is the mean of R_{it}, that is $m = E\, R_{it} = (T+1)/2$. Now, from (1.4), with $s_{ii} = (T^2 + T)/12$ and $\Sigma_t\, Z_{i,t,u} = 0$, we have

$$R_{AVE}^2 = 144 \,/\, (\, n(n-1)\,(T^3 - T)^2)\; \Sigma_{i\neq j}\, (\Sigma_t\, (R_{it} - m)(R_{jt} - m))^2$$

$$= (\, n(n-1)\,)^{-1}\; \Sigma_{i\neq j}\, \Sigma_{t,u}\, Z_{i,t,u}\, Z_{j,t,u}$$

$$= (\, n(n-1)\,)^{-1}\; \Sigma_{t,u}\, (\, (\Sigma_i\, Z_{i,t,u})^2 - \;\Sigma_i\, Z_{i,t,u}^2\,). \qquad (3.1)$$

For small T and large n, the expression of R_{AVE}^2 in (3.1) is faster to compute than the definition in (1.4). The expression in (3.1) also provides some insights in the form of the limit distribution in Theorem 2. That is, the sums in (3.1) suggest limiting normal distributions, and when squared, suggest limiting weighted sums of χ^2 distributions as appear in Theorem 2.

I now establish the following

THEOREM 2. Assume A1 and that the random variables $\{e_{it}\}$ have a continuous distribution. Then,

$$n\,(R^2_{AVE} - (T-1)^{-1}) \to_D Y = a(T)(\chi^2_{1,T-1} - (T-1)) + b(T)(\chi^2_{2,T(T-3)/2} - T(T-3)/2).$$

Here, χ^2_1 and χ^2_2 are independent χ^2 random variables with T-1 and T(T-3)/2 degrees of freedom, respectively, and $a(T) = 4(T+2) / (5(T-1)^2 (T+1))$ and $b(T) = 2(5T+6) / (5T(T-1)(T+1))$.

Remarks: Note that E Z = 0. Further, a(T) and b(T) are constants which depend on T such that Var Z = $4(T-2)(25T^2 - 7T - 54) / (25T(T-1)^3 (T+1)) = 2 \text{ Var } h_4(Y_1, Y_2)$. Since the limiting standardized asymptotic variance equals the variance of the limiting distribution, under uniform integrability, the fact that Var Z ~ n Var R^2_{AVE} is expected.

The method of proof below utilizes a conversion of the class of statistics to a quadratic form and then applies classical techniques. An alternative method of proof would have been to appeal to results on degenerate U-statistics with limiting distributions which are weighted sums of χ^2, due to Gregory (1977) and Serfling (1980). However, these results were only explicitly stated for the case in which the arguments of the kernels are real-valued instead of vector-valued as is the case here. Presumably, this extension is minor but it seems that the classical techniques provides more insights into what makes the theorem work and computations needed to check the conditions underlying each technique are the same.

<u>Proof of Theorem 2:</u> Define the centered version of Z, $\dot{Z}_{i,t,u} = Z_{i,t,u} - E\,Z_{i,t,u}$ and note that $E\,Z_{i,t,u} = T^{-1}$ for t=u and $= -1/(T(T-1))$ for t≠u. Starting with the second line of (3.1), with some algebra we have

$$n\,(R^2_{AVE} - (T-1)^{-1}) = n\,((n(n-1))^{-1} \Sigma_{i\neq j} \{\Sigma_{t,u} Z_{i,t,u} Z_{j,t,u} - (T-1)^{-1}\}$$

$$= (n-1)^{-1} \Sigma_{i\neq j} \Sigma_{t,u} \dot{Z}_{i,t,u} \dot{Z}_{j,t,u}$$

$$= (n-1)^{-1} \Sigma_{t,u} ((\Sigma_i \dot{Z}_{i,t,u})^2 - \Sigma_i \dot{Z}^2_{i,t,u}). \qquad (3.2)$$

