TESTING FOR AUTOCORRELATION IN UNEQUALLY REPLICATED FUNCTIONAL MEASUREMENT ERROR MODELS

A. Rasekh*

Department of Statistics, Shahid Chamran University, Ahwaz, Islamic Republic of Iran

Abstract

In the ordinary linear models, regressing the residuals against lagged values has been suggested as an approach to test the hypothesis of zero autocorrelation among residuals. In this paper we extend these results to the both equally and unequally replicated functionally measurement error models. We consider the equally and unequally replicated cases separately, because in the first case the residuals of the means of replicate groups of observations in both *X* and *Y* directions are functions of the same residual while in the second case we have no analogous result and so we have to deal with the residuals in each direction. We derive the asymptotic validity of these tests and we carry out a bootstrap simulation study to determine how well the asymptotic theory of the proposed test works for different size of samples.

1. Introduction

In ordinary linear models, plotting residuals against time has been recommended to assess the model assumptions with respect to independence of successive observations. In particular, such a plot should show no autocorrelation. Furthermore, to test the hypothesis that the errors have zero autocorrelation, we can regress the residuals against lagged values of the regression [1-4]. In this paper we extend these results to the measurement error models in a general case of unequally functional replicated case. The unequally replicated functional measurement error model is defined by

Keywords: Diagnostic; Errors in variables; Lag; Unbalanced replicate

$$
\begin{aligned}
Y_{il} &= y_i + e_{il} && i = 1, \dots, n \\
X_{ij} &= x_i + u_{ij} && j = 1, \dots, r_i \qquad (1) \\
y_i &= \beta' x_i && l = 1, \dots, s_i
\end{aligned}
$$

if $s_i \neq r_i$ for at least one i ($i = 1, \dots, n$). x_i and y_i are the vectors of unobservable fixed values with k and p dimensions, respectively, and β is the matrix of coefficients. For each unobservable x_i and y_i we have more than one observable random vector X_{ij} and Y_{il}. Furthermore, e_{il} and u_{ij} are random errors which have zero mean and covariance matrices Σ_{ee} and Σ_{uu}, respectively. In this case there is no natural pairing among the individual observations X_{ij} and Y_{il}. Thus

* *E-mail: a_rasekh@hotmail.com*

we assume that $\text{cov}(e_{il}, u_{mj}) = 0$ for all *i, j, l*, and *m*. We consider a simple case, $p = 1$ and then extend the results for general cases. We use σ_{ee} as the variance of the error e_{il}.

Suppose that $s_i = r_i = 1$, $i = 1, ..., n$.[5], recommends plot of the residuals against i (i.e. time) to assess the existence of any autocorrelation in the measurement error models. We construct a test based on the regressing residuals against lagged values for both equally and unequally replicated, in the same manner as ordinary linear models. We consider the equally and unequally replicated cases separately because in the first case the residuals of the means of replicate groups of observations in both *X* and *Y* directions are functions of the same residual, while in the second case we do not have any analogous result and so we have to deal with the residuals in each direction [6]. We also present the asymptotic validity of the proposed test.

2. Equally Replicated Case

Suppose that $s_i = r_i$ ($i = 1, ..., n$), then the model (1) will be an equally replicated model. If we define $\nu_i = \overline{Y}_i - \beta' \overline{X}_i$ and $\hat{\nu}_i = \overline{Y}_i - \hat{\beta}' \overline{X}_i$, ($i = 1, ..., n$), in which $\hat{\beta}$ is an estimate of β (see [5]), $\overline{Y}_i$ and $\overline{X}_i$ are means of observations at *i*th level, then it can be shown [6] that in this case the residuals of mean observations in both *X* and *Y* directions are functions of the $\hat{\nu}_i$. Therefore, to assess the existence of any autocorrelation we will use the residual $\hat{\nu}_i$ and for convenience we examine regressing $\hat{\nu}_i$ on the first *t* lagged values of $\hat{\nu}_i$ (i.e. against $\hat{\nu}_{i-1}, \hat{\nu}_{i-2}, ..., \hat{\nu}_{i-t}$ for $i = t+1, ..., n$). It will be clear that the procedure generalises to other situations including non-consecutive lags. First we define $\hat{W}_i = (\hat{\nu}_{i-1}, \hat{\nu}_{i-2}, ..., \hat{\nu}_{i-t})'$, $i = t+1, ..., n$. Then the regression coefficient of $\hat{\nu}_i$ on $\hat{W}_i$ is given by

