ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 48 — 1997

Monte Carlo Results on Pure and Pretest Estimators of an Error Component Model with Autocorrelated Disturbances

Badi H. BALTAGI, Qi LI*

ABSTRACT. – A Monte Carlo study is used to compare the finite sample relative efficiency of a number of pure and pre-test estimators for an error component model with first-order autocorrelated remainder disturbances.

Quelques résultats de la méthode de Monte Carlo sur des estimateurs pures et pre-test dans le cadre d'un modèle à erreurs composées avec autocorrélation des résidus.

RÉSUMÉ. – La méthode de Monte Carlo est utilisée pour comparer l'efficacité relative à distance finie de différents estimateurs pures et pre-test dans le cadre d'un modèle à erreurs composées avec autocorrélation au premier ordre des résidus.

* B. H. BALTAGI : Texas A&M University; Q. LI University of Guelph, Ontario. An earlier version of this paper was presented at the fifth conference on panel data held at the University of Paris, Val de Marne, June, 1994. We would like to thank the editor and two referees for their helpful comments and suggestions.

1 Introduction

GRIFFITHS and BEESLEY [1984] questioned the usual practice of estimating AR(1) models after a Durbin-Watson test especially if the model is misspecified. Monte Carlo experiments were performed to compare the relative efficiency of various pure and pre-test estimators for a time-series regression when the disturbances follow an AR(1) or MA(1) process. This study considers a similar problem for the panel data error component regression with autocorrelated remainder disturbances. Some of the main differences are that individual effects are now present and brute force GLS cannot be performed when the panel size is huge. BALTAGI and LI [1991, 1994] derived a simple FULLER and BATTESE [1974] type transformation to perform GLS as weighted least squares when the remainder disturbances follow an AR(p) or an MA(q) process. Also, BALTAGI and LI [1995] derived a simple Lagrange Multiplier test to test for first order serial correlation given the presence of individual effects and another simple test to distinguish between the AR(1) and MA(1) remainder error process. This study performs Monte Carlo Experiments to compare the small sample properties of eight alternative estimators, when the remainder disturbances are generated by an AR(1) or an MA(1) process. These estimators are: (1) Ordinary Least Squares (OLS); (2) Least Squares with Dummy Variables or the 'Within' estimator which assumes that the individual effects are fixed parameters to be estimated; (3) Conventional GLS which ignores the serial correlation in the remainder disturbances but accounts for the random error components structure. This is denoted by CGLS; (4) GLS assuming random error components with the remainder disturbances following an MA(1) process. This is denoted by GLSM; (5) GLS assuming random error components with the remainder disturbances following an AR(1) process. This is denoted by GLSA; (6) A pre-test estimator which is based on the results of two tests. This is denoted by PRE. The first test is for serial correlation given the existence of individual effects. If the null is not rejected, this estimator reduces to conventional GLS. In case serial correlation is found, a second test is performed to distinguish between the AR(1) and MA(1) process and GLSA or GLSM is performed. (7) A Generalized Method of Moments (GMM) estimator, where the error component structure of the disturbances is ignored and a general variance-covariance matrix is estimated across the time-dimension. Finally; (8) True GLS which is denoted by TGLS is obtained for comparison purposes. In fact, the relative efficiency of each estimator is obtained by dividing its MSE by that of TGLS. It is important to emphasize that all the estimators considered are consistent as long as the explanatory variables and the disturbances are uncorrelated, as $N \to \infty$, with *T* fixed. The primary concern in this paper is with their small sample properties.

The literature on pre-test estimators is surveyed in JUDGE *et al.* [1985], while the error component literature is extensively studied in HSIAO [1986] and MÁTYÁS and SEVESTRE [1996]. Testing AR(1) versus MA(1) is well studied in the time-series literature by KING and MCALEER [1987] and

BURKE, GODFREY and TERMAYNE [1990]. In the panel data literature, this issue is considered by BALTAGI and LI [1995]. Pre-test in the autocorrelation context is considered in the time series literature by KING and GILES [1984] and GRIFFITHS and BEESLEY [1984] to mention a few. This paper takes up this issue in a panel data context. More specifically, this paper considers the small sample performance of pure and pre-test estimators in a panel data regression when the remainder error process can be misspecified.

Section 2 describes the model and the estimators considered. Section 3 gives the Monte Carlo design and the results of the experiments, while Section 4 gives our summary and conclusion.