Now, by the weak law of large numbers, $(n-1)^{-1} \Sigma_i \dot{Z}^2_{i,t,u} \to_p \text{Var}(Z_{t,u})$. Thus,

$$(n-1)^{-1} \ \Sigma_{t,u} \ \Sigma_i \dot{Z}^2_{i,t,u} = T \ \mathrm{Var}\ (Z_{1,1}) + T(T-1)\ \mathrm{Var}\ (Z_{1,2}) + o_p(1)$$

$$= (T-2) / (T-1) + o_p(1), \tag{3.3}$$

after some algebra similar to that in Corollary 2. Now define Q_n to be a T x T matrix with $n^{-1/2} \Sigma_i \dot{Z}_{i,t,u}$ to be the element in the t^{th} row and u^{th} column. Further define $X_n = \mathrm{vec}\ (Q_n)$, the T^2 x 1 vector built by stacking the columns of Q_n. Then, using (3.2) and (3.3), we have

$$n\ (R^2_{AVE} - (T-1)^{-1}) = (1 + o(1))\ X'_{n}{}_{n} - (T-2)/(T-1) + o_p(1). \tag{3.4}$$

By the multivariate central limit theorem, $X_n \rightarrow_D X =_D N(0, \Sigma)$, where $\Sigma = E\ (X_1\ X_1')$. Putting this into (3.4), we have

$$n\ (R^2_{AVE} - (T-1)^{-1}) \rightarrow_D X'\ X - (T-2) / (T-1).$$

Thus, we only need to show that $Y =_D X'\ X - (T-2)/(T-1)$. To this end, first compute the eigenvalues of Σ, defined by

$$\lambda\ a = \Sigma\ a. \tag{3.5}$$

Here, 'a' is a T^2 x 1 vector built from a = vec(A) where A is a T x T matrix with a_{tu} is the t^{th} row and u^{th} column. The $(t,u)^{th}$ component of (3.5) is

$$\lambda\ a_{tu} = \Sigma_{r,s}\ a_{rs}\ \mathrm{Cov}\ (Z_{tu}, Z_{rs}) = E\ (\dot{Z}_{tu}\ \Sigma_{rs}\ a_{rs}\ \dot{Z}_{rs}). \tag{3.6}$$

Recall $\Sigma_t \dot{Z}_{tu} = 0$ and note that $\Sigma_t \dot{Z}_{tt} = 0$. Thus, from (3.6), $a_{tu} \equiv 1$ yields $\lambda = 0$ as an eigenvalue. To find non-zero eigenvalues, from (3.6), assuming $\lambda \neq 0$, we have $\Sigma_u a_{ut} = \Sigma_u a_{tu} = \Sigma_u a_{uu} = 0$ for each t. With these constraints, now evaluate (3.6) for the case t=u. This yields

$$\lambda\ a_{tt} = E\ (\dot{Z}_{tt}\ \Sigma_{rs}\ a_{rs}\ \dot{Z}_{rs})$$

$$= a_{tt}\ \mathrm{Cov}\ (\dot{Z}_{11}, \dot{Z}_{11}) + \mathrm{Cov}\ (\dot{Z}_{11}, \dot{Z}_{12})\ (\Sigma_{r,d(r,t)}\ a_{rt} + a_{tr})$$
$$+ \mathrm{Cov}\ (\dot{Z}_{11}, \dot{Z}_{22})\ (\Sigma_{r,d(r,t)}\ a_{rr}) + \mathrm{Cov}\ (\dot{Z}_{11}, \dot{Z}_{23})\ (\Sigma_{s,r,d(r,s,t)}\ a_{rs})$$

$$= a_{tt} \{\mathrm{Cov}(\dot{Z}_{11}, \dot{Z}_{11}) - 2\,\mathrm{Cov}(\dot{Z}_{11}, \dot{Z}_{12}) - \mathrm{Cov}(\dot{Z}_{11}, \dot{Z}_{22}) - 2\,\mathrm{Cov}(\dot{Z}_{11}, \dot{Z}_{23})\}$$

$$= a_{tt}\, a(T). \tag{3.7}$$

Here, $\Sigma_{r,d(r,t)}$ means the sum over r =1, ..., T with distinct r and t. Similarly, $\Sigma_{s,r,d(r,s,t)}$ means the sum over r and s with {r, s, t} distinct. Calculations similar to those in Corollary 2 show that the last equality in (3.7) holds. Now, for $t \neq u$,