$$\hat{\gamma} = \left[\sum_{i=t+1}^{n} (\hat{W}_i - \overline{\hat{W}})(\hat{W}_i - \overline{\hat{W}})' \right]^{-1} \sum_{i=t+1}^{n} (\hat{W}_i - \overline{\hat{W}}) \hat{\nu}_i \qquad (2)$$

where $\overline{\hat{W}} = (n-t)^{-1} \sum_{i=t+1}^{n} \hat{W}_i$. In the following theorem we derive the convergence properties of the $\hat{\gamma}$ as a basis for testing zero autocorrelation among errors of the model.

Theorem 2.1. Let $\hat{\gamma}$ be the vector of estimated coefficients from regressing $\hat{\nu}_i$ on $\hat{W}_i$, defined by (2). Then $\hat{\gamma} = r\sigma_{\nu\nu}^{-1} n^{-1} \sum_{i=t+1}^{n} W_i \nu_i + o_p(n^{-\frac{1}{2}})$ and $n^{\frac{1}{2}} \hat{\gamma}$ converges in Law to the standard normal distribution as *n* tends to infinity.

Proof. We define $\hat{x}_i = \overline{X}_i + \hat{\sigma}_{\nu\nu}^{-1} \hat{\Sigma}_{uu} \hat{\beta} \hat{\nu}_i$ as an estimate of x_i [5], then it is not difficult to show that

$$\begin{aligned} \hat{\nu}_i &= (1 + C_n)\nu_i - (\hat{\beta} - \beta)' \ddot{x}_i \\ \hat{x}_i &= (1 + C_n)\ddot{x}_i + d_n \nu_i \end{aligned} \qquad i = 1, ..., n \qquad (3)$$

in which $\ddot{x}_i = \overline{X}_i + \sigma_{\nu\nu}^{-1} \Sigma_{uu} \beta \nu_i$, $C_n = \text{O}_p(n^{-\frac{1}{2}})$ and $d_n = \text{O}_p(n^{-\frac{1}{2}})$. We have

$$\hat{\nu}_i - \overline{\hat{\nu}} = (1 + C_n)(\nu_i - \overline{\nu}) - E_n(\ddot{x}_i - \overline{\ddot{x}}) \quad i = 1, ..., n \qquad (4)$$

where $E_n = (\hat{\beta} - \beta)' = \text{O}_p(n^{-\frac{1}{2}})$ and $\overline{\hat{\nu}} = n^{-1} \sum_{i=1}^{n} \hat{\nu}_i$. This expression holds for lagged values and derivations of mean when the mean $\overline{\hat{\nu}}$ is calculated over only $(n-t)$ groups of the observations for fixed t. For example, expression (3) holds for

$$\hat{\nu}_{i-k} - \overline{\hat{\nu}}_{(-k)} = (1 + C_n)\left[\nu_{i-k} - \overline{\nu}_{(-k)}\right] - E_n\left[\ddot{x}_{i-k} - \overline{\ddot{x}}_{(-k)}\right] \qquad (5)$$

where $\overline{\hat{\nu}}_{(-k)} = (n-t)^{-1} \sum_{i=t+1}^{n} \hat{\nu}_{i-k}$, $k = 1, ..., t$, and with the analogous definition for $\overline{\nu}_{(-k)}$ and $\overline{\ddot{x}}_{(-k)}$. To assess the asymptotic distribution of the $\hat{\gamma}$, we first examine expressions of the form,

$$n^{-1} \sum_{i=t+1}^{n} \left[\hat{\nu}_{i-k} - \overline{\hat{\nu}}_{(-k)}\right](\hat{\nu}_i - \overline{\hat{\nu}}) \quad k = 1, ..., t \qquad (6)$$

which are elements of the $n^{-1} \sum_{i=t+1}^{n} (\hat{W}_i - \overline{\hat{W}}) \hat{\nu}_i$. Substituting expressions of the $(\hat{\nu}_i - \overline{\hat{\nu}})$ and $(\hat{\nu}_{i-k} - \overline{\hat{\nu}}_{(-k)})$ from (4) and (5) into (6) we can show that