2 The Model

Consider the following panel data regression model

$$y_{it} = x'_{it}\,\beta + u_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{1}$$

where β is $K \times 1$ vector of regression coefficients including the intercept. x_{it} is assumed to be exogenous with $E(u_{it}/x_{js}) = 0$ for $i, j = 1, ..., N$ and $s, t = 1, ..., T$. The disturbances follow an one-way error component model, see MÁTYÁS and SEVESTRE [1996],

$$u_{it} = \mu_i + \nu_{it} \tag{2}$$

with individual effects $\mu_i \sim IIN\,(0,\ \sigma^2_\mu)$ and the remainder disturbances ν_{it} following either a stationary AR(1), *i.e.,*

$$\nu_{it} = \rho\nu_{i,t-1} + \varepsilon_{it} \tag{3}$$

or a stationary MA(1) process, *i.e.,*

$$\nu_{it} = \varepsilon_{it} + \lambda\varepsilon_{i,t-1} \tag{4}$$

where ε_{it} is $IIN(0,\ \sigma^2_\varepsilon)$. BALTAGI and LI (1991, 1994), derived the spectral decomposition of the variance-covariance matrix of the composite error u_{it} when the remainder error ν_{it} follows an AR(p) or an MA(q) process. This allows the derivation a simple transformation that will reduce the disturbances into spherical errors. Briefly, this transformation can be summarized as follows:

Let C be a $T \times T$ matrix that removes the serial correlation in $\nu_i = (\nu_{i1},\ \ldots,\ \nu_{iT})'$, in the sense that $C\nu_i \sim (0,\ \sigma^2 I_T)$.

Step (i): Perform the C transformation on the observations of each individual in equation (1) to remove the serial correlation. We obtain, say $y^*_i = Cy_i$, where $y'_i = (y_{i1},\ \ldots,\ y_{iT})$.

Step (ii): Perform another transformation on the y_{it}^*'s, obtained in step 1, which subtracts from y_{it}^* a fraction of a weighted average of observations on y_i^*, *i.e.*,

$$y_{it}^* = y_{it}^* - \delta\alpha_t\,(\Sigma_{s=1}^T \alpha_s y_{is}^*)/(\Sigma_{s=1}^T \alpha_s^2) \tag{5}$$

where the α_t's are the elements of e_T^α which is defined by $e_T^\alpha = Ce_T \equiv (\alpha_1,\ \alpha_2, \ldots,\ \alpha_T)'$ and e_T is a $T \times 1$ vector with all the elements equal one. $\delta = 1 - (\sigma/\sigma_\alpha)$ and $\sigma_\alpha^2 = \sigma_\mu^2\,(\Sigma_{t=1}^T \alpha_t^2) + \sigma^2$.

Case (1): ν_{it} follows an AR(1) process.

Step (i): The C transformation for the AR(1) model is the familiar Prais-Winsten (PW) transformation on each individual *i.e.*, $y_i^* = Cy_i$, where $y_{i1}^* = \sqrt{1-\rho^2}\,y_{i1}$ and $y_{it}^* = y_{it} - \rho y_{i,t-1}$ for $t = 2, \ldots,\ T$.

Step (ii): The second step transformation is the same as the above general (**) transformation given in (5) with $e_T^\alpha = Ce_T = (1-\rho)(\alpha,\ 1,\ \ldots,\ 1)'$ where $\alpha = \sqrt{(1+\rho)/(1-\rho)}$. Also $\sigma_\alpha^2 = (1-\rho)^2[\alpha^2 + (T-1)]\,\sigma_\mu^2 + \sigma^2$ and $\sigma^2 = \sigma_\varepsilon^2$.

After the two step transformation we get, in vector form

$$y^{**} = x^{**}\,\beta + u^{**} \tag{6}$$

u^{**} now is a classical error, hence OLS on (6) is equivalent to GLS on (1). When ν_{it} follows an AR(1) process, we denote this estimator by $\hat{\beta}_{\mathrm{GLSA}}$.

BALTAGI and LI (1991) also suggest some simple consistent estimators (when both N and T are large) for ρ, σ_α^2 and σ_ε^2. The ρ estimator proposed by BALTAGI and LI [1991] is $\tilde{\rho} = \Sigma_{i=1}^N \Sigma_{t=2}^T \tilde{u}_{it}\tilde{u}_{i,t-1}/\Sigma_{i=1}^N \Sigma_{t=2}^T \tilde{u}_{i,t-1}^2$, where $\tilde{u}_{it}$ denotes the within residuals. $\hat{\sigma}_\alpha^2 = \hat{u}^{*\prime}(I_N \otimes \bar{J}_T^\alpha)\,\hat{u}^*/N$ and $\hat{\sigma}^2 \equiv \hat{\sigma}_\varepsilon^2 = \hat{u}^{*\prime}(I_N \otimes E_T^\alpha)\,\hat{u}^*/N(T-1)$, where $\bar{J}_T^\alpha = e_T^\alpha e_T^{\alpha\prime}/\Sigma\alpha_t^2$; $E_T^\alpha = I_T - \bar{J}_T^\alpha$ and $\hat{u}^*$ denotes the OLS residuals on the (*) transformed equation. $\tilde{\rho}$ is a consistent estimator for ρ only when T is large. This is because the within residual $\tilde{\nu}_{it}$ is a consistent estimator for ν_{it} when T is large. For typical labor or consumer panels, N is large and T is small. In this case, $\tilde{\rho}$ is biased and inconsistent if T is fixed, even when $N \to \infty$.