$$\lambda\, a_{tu} = E(\dot{Z}_{tu}\, \Sigma_{rs}\, a_{rs}\, \dot{Z}_{rs})$$

$$= \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{12})(a_{tu} + a_{ut}) + \mathrm{Cov}(\dot{Z}_{11}, \dot{Z}_{12})(a_{tt} + a_{uu})$$
$$+ \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{33})(\Sigma_{r,d(r,u,t)}\, a_{rr}) + \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{23})(\Sigma_{r,d(r,u,t)}\, a_{ru} + a_{ur} + a_{tu} + a_{ut})$$
$$+ \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{34})\, \Sigma_{r,s,d(r,s,t,u)}\, a_{rs}$$

$$= (a_{tu} + a_{ut})\{\mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{12}) - \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{23}) + \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{34})\}$$
$$+ (a_{tt} + a_{uu})\{\mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{11}) - \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{33}) - \mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{23}) + 2\,\mathrm{Cov}(\dot{Z}_{12}, \dot{Z}_{34})\}$$

$$= (a_{tu} + a_{ut})/2\;\; b(T) + (a_{tt} + a_{uu})\, c(T). \tag{3.8}$$

Here, $\Sigma_{r,s,d(r,s,t,u)}$ means the sum over r and s with {r, s, t, u} distinct. Again, calculations similar to those in Corollary 2 yield the expressions b(T) and c(T). By (3.8), we have $a_{tu} = a_{ut}$.

For $\lambda \neq 0$, solutions to (3.7) - (3.8) are of two forms. First, if $a_{tt} \neq 0$ for some t, then $\lambda = a(T)$. Second, if $a_{tt} = 0$ for each t, then $\lambda = b(T)$. Thus the eigenvalues of Σ are either 0, a(T) or b(T). It is also useful to recall that the sum of eigenvalues = trace(Σ) = $\Sigma_{t,u}$ Var (Z_{tu}) = (T-2) / (T-1), from (3.3). Now, there are T-1 free diagonal terms a_{tt} since we have the restriction $\Sigma_t\, a_{tt} = 0$. If any of these terms are non-zero, by (3.8), (a(T) - b(T)) $a_{tu} = (a_{tt} + a_{uu})$ c(T) . Thus, fixing the diagonal terms of A determines the remaining elements of the matrix and thus there are T-1 multiplicities at the root λ=a(T). If the diagonal terms of A are all zero, then there are T(T-1) off-diagonal terms to contend with. Due to the symmetry of A noted above and the restrictions $\Sigma_u\, a_{tu} = 0$ for each t, it is easy to verify that there are T(T-1)/2 - T = T(T-3)/2 free off-diagonal terms of A. This then establishes that there are T(T-3)/2 multiplicities at root λ=b(T). To establish that the only other roots are λ=0, we have that (T-1) a(T) + T(T-3)/2 b(T) = $\Sigma_i\, \lambda_i$ = (T-2)/(T-1) = trace(Σ). That is, the sum of eigenvalues equals the trace, as it should.

The remainder of the proof is similar to the proof for R_{AVE}, cf., Hettmansperger (1984, page 184). There exists a $T^2 \times T^2$ orthogonal matrix G such that $G\Sigma G' = D$, a diagonal matrix with the eigenvalues of Σ on the diagonal. With $U = G X =_D N(0, D)$, we have

$$X' X - (T-2)/(T-1) = X' G' G X - (T-2)/(T-1) = U' U - (T-2)/(T-1)$$

$$= \Sigma_i \lambda_i \chi^2_{i,1} - (T-2)/(T-1) \quad = a(T)\, \chi^2_{1,T-1} + b(T)\, \chi^2_{2,T(T-3)/2} - (T-2)/(T-1)$$

$$=_D Y.$$

Here, $\chi^2_{i,1}$ are independent χ^2 random variables with one degree of freedom. Q. E. D.

4. RESIDUAL BASED CORRELATIONS

As an extension of (1.1), now consider the model

$$Y_{it} = \mu_i + g_{it}(\theta) + \sigma_i e_{it} \qquad i=1, ..., n \qquad t=1, ..., T \qquad (4.1)$$

where g_{it} is known up to θ, a p x 1 vector of parameters. This is a nonlinear, heteroscedastic version of the model independently introduced by Hausman and Taylor (1981) and Laird and Ware (1982). Again, μ_i is a nuisance parameter and may be taken to be fixed or random. Let $\hat\theta$ be a root-n consistent estimate of θ, i.e., assume