$$n^{-1}\sum_{i=t+1}^{n}\left[\hat{\nu}_{i-k}-\overline{\hat{\nu}}_{(-k)}\right](\hat{\nu}_i-\overline{\hat{\nu}})=\sum_{i=1}^{4}\Pi_i \qquad (7)$$

where Π_i 's $i=1,\ldots,4$ are

$$\Pi_1=(1+C_n)^2 n^{-1}\sum_{i=t+1}^{n}\left[\nu_{i-k}-\overline{\nu}_{(-k)}\right](\nu_i-\overline{\nu})$$

$$\Pi_2=(1+C_n)n^{-1}\sum_{i=t+1}^{n}\left[\nu_{i-k}-\overline{\nu}_{(-k)}\right]\left[E_n(\ddot{x}_i-\overline{\ddot{x}})\right] \qquad (8)$$

$$\Pi_3=(1+C_n)n^{-1}\sum_{i=t+1}^{n}(\nu_i-\overline{\nu})\left[E_n(\ddot{x}_{i-k}-\overline{\ddot{x}}_{(-k)})\right]$$

$$\Pi_4=n^{-1}\sum_{i=t+1}^{n}\left[E_n(\ddot{x}_{i-k}-\overline{\ddot{x}}_{(-k)})\right]'\left[E_n(\ddot{x}_i-\overline{\ddot{x}})\right].$$

Using Hölder's inequality repeatedly for the Π_4, we conclude that $\Pi_4=O_p(n^{-1})$, while the independence of ν_i and x_i 's implies $\Pi_2=o_p(n^{-\frac{1}{2}})$ and $\Pi_3=o_p(n^{-\frac{1}{2}})$. Also for Π_1 we have

$$\Pi_1=n^{-1}\sum_{i=t+1}^{n}\nu_{i-k}\nu_i+o_p(n^{-\frac{1}{2}}) \qquad (9)$$

However, as n tends to infinity, (5) implies that

$$(n-t)^{-1}\sum_{i=t+1}^{n}(\hat{W}_i-\overline{\hat{W}})(\hat{W}_i-\overline{\hat{W}})'$$
$$=r^{-1}\sigma_{\nu\nu}\mathrm{I}_t+O_p(n^{-\frac{1}{2}}) \qquad (10)$$

where $r^{-1}\sigma_{\nu\nu}$ is variance of the ν_i. Combining results from (2), (7), (9) and (10) we have $\hat{\gamma}=r\sigma_{\nu\nu}^{-1}n^{-1}\sum_{i=t+1}^{n}W_i\nu_i+o_p(n^{-\frac{1}{2}})$. Thus, it follows from Theorem (8.2.1) of [7], that

$$n^{\frac{1}{2}}\hat{\gamma}\xrightarrow{L}N(0,\mathrm{I}_t) \quad \text{as} \quad n\to\infty. \qquad (11)$$

Lemma 2.2. Let $SSR1=(n-t)^{-1}\sum_{i=t+1}^{n}\left[(\hat{\nu}_i-\overline{\hat{\nu}})-\hat{\gamma}'(\hat{W}_i-\overline{\hat{W}})\right]^2$ be the mean squares residuals from regressing $\hat{\nu}_i$ on $\hat{W}_i$. $SSR1$ will converge to $r^{-1}\sigma_{\nu\nu}$ as n tends to infinity.

Proof. From (11) we conclude that $\hat{\gamma}=O_p(n^{-\frac{1}{2}})$ (see Theorem 14.4-2 of [8]). Therefore, using expression (10), the mean squares residuals is equal to

$$(n-t)^{-1}\sum_{i=t+1}^{n}(\hat{\nu}_i-\overline{\hat{\nu}})^2+o_p(1) \qquad (12)$$

and will converge to the $r^{-1}\sigma_{\nu\nu}$ as n tends to infinity.

Expressions (10), (11) and (12) imply that the common t statistic for testing each element of γ ($\gamma_j=0$) and F statistic for simultaneously testing of the elements of γ converge to the standard normal and Chi-square distributions, respectively. In large samples, such tests can be used to test hypothesis of zero autocorrelation of the ν_i 's.

