Testing for Autocorrelation in unequally replicated functional measurement error

Abdolrahman Rasekh

Department of Statistics, Shahid Chamran University, Ahwaz, Iran
A_Rasekh@Hotmail.com

1. Introduction

In ordinary linear models, regressing residuals against lagged values has been suggested to test the hypothesis of zero autocorrelation among the errors (Durbin and Watson, 1971). 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

$$Y_{il} = y_i + e_{il} \qquad i = 1, \ldots, n$$
$$X_{ij} = x_i + u_{ij} \qquad j = 1, \ldots, r_i$$
$$y_i = b' x_i \qquad l = 1, \ldots, s_i$$

if $s_i \neq r_i$ for at least one i. x_i and y_i are the vectors of unobservable fixed values with k and p dimensions, respectively, and b 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. We start with the simple case, $p = 1$ and we use s_{ee} as the variance of the error e_{il}. Then we extend the results to the general case. We derive a test based on regressing residuals against lagged values, in the same manner as ordinary linear models and present the asymptotic validity of the proposed test.

2. Test in unequally replicated case

In measurement error models we have errors and so residuals in both X and Y directions. Therefore, we have to consider existence any autocorrelation among the errors in both directions. However, to minimize the length of paper, we only look at this problem in Y direction. We define $\hat{e}_i = \bar{Y}_i - \hat{b}'\hat{x}_i$, as the residual in this direction in which $\hat{b}$ is an estimate of b, $\hat{x}_i = \bar{X}_i - s^{-1}_{n_i n_i} \hat{\Sigma}_{u_i n_i} \hat{n}_i$, $n_i = \bar{Y}_i - b'\bar{X}_i$, $s_{n_i n_i} = s_i^{-1} s_{ee} + r_i^{-1} b' \Sigma_{uu} b$ and $\Sigma_{u_i v_i} = -r_i^{-1}(b'\Sigma_{uu})$ (see Fuller, 1987). Then it is not difficult to show that $\hat{e}_i = \tilde{e}_i + O_p(n^{-\frac{1}{2}})$, in which $\tilde{e}_i = (1 + b' d_i) n_i$, $d_i = -r_i^{-1} s^{-1}_{n_i n_i} \Sigma_{uu} b$. Furthermore, the asymptotic variance of the $\hat{e}_i$, is $s_{\tilde{e}_i \tilde{e}_i} = (1 + b' d_i)^2 s_{n_i n_i}$. Clearly $\hat{e}_i$s have different asymptotic variances. Therefore, to avoid using heteroscedastic regression of the residuals, we define $\hat{e}_i^* = s^{-\frac{1}{2}}_{\tilde{e}_i \tilde{e}_i} \hat{e}_i$, as the studentised residuals and we use $\hat{e}_i^*$ instead of $\hat{e}_i$. Thus, we have

$$\text{(1)} \qquad \hat{e}_i^* = s^{-\frac{1}{2}}_{\tilde{e}_i \tilde{e}_i} \tilde{e}_i + O_p(n^{-\frac{1}{2}}) = \tilde{e}_i^* + O_p(n^{-\frac{1}{2}}) \qquad i = 1, \ldots, n$$

We define $\hat{\Xi}_i = (\hat{e}^*_{i-1}, \hat{e}^*_{i-2}, \ldots, \hat{e}^*_{i-t})'$, $i = t+1, \ldots, n$, 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

$$\text{(2)} \qquad \hat{g}^* = \left[\sum_{i=t+1}^{n} (\hat{\Xi}_i - \bar{\hat{\Xi}})(\hat{\Xi}_i - \bar{\hat{\Xi}})'\right]^{-1} \sum_{i=t+1}^{n} (\hat{\Xi}_i - \bar{\hat{\Xi}}) \hat{e}_i^*$$

where $\bar{\hat{\Xi}} = (n-t)^{-1} \sum_{i=t+1}^{n} \hat{\Xi}_i$. Next we derive the convergence properties of the $\hat{g}^*$.

Theorem 2.1: Let $\tilde{g}^*$ be the vector defined by (2). Then $\hat{g}^* = n^{-1}\sum_{i=t+1}^{n}\tilde{\Xi}_i^*\tilde{e}_i^* + o_p(n^{-\frac{1}{2}})$ and $n^{\frac{1}{2}}\tilde{g}^*$ converges in Law to the standard normal distribution as n tends to infinity.

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

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

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

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

Combining results from (3) and (4) we obtain

(5) $$\tilde{g}^* = n^{-1}\sum_{i=t+1}^{n}\tilde{\Xi}_i^*\tilde{e}_i^* + o_p(n^{-\frac{1}{2}}) ,$$

where $\tilde{\Xi}_i^* = (\tilde{e}_{i-1}^*, \tilde{e}_{i-2}^*, ..., \tilde{e}_{i-t}^*)'$. Thus, it follows from Fuller (1976, theorems, 8.2.1 and 8.2.2) that

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

Lemma 2.1: Let $SSR = (n-t)^{-1}\sum_{i=t+1}^{n}\left[(\hat{e}_i^* - \bar{\hat{e}}^*) - \hat{g}^{*\prime}(\hat{\Xi}_i - \bar{\hat{\Xi}})\right]^2$ be the mean square residuals from regressing $\tilde{e}_i^*$ on $\hat{\Xi}_i$. SSR will converge to 1 as n tends to infinity.

Proof: From (6) we have $\hat{g}^* = O_p(n^{-\frac{1}{2}})$ (see Bishop, Fienberg and Holland, 1975 theorem 14.4-2) which implies that SSR is $(n-t)^{-1}\sum_{i=t+1}^{n}(\hat{e}_i^* - \bar{\hat{e}}^*)^2 = (n-t)^{-1}\sum_{i=t+1}^{n}(\tilde{e}_i^* - \bar{\tilde{e}}^*)^2 + o_p(1)$ and $(n-t)^{-1}\sum_{i=t+1}^{n}(\tilde{e}_i^* - \bar{\tilde{e}}^*)^2$ converges to 1 as n tends to infinity.•

Expressions (4), (5) and (6) imply that the common t statistic for testing each element of g^* ($g_j^* = 0$) and F statistic for simultaneously testing of elements of g^* 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.

The procedure for testing autocorrelation can be easily extended to the cases in which Y_{ij} is a random vector. We define $\tilde{e}_i^* = (\tilde{e}_{i1}^*, ..., \tilde{e}_{ip}^*)'$ as the vector of ith residual and we examine elements of $\tilde{e}_i^*$ and we regress $\hat{e}_{ij}^*$ $(j = 1, ..., p)$ versus their first *t-lags*. Then we can test for zero regression coefficients in each case. The asymptotic validity of the t and F statistics can be derived using exactly the same arguments given above.

References

Durbin, J. & Watson, G. S. (1971), Testing for serial correlation in least squares regression III, *Biometrika*, **58**, 1-19.

Fuller, W. A.(1987), *Measurement error models*. Wiley, New York .

Fuller, W. A. (1976), *Introduction to statistical time series*. Wiley, New York.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, The MIT press.