$$n^{1/2} (\hat\theta - \theta) = O_p(1). \qquad (4.2)$$

Note that I do not require consistency of $\hat\mu_i$. Using these estimates and estimates of μ_i, $\hat\mu_i$, define the residuals $\hat e_{it} = Y_{it} - \hat Y_{it} = \mu_i - \hat\mu_i + g_{it}(\theta) - g_{it}(\hat\theta) + \sigma_i e_{it}$ and the corresponding ranks $\hat R_{it} = \Sigma_{r=1}^{T} I(\hat e_{ir} \le \hat e_{it})$. Analogous to (3.1), the contemporaneous correlation statistic based on residuals considered in this section is

$$\hat R^2_{AVE} = 144 / (n(n-1) (T^3 - T)^2) \Sigma_{i \neq j} (\Sigma_t (\hat R_{it} - m)(\hat R_{jt} - m))^2. \qquad (4.3)$$

Let F(.) and f(.) be the distribution function and probability density function, assumed to exist, of the errors $\{e_{it}\}$. Some additional regularity conditions on the regression function g_{it} are given in assumption A2 below. I am now in a position to state the main result of this section.

THEOREM 3. Assume A1, A2, (4.2), $\inf_i \sigma_i > 0$, and that the probability density function of $\{e_{it}\}$ exists. Then, under the model (4.1),

$$n(R^2_{AVE} - \hat{R}^2_{AVE}) = \Sigma_{t=1}^{T} (W_{1,n}(t) + W_{2,n}(t))\, W_{2,n}(t) + o_p(1)$$

where $W_{1,n}(t) = 2\, n^{-1/2} \Sigma_i \dot{Z}_{i,t,t}$, $W_{2,n}(t) = n^{1/2} (\hat{\theta} - \theta)' (n^{-1} \Sigma_i \nabla M_{i,t}(\theta))\, (12/(T^3-T))$ and $\nabla M_{i,t}(\theta)$ is the gradient vector of $E_\theta\, (R_t(\theta)\text{-}m)^2$ which turns out to be

$$\nabla M_{i,t}(\theta) = 2\, (T-2)\, (\textstyle\int (F(x) - 1/2)\, f(x)\, dF(x)\,)/\sigma_i\, \Sigma_{r=1}^{T} (\partial/\partial\gamma)\, [g_{ir}(\gamma) - g_{it}(\gamma)]_{\gamma=\theta}. \qquad (4.4)$$

Perhaps the most useful aspect of Theorem 2 is to give precise conditions when residuals behave as do i.i.d. error terms in the limiting distribution of R^2_{AVE}. An important special case is when f(.) is symmetric. Here, $\nabla M_{i,t}(\theta) = 0$ and we have

COROLLARY 1. Assume A1, A2, (4.2) and that random variables $\{e_{it}\}$ have a symmetric probability density function f(.). Then,

$$n\, (\hat{R}^2_{AVE} - (T-1)^{-1}\,) \to_D\; Z \;=\; a(T)(\chi^2_{1,T-1} - (T-1)) + b(T)\, (\chi^2_{2,T(T-3)/2} - T(T-3)/2).$$

The method of analysis employs the "estimated parameter" approach advocated by Randles (1984). To this end, I use Randles' (1984, Condition 1.10) regularity condition on the regression function, as follows.

A2. Let $K(\theta)$ be a neighborhood of θ in R^p. For each $\gamma \in K(\theta)$, assume that $g_{it}(\gamma)$ is differentiable uniformly in i and the gradient vector, $\partial/\partial\gamma\, g_{it}(\gamma)$ is uniformly bounded in i.

Implicit in A2, as noted by Randles, is that if g_{it} contains covariates then the covariate space is bounded. Now, for $\gamma \in R^p$, define the perturbed errors, $e_{it}(\gamma) = g_{it}(\theta) - g_{it}(\gamma) + \sigma_i\, e_{it}$, so that $e_{it}(\hat{\theta}) = \hat{e}_{it}$ up to $\mu_i - \hat{\mu}_i$ which, being merely location shifts, do not affect the subsequent analysis. We also need the perturbed ranks

$$R_{it}(\gamma) = \Sigma_{r=1}^{T} I(e_{ir}(\gamma) \le e_{it}(\gamma)) = \Sigma_{r=1}^{T} I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) \qquad (4.5)$$

where $\eta_{i,r,t}(\gamma) = \{g_{it}(\theta) - g_{it}(\gamma) - g_{ir}(\theta) + g_{ir}(\gamma)\} / \sigma_i$. Note that $\eta_{i,r,t}(\gamma) = -\eta_{i,t,r}(\gamma)$. Similar to Section 3, define $Z_{i,t,u}(\gamma) = 12\ (R_{it}(\gamma) - m)(R_{iu}(\gamma) - m) / (T^3 - T)$. The main work in proving Theorem 3 is in