3. Unequally Replicated Case

In the unequally replicated case residuals of the means of observations in both X and Y directions are not solely depend on $\hat{\nu}_i$ [6]. Therefore, we have to consider existence of any autocorrelation among the errors in both directions. However, to minimise the length of paper, we only look at this problem in Y direction and we define $\hat{e}_i=\overline{Y}_i-\hat{\beta}'\hat{x}_i$, $i=1,\ldots,n$, as the residual in this direction. It is not difficult to show that $\hat{e}_i=\ddot{e}_i+O_p(n^{-\frac{1}{2}})$, in which $\ddot{e}_i=(1+\beta'\delta_i)\nu_i$, $\delta_i=-r_i^{-1}\sigma_{\nu_i\nu_i}^{-1}\Sigma_{uu}\beta$ and $\sigma_{\nu_i\nu_i}=n_i^{-1}\sigma_{ee}+r_i^{-1}\beta'\Sigma_{uu}\beta$. Furthermore, the asymptotic variance of the $\hat{e}_i$, is $\sigma_{\ddot{e}_i\ddot{e}_i}=(1+\beta'\delta_i)^2\sigma_{\nu_i\nu_i}$. Clearly $\hat{e}_i$, $i=1,\ldots,n$, have different asymptotic variances. Therefore, to avoid using heteroscedastic regression of the residuals, we define $\hat{e}_i^*=\sigma_{\ddot{e}_i\ddot{e}_i}^{-\frac{1}{2}}\hat{e}_i$, $i=1,\ldots,n$, as the studentised residuals and we use $\hat{e}_i^*$ instead of $\hat{e}_i$. Thus, we have

$$\hat{e}_i^*=\sigma_{\ddot{e}_i\ddot{e}_i}^{-\frac{1}{2}}\ddot{e}_i+O_p(n^{-\frac{1}{2}}) \qquad i=1,\ldots,n \qquad (13)$$
$$=\ddot{e}_i^*+O_p(n^{-\frac{1}{2}}).$$

We define $\hat{\Xi}_i = (\hat{e}^*_{i-1}, \hat{e}^*_{i-2}, ..., \hat{e}^*_{i-t})'$, $i = t+1, ..., n$, as the vector of first t-lags. If we regress $\hat{e}^*_i$ on $\hat{\Xi}_i$, then the regression coefficient can be given by

$$\hat{\gamma}^* = \left[\sum_{i=t+1}^{n} (\hat{\Xi}_i - \bar{\hat{\Xi}})(\hat{\Xi}_i - \bar{\hat{\Xi}})'\right]^{-1} \sum_{i=t+1}^{n} (\hat{\Xi}_i - \bar{\hat{\Xi}})\hat{e}^*_i \qquad (14)$$

in which $\bar{\hat{\Xi}} = (n-t)^{-1}\sum_{i=t+1}^{n} \hat{\Xi}_i$. We derive the convergence properties of the $\hat{\gamma}^*$ in the following theorem.

Theorem 3.1. Let $\hat{\gamma}^*$ be the vector of estimated coefficients from regressing $\hat{e}^*_i$ on $\hat{\Xi}_i$ defined by (14). Then $\hat{\gamma}^* = n^{-1}\sum_{i=t+1}^{n} \ddot{\Xi}^*_i \ddot{e}^*_i + o_p(n^{-\frac{1}{2}})$ and $n^{\frac{1}{2}}\hat{\gamma}^*$ converges in Law to the standard normal distribution as n tends to infinity.

Proof. From (13) we have $\hat{e}^*_i - \bar{\hat{e}}^* = \ddot{e}^*_i - \bar{\ddot{e}}^* + O_p(n^{-\frac{1}{2}})$. This expression also holds for the lagged values and for derivations from the mean when the mean $\bar{\hat{e}}^*$ is calculated over only $(n-t)$ groups of the observations for fixed t and we have $\hat{e}^*_{i-k} - \bar{\hat{e}}^*_{(-k)} = \ddot{e}^*_{i-k} - \bar{\ddot{e}}^*_{(-k)} + O_p(n^{-\frac{1}{2}})$, where $\bar{\hat{e}}^*_{(-k)} = (n-t)^{-1}\sum_{i=t+1}^{n} \hat{e}^*_{(i-k)}$ and with the same definition for the $\bar{\ddot{e}}^*_{(-k)}$. To assess the asymptotic distribution of the $\hat{\gamma}^*$, we examine expressions $n^{-1}\sum_{i=t+1}^{n}\left[\hat{e}^*_{i-k} - \bar{\hat{e}}^*_{(-k)}\right](\hat{e}^*_i - \bar{\hat{e}}^*)$ which are elements of the $n^{-1}\sum_{i=t+1}^{n}(\hat{\Xi}_i - \bar{\hat{\Xi}})\hat{e}^*_i$. We have