LEMMA 2. Under the assumptions of Theorem 3, for each t, u,

$$n^{-1/2} \left(\Sigma_i Z_{i,t,u}(\hat{\theta}) - Z_{i,t,u}(\theta) \right) - I(t{=}u)\ W_{2,n}(t) \to_p 0.$$

Proof of Lemma 2: The lemma is a consequence of Theorem A.9 of Randles' (1984). To verify the conditions of that theorem, first note that (4.2) and Assumption A2 satisfy Randles' A.4 and Condition 1.10, respectively. Define the kernel of order 1 to be

$$h_i(e_i, \gamma) = Z_{i,t,u}(\gamma).$$

It is easy to see that the kernel is uniformly bounded (satisfying Randles' A.8a). Now, following Randles' notation, for $K(\theta)$ in A2, let $D(\gamma,d) \in R^p$ be a sphere centered at γ with radius such that D is contained in $K(\theta)$. We wish to prove that there exists K>0 so that

$$q_i(d) \equiv E_\theta \sup_{\gamma' \in D} |\ h_i(e_i, \gamma) - h_i(e_i, \gamma')\ | \le K\ d. \qquad (4.6)$$

By A2, there exists $K_1 > 0$ such that $\sup_{\gamma' \in D} |\ g_{it}(\gamma) - g_{it}(\gamma')\ | \le K_1\ d$. This, and since $\inf_i \sigma_i > 0$, implies that there exists $K_2 > 0$ such that $\sup_{\gamma' \in D} |\ \eta_{i,r,t}(\gamma) - \eta_{i,r,t}(\gamma')\ | \le K_2\ d$. From this and (4.5),

$$E_\theta \sup_{\gamma' \in D} |\ R_{it}(\gamma) - R_{it}(\gamma')\ | = E_\theta \sup_{\gamma' \in D} |\ \Sigma_r I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma'))\ |$$

$$\le E_\theta \sup_{\gamma' \in D} \Sigma_{r,d(r,t)} I(\ |e_{ir} - e_{it}\ | \le |\eta_{i,r,t}(\gamma) - \eta_{i,r,t}(\gamma')|\ )$$

$$\le \Sigma_{r,d(r,t)} E_\theta I(\ |e_{ir} - e_{it}\ | \le K_2\ d\ )$$

$$\le (T-2)\ K_2\ d.$$

Recall that $\Sigma_{r,d(r,t)}$ means the sum over r with distinct {r,t}. This and the triangle inequality are sufficient for (4.6).

We only need now calculate the gradient of $E_\theta\, h_i(e_i, \gamma) = M_{i,u}(\gamma)$ and check that it is uniformly (in i) achieved at $\gamma = \theta$. First consider the case t=u. From (4.5), we have

$$R_{it}(\gamma) - m = \Sigma_{r,d(r,t)}\, (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2). \tag{4.7}$$

Thus,

$$M_{i,t}(\gamma) \equiv E_\theta\, (R_{it}(\gamma) - m)^2$$

$$= E_\theta\, \Sigma_{r,d(r,t)}\, (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2)^2$$
$$+ E_\theta\, \Sigma_{r,s,d,(r,s,t)}\, (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2)\, (I(e_{is} \le e_{it} + \eta_{i,s,t}(\gamma)) - 1/2)$$

$$= (T-1)/4 + \Sigma_{r,s,d,(r,s,t)}\, E_\theta\, (F(e + \eta_{i,r,t}(\gamma)) - 1/2)\, (F(e + \eta_{i,s,t}(\gamma)) - 1/2). \tag{4.8}$$

Note that $(\partial/\partial\gamma)\, \eta_{i,r,t}(\gamma) = (\partial/\partial\gamma)\, [g_{ir}(\gamma) - g_{it}(\gamma)]/\sigma_i$. By the chain rule, for fixed e, i, r, s, and u, we have

$$(\partial/\partial\gamma)\, [(F(e + \eta_{i,r,t}(\gamma)) - 1/2)\, (F(e + \eta_{i,s,t}(\gamma)) - 1/2)]_{\gamma=\theta}$$

$$= (F(e) - 1/2)\, f(e) / \sigma_i\, (\partial/\partial\gamma)\, [g_{ir}(\gamma) + g_{is}(\gamma) - 2g_{it}(\gamma)]_{\gamma=\theta}. \tag{4.9}$$

Now take the vector derivative of (4.8). Since r and s are finite, an application of the Bounded Convergence Theorem, with A2 and (4.9), yields $\nabla M_{i,t}(\theta)$ in (4.4). Showing that this is achieved uniformly in i is another application of the Bounded Convergence Theorem.

For $t \ne u$, first recall that $\eta_{i,u,t} = -\eta_{i,t,u}$ and thus $I(e_{iu} \le e_{it} + \eta_{i,u,t}(\gamma)) - 1/2 = 1/2 - I(e_{it} \le e_{iu} + \eta_{i,t,u}(\gamma))$. Thus, from (4.7)

$E_\theta\ (R_{it}(\gamma)\text{-}m)(R_{iu}(\gamma)\text{-}m)$

$$= E_\theta\ [\Sigma_{r,d(r,t,u)}\ (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2) + (I(e_{iu} \le e_{it} + \eta_{i,u,t}(\gamma)) - 1/2)]$$
$$[\Sigma_{r,d,(r,t,u)}\ (I(e_{ir} \le e_{iu} + \eta_{i,r,u}(\gamma)) - 1/2) + (I(e_{it} \le e_{iu} + \eta_{i,t,u}(\gamma)) - 1/2)]$$

$$= E_\theta\ \{[\Sigma_{r,d(r,t,u)}\ (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2)]\ [\Sigma_{r,d,(r,t,u)}\ (I(e_{ir} \le e_{iu} + \eta_{i,r,u}(\gamma)) - 1/2)]$$
$$+ [(I(e_{iu} \le e_{it} + \eta_{i,u,t}(\gamma)) - 1/2)]$$
$$[\Sigma_{r,d,(r,t,u)}\ (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - I(e_{ir} \le e_{iu} + \eta_{i,r,u}(\gamma))]\} - 1/4$$

$$= \Sigma_{r,d(r,t,u)}\ E_\theta\ (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2)\ (I(e_{ir} \le e_{iu} + \eta_{i,r,u}(\gamma)) - 1/2)]$$
$$+ \Sigma_{r,s,d,(r,s,t)}\ E_\theta\ (F(e + \eta_{i,r,t}(\gamma)) - 1/2)\ E_\theta(F(e + \eta_{i,r,u}(\gamma)) - 1/2)$$
$$+ \Sigma_{r,d(r,t,u)}\ E_\theta\ (I(e_{iu} \le e_{it} + \eta_{i,u,t}(\gamma)) - 1/2)[F(e_{it} + \eta_{i,r,t}(\gamma)) - (F(e_{iu} + \eta_{i,r,u}(\gamma))] - 1/4.$$

Since $E_\theta\ (F(e) - 1/2) = 0$, by the chain rule, we have

$(\partial/\partial\gamma)\ [E_\theta\ (R_{it}(\gamma)\text{-}m)(R_{iu}(\gamma)\text{-}m)]_{\gamma=\theta}$

$$= (\partial/\partial\gamma)\ [\Sigma_{r,d(r,t,u)}\ E_\theta\ (I(e_{ir} \le e_{it} + \eta_{i,r,t}(\gamma)) - 1/2)\ (I(e_{ir} \le e_{iu} + \eta_{i,r,u}(\gamma)) - 1/2)$$
$$+ \Sigma_{r,d(r,t,u)}\ E_\theta\ (I(e_{iu} \le e_{it} + \eta_{i,u,t}(\gamma)) - 1/2)\ (F(e_{it} + \eta_{i,r,t}(\gamma)) - (F(e_{iu} + \eta_{i,r,u}(\gamma)))]_{\gamma=\theta}$$