$$n^{-1}\sum_{i=t+1}^{n}\left[\hat{e}^*_{i-k} - \bar{\hat{e}}^*_{(-k)}\right](\hat{e}^*_i - \bar{\hat{e}}^*)$$

$$= n^{-1}\sum_{i=t+1}^{n}\left[\ddot{e}^*_{i-k} - \bar{\ddot{e}}^*_{(-k)}\right](\ddot{e}^*_i - \bar{\ddot{e}}^*) + O_p(n^{-\frac{1}{2}}) \qquad (15)$$

$$= n^{-1}\sum_{i=t+1}^{n}\ddot{e}^*_{i-k}\ddot{e}^*_i + o_p(n^{-\frac{1}{2}})$$

since $n^{-1}\bar{\ddot{e}}^*_{(-k)}\sum_{i=t+1}^{n}\ddot{e}^*_i = o_p(n^{-\frac{1}{2}})$. On the other hand, as n tends to infinity, we have

$$(n-t)^{-1}\sum_{i=t+1}^{n}(\hat{\Xi}_i - \bar{\hat{\Xi}})(\hat{\Xi}_i - \bar{\hat{\Xi}})' = \mathrm{I} + O_p(n^{-\frac{1}{2}}) \qquad (16)$$

Combining results from (15) and (16) we obtain

$$\hat{\gamma}^* = n^{-1}\sum_{i=t+1}^{n}\ddot{\Xi}^*_i\ddot{e}^*_i + o_p(n^{-\frac{1}{2}}) , \qquad (17)$$

in which $\ddot{\Xi}^*_i = (\ddot{e}^*_{i-1}, \ddot{e}^*_{i-2}, ..., \ddot{e}^*_{i-t})'$. Thus, it follows from Theorems (8.2.1) and (8.2.2) of [7], that

$$n^{\frac{1}{2}}\hat{\gamma}^* \xrightarrow{L} N(0, \mathrm{I}_t) \quad \text{as} \quad n \to \infty . \square \qquad (18)$$

Lemma 3.2. Let $SSR2 = (n-t)^{-1}\sum_{i=t+1}^{n}\left[(\hat{e}^*_i - \bar{\hat{e}}^*) - \hat{\gamma}^{*\prime}(\hat{\Xi}_i - \bar{\hat{\Xi}})\right]^2$ be the mean square residuals from regressing $\hat{e}^*_i$ on $\hat{\Xi}_i$. $SSR2$ will converge to 1 as n tends to infinity.

Proof. From (18) we have $\hat{\gamma}^* = O_p(n^{-\frac{1}{2}})$ (see Theorem 14.4-2 of [8]) which implies that the mean square residuals is equal to

$$(n-t)^{-1}\sum_{i=t+1}^{n}(\hat{e}^*_i - \bar{\hat{e}}^*)^2 = (n-t)^{-1}\sum_{i=t+1}^{n}(\ddot{e}^*_i - \bar{\ddot{e}}^*)^2 + o_p(1)$$

and $(n-t)^{-1}\sum_{i=t+1}^{n}(\ddot{e}^*_i - \bar{\ddot{e}}^*)^2$ will converge to 1 as n tends to infinity.$\square$

Expressions (16), (17) and (18) imply that the common t statistic for testing each element of γ^* ($\gamma^*_j = 0$) and F statistic for simultaneously testing of elements of γ^* converge to the standard normal and Chi-square distribution, respectively. Thus, in large samples, such tests can be used to test hypothesis of zero autocorrelation among the errors in Y direction. In practice we can use an estimate of the $\sigma_{\ddot{e}_i\ddot{e}_i}$ which is $\hat{\sigma}_{\ddot{e}_i\ddot{e}_i} = (1+\hat{\beta}'\hat{\delta}_i)^2\hat{\sigma}_{\nu_i\nu_i}$, $i=1,...,n$, in the definition of the $\hat{e}_i^*$ instead of $\sigma_{\ddot{e}_i\ddot{e}_i}$ which is unknown.