$$= (\partial/\partial\gamma)\ \ [\Sigma_{r,d(r,t,u)}\ \{E_\theta\{(F(e + \eta_{i,t,r}(\gamma)) - 1/2)\ (F(e + \eta_{i,u,r}(\gamma)) - 1/2)$$
$$+ (F(e + \eta_{i,u,t}(\gamma)) - 1/2)\ (F(e + \eta_{i,r,t}(\gamma)) - 1/2)$$
$$+ (F(e + \eta_{i,t,u}(\gamma)) - 1/2)\ (F(e + \eta_{i,r,u}(\gamma)) - 1/2)\}]_{\gamma=\theta}$$

$$= (E_\theta\ (F(e) - 1/2)\ f(e)\ /\ \sigma_i)\ (\partial/\partial\gamma)\ [g_{it}(\gamma) + g_{iu}(\gamma) - 2g_{ir}(\gamma)$$
$$+ g_{iu}(\gamma) + g_{ir}(\gamma) - 2\ g_{it}(\gamma) + g_{it}(\gamma) + g_{ir}(\gamma) - 2\ g_{iu}(\gamma)]_{\gamma=\theta}$$

$$= 0,$$

since $(\partial/\partial\gamma)\ E_\theta\ [(F(e + \eta_{i,u,t}(\gamma)) + F(e + \eta_{i,t,u}(\gamma))]_{\gamma=\theta} = 0$. This establishes that the gradient of $E\ Z_{i,t,u}(\gamma)$ at $\gamma=\theta$ is zero. The fact that it is uniformly achieved in 'i' can be established similarly to the case t=u. Q. E. D.

Proof of Theorem 3: From (4.3) and (3.1), we have

$$n(R^2_{AVE} - \hat{R}^2_{AVE}) = (1 + o(1))\ n^{-1} \Sigma_{t,u}\ \{(\Sigma_i \dot{Z}_{i,t,u})^2 - (\Sigma_i \dot{Z}_{i,t,u}(\hat{\theta}))^2 - (\Sigma_i \dot{Z}_{i,t,u}{}^2 - \dot{Z}_{i,t,u}(\hat{\theta})^2)\}$$

where $\dot{Z}_{i,t,u}(\hat{\theta}) \equiv Z_{i,t,u}(\hat{\theta}) - E\ Z_{i,t,u}(\theta)$. By Lemma 2, $n^{-1} \Sigma_i (\dot{Z}_{i,t,u} - \dot{Z}_{i,t,u}(\hat{\theta})) \rightarrow_p 0$. Similar arguments can be used to show that $n^{-1} \Sigma_i (\dot{Z}_{i,t,u}{}^2 - \dot{Z}_{i,t,u}(\hat{\theta})^2) \rightarrow_p 0$. Thus,

$$n(R^2_{AVE} - \hat{R}^2_{AVE}) = n^{-1} \Sigma_{t,u}\ \{(\Sigma_i \dot{Z}_{i,t,u} - \dot{Z}_{i,t,u}(\hat{\theta}))\ (\Sigma_i \dot{Z}_{i,t,u} + \dot{Z}_{i,t,u}(\hat{\theta}))\} + o_p(1)$$

$$= \Sigma_{tu}\ (I(t{=}u)\ W_{2n}(t) + o_p(1))\ (W_{1n}(t) + W_{2n}(t) + o_p(1)) + o_p(1)$$

which, with A2, (4.2) and Lemma 2, is sufficient for the proof. Q. E. D.

REFERENCES

Bartlett, M. S. (1954), "A note on multiplying factors for various Chi-squared approximations," *Journal of the Royal Statistical Society (B)* 16, 296-298.

Breusch, T. S. and Pagan, A. R. (1980), "The Lagrange multiplier test and its applications to model specification in econometrics," *The Review of Economic Studies* 47, 239-253.

Gregory, G. G. (1977), "Large sample theory for U-statistics and test of fit," *Annals of Statistics* 5, 110-123.

Hausman, J. A. and Taylor, W. E. (1981), "Panel data and unobservable effects," *Econometrica* 49, 1377-1398.

Hettmansperger, T. P. (1984), *Statistical Inference Based on Ranks*, Wiley, New York.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press, New York.

Laird, N. M. and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics* 38, 963-974.

Randles, R. H. (1984), "On test applied to residuals," *Journal of the American Statistical Association* 79, 349-354.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.