4. Multivariate Extensions

In previous sections we concentrated on the univariate model in which Y_{ij} is a random variable. However, the procedure for testing autocorrelation can be easily extended to the multivariate models in which Y_{ij} is a random vector. We define $\hat{\nu}_i = (\hat{\nu}_{i1},...,\hat{\nu}_{ip})'$ and $\hat{e}_i^* = (\hat{e}_{i1}^*,...,\hat{e}_{ip}^*)'$ as the random vectors of ith residual for equally and unequally replicated cases, respectively. We have $\hat{e}_i^* = \hat{\Sigma}_{\ddot{e}_i\ddot{e}_i}^{-\frac{1}{2}}\hat{e}_i$, in which $\Sigma_{\ddot{e}_i\ddot{e}_i} = (\mathrm{I}+\beta'\delta_i)\Sigma_{\nu_i\nu_i}(\mathrm{I}+\beta'\delta_i)'$, $\delta_i = -r_i^{-1}\Sigma_{uu}\beta\Sigma_{\nu_i\nu_i}^{-1}$ and $\Sigma_{\nu_i\nu_i} = n_i^{-1}\Sigma_{ee} + r_i^{-1}\beta'\Sigma_{uu}\beta$.

To investigate existence of any autocorrelation among errors of the model, we examine elements of $\hat{\nu}_i$ or $\hat{e}_i^*$ and we regress $\hat{\nu}_{ij}$ or $\hat{e}_{ij}^*$ ($j=1,...,p$) versus their first *t-lags*. Then we can test for zero regression coefficient in each case. The asymptotic validity of the t and F statistics can be derived using exactly the same arguments given in sections (2) and (3) and so we are not going to go through further details.

5. Parametric Bootstrap Simulation Study

We derived the theoretical justifications of using statistical techniques for testing autocorrelation analogue to those given in ordinary linear models. These results are only hold if n tends to infinity. However, in practice there are situations in which sample size is medium or small. Therefore, the aim of parametric bootstrap simulation study here is to determine how well the asymptotic theory of the proposed test works for the different sample sizes. The study is constructed so as to simulate an actual data set as much as possible. In order to do this, we simulated data in accordance with a set of real data. First we introduce this data set and then we perform the simulation study.

5.1. Data Analysis

The data set in this example arises from a series of experiments in 1985 at the Animal Research Institute (Werribee), Victoria, Australia and is known as "digestibility data". The objective of experiments was to assess a new and more convenient method of assaying the digestibility of various diets fed to animals. The new method (the "nylon bag" technique) involved putting the food in a loose meshed nylon and weighing it before and after digestion. The old or "conventional assay" entailed sacrifice of the animal.

The collected data set contains the digestibility values of the thirty-nine diets fed to animals as determined by conventional and nylon bag assays. Of the thirty-nine diets, eighteen are pellet and grain and the remaining are from another diets (The original data set exists from author on request). The question of interest is to determine the relationship between digestibility values as determined by the two kinds of assays. In each diet there are different numbers of replications for the conventional assay (X_{ij}) and nylon bag assay (Y_{il}) and so we have unequal number of replicated data at each level [9]. Preliminary analysis of this data set shows that it is reasonable to investigate linear relationship between the nylon bag and the conventional assay and so we fitted the functional measurement error model. The estimators of the parameters are $\hat{\beta}_0 = -5.009$, $\hat{\beta}_1 = 1.005$, $\hat{\sigma}_{uu} = 2.89$ and $\hat{\sigma}_{ee} = 17.56$. Furthermore, the value of the F-statistic for testing a zero regression coefficient from regressing $\hat{e}_i^*$'s on one lagged value is given by $F = 2.986$ which is significant at the 10% level (but not at more stringent levels). This implies that there is some evidence of the existence autocorrelation in the Y direction.

Furthermore, as a small data set, we also considered a subset of eighteen groups of the digestibility data, which are pelleted and grains diets. The estimators of the parameters of the model from fitting a functional measurement error model to this subset are $\hat{\beta}_0 = 5.396$, $\hat{\beta}_1 = 0.871$, $\hat{\sigma}_{ee} = 23.393$ and $\hat{\sigma}_{uu} = 2.042$. The value of the F-statistic for testing a zero regression coefficient from regressing $\hat{e}_i^*$'s on one lagged value is given by $F = 3.982$ which is significant at the 10% level and thus gives some slight evidence of existence autocorrelation among the Y values. In the next section we use this data set to conduct our simulation study.

5.2. Simulation Results

In this section we shall use the digestibility data set to simulate data accordance to the model (1). The group number is 39, which is relatively medium, and the number of replications in each group is exactly the same as those for the original digestibility data. At each step of the bootstrap replication we generated data for the model

$$\begin{aligned} Y_{il} &= \hat{\beta}_0 + \hat{\beta}_1\hat{x}_i + e_{il} && i=1,...,39 \\ & && j=1,...,r_i \\ X_{ij} &= \hat{x}_i + u_{ij} && l=1,...,n_i \end{aligned}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of the β_0 and β_1 and the values $\hat{x}_i$, $i = 1, ..., 39$, are the estimated values of the x_i, $i = 1, ..., 39$, based on the original data. In addition we assumed that e_{il} and u_{ij} have normal distribution with zero mean and variances as $\hat{\sigma}_{ee}$ and $\hat{\sigma}_{uu}$ based on the original data. We simulated a total of 1000 data sets and repeated the procedure of testing first order autocorrelation for the simulated data and we calculated the value of the *F*-statistic in each replication and compared it with the values of *F*-distribution for different critical regions 0.10, 0.05, 0.025 and 0.01.

Table (1) summarises the results of the bootstrap study. The bottom row of the table gives the proportions of time that the calculated *F*-statistic fell beyond the critical values for different critical regions. This table presents relatively good evidence about the behaviour of the test. For the critical regions 0.10, 0.05 and 0.025 the calculated proportions are higher than the theoretical values, while we remember that we have rejected the assumption of no autocorrelation for the original data at the 0.10 level. The calculated proportion for the 0.01 is less than the theoretical value, which indicates that the test procedure will be conservative for the small critical regions.

Table 1. Simulation results of the first order autocorrelation test

Critical region	0.10	0.05	0.025	0.01
Calculated proportion	0.130	0.070	0.035	0.008

Table 2. Simulation results of the first order autocorrelation test using pelleted and grains diets data

Critical region	0.10	0.05	0.025	0.01
Calculated proportion	0.085	0.048	0.027	0.008

A second bootstrap study was conducted with a subset of eighteen groups of the digestibility data, which are pelleted and grains diets. The aim of the second study is to determine the effect of the small sample sizes on the test procedure. While the other aspects of the bootstrap process were unchanged, we simulated 1000 data sets with a procedure exactly the same as before.

The results of the second study are summarised in Table (2). From this table we can see that, despite the possible rejection of the hypothesis of no autocorrelation for the original data at the level of 0.10, the calculated proportions are less than the theoretical values, which shows that for the small sample sizes the test procedure is conservative. Finally, while our simulation study is restricted to the first order autocorrelation, however, we could also extend our study to the higher orders and to see how the procedure works for small sample sizes.

Acknowledgements

I am indebted to the anonymous referee for a number of useful comments and suggestions, which help me to improve my paper.

References

1. Durbin, J. and Watson, G. S., Testing for serial correlation in least squares regression III. *Biometrika*, **58**: 1-19 (1971).
2. Godfery, L., Testing for Multiplicative Heterosce-dasticity. *J. Econometrics*, 227-36 (1978).
3. Farebrother, R., The Durbin-Watson Test for Serial Correlation when there is no intercept in the Regression. *Econometrica*, 1553-63 (1980).
4. Engle, R. F., *Wald, Likelihood-Ratio and Lagrangian Multiplier Tests in Econometrics*. In: Z. Griliches and M. D. Intrilligator (Eds.), Annals of Econometrics, North-Holland, New York, (1984).
5. Fuller, W. A., *Measurement Error Models*. Wiley, New York, (1987).
6. Rasekh, A. R., Residuals in unequally replicated functional measurement error models. (Submitted to Communication is Statistics, Theory & Methods) (1998).
7. Fuller, W. A., *Introduction to Statistical Time Series*. Wiley, New York, (1976).
8. Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W., *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, (1975).
9. Fieller, N. R. J., Analysis and comparison of unbalanced linear function relationships obtained from various forms of diets and digestibility assays, *Statistical Consulting Centre, Research Report No. 25, University of Melbourne*, (1985).