In this paper we suggest another estimator for ρ, which relies on N being large for consistency. Let $Q_s = E(u_{it}u_{i,t-s})$ denote the autocovariance function of u_{it}, then it is easy to see that $Q_s = \sigma_\mu^2 + \sigma_\nu^2\rho^s$. From Q_0, Q_1 and Q_2, one can easily show that $\rho + 1 = (Q_0 - Q_2)/Q_0 - Q_1)$. Hence a consistent estimator of ρ (for large N) is given by

$$\hat{\rho} = \frac{\tilde{Q}_0 - \tilde{Q}_2}{\tilde{Q}_0 - \tilde{Q}_1} - 1 = \frac{\tilde{Q}_1 - \tilde{Q}_2}{\tilde{Q}_0 - \tilde{Q}_1} \tag{7}$$

where $\tilde{Q}_s = \sum_{i=1}^{N}\sum_{t=s+1}^{T} \hat{u}_{it}\hat{u}_{i,t-s}/N(T-s)$ and $\hat{u}_{it}$ denotes the OLS residuals on (1).

With $\hat{\rho}$ given in (7) instead of $\tilde{\rho}$, it is easy to show that the estimators of $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\varepsilon^2$ given above are consistent estimators of σ_α^2 and σ_ε^2 when N tends to infinity.

Case (2): ν_{it} follows a MA(1) process.

Let $\gamma_s = E(\nu_{it}\nu_{i,t-s})$ denote the autocovariance function of ν_{it} and $r = \gamma_1/\gamma_0$. Note that when ν_{it} follows a MA(1) process, we have $Q_s = \sigma_\mu^2 + \gamma_s$ for $s = 0, 1$ and $Q_s = \sigma_\mu^2$ for $s > 1$. Hence we have $\gamma_\tau = Q_\tau - Q_s\,(\tau = 0, 1)$ for some $s > 1$.

Step (i): First perform the C transformation suggested by LI [1992] to remove the MA(1) serial correlation, *i.e.*, $y_{i1}^* = y_{i1}/\sqrt{a_1}$ and $y_{it}^* = [y_{it} - (r y_{i,t-1}^*)/\sqrt{a_{t-1}}]/\sqrt{a_t}$ for $t = 2, \ldots, T$, where $a_1 = 1$ and $a_t = 1 - r^2/a_{t-1}$ for $t = 2, \ldots, T$. Note that this transformation depends only on r, which can be estimated by $\hat{r} = \hat{\gamma}_1/\hat{\gamma}_0 = (\tilde{Q}_1 - \tilde{Q}_s)/(\tilde{Q}_0 - \tilde{Q}_s)$ for some $s > 1$.

Step (ii): The second step transformation is the same as the general formula (**) given in (5) above. Now $e_T^\alpha = Ce_T = (\alpha_1, \ldots, \alpha_T)'$ with $\alpha_1 = 1$ and $\alpha_t = [1 - r/\sqrt{a_{t-1}}]/\sqrt{a_t}$ for $t = 2, \ldots, T$. Note that in this case $\sigma^2 = \gamma_0$ [1]. The estimators of σ_α^2 and σ^2 are simply given by $\hat{\sigma}_\alpha^2 = (\Sigma_{t=1}^T \hat{\alpha}_t^2)\,\hat{\sigma}_\mu^2 + \hat{\sigma}^2$, and $\hat{\sigma}^2 = \hat{\gamma}_0 = \tilde{Q}_0 - \tilde{Q}_s$ for some $s > 1$ with $\hat{\sigma}_\mu^2 = \tilde{Q}_s$ for some $s > 1$. Finally $\hat{\delta} = 1 - \sqrt{\hat{\gamma}_0/\hat{\sigma}_\alpha^2}$. Again, the OLS estimator on the (**) transformed equation is equivalent to GLS on (1). When ν_{it} follows a MA(1) process, we denote this estimator by $\hat{\beta}_{GLSM}$.

Recently BALTAGI and LI [1995] derived a Lagrange Multiplier (LM) test for testing first order serial correlation in the remainder error term ν_{it} (assuming the existence of random individual effects, *i.e.*, $\sigma_\mu^2 > 0$). This LM test statistic is invariant to the form of the first-order serial correlation, *i.e.*, whether the null hypothesis is H_0^a; $\lambda = 0$ (given $\sigma_\mu^2 > 0$) or H_0^b; $\rho = 0$ (given $\sigma_\mu^2 > 0$). The alternative hypothesis corresponds to H_1^a; $\lambda > 0$ (given $\sigma_\mu^2 > 0$) or H_1^b; $\rho > 0$ (given $\sigma_\mu^2 > 0$), respectively. The explicit form of this LM test statistic is given in BALTAGI and LI (1995, p. 137) and is reproduced here for convenience:

$$LM = \hat{D}_\rho \sqrt{\hat{J}^{11}} \tag{8}$$

where D_ρ is the score with respect to ρ (or λ) evaluated at the null

$$D_\rho = [N(T-1)/T]\,\frac{\sigma_1^2 - \sigma_\varepsilon^2}{\sigma_1^2} + (\sigma_\varepsilon^2/2)\,u'\{I_N \otimes [(\bar{J}_T/\sigma_1^2 + E_T/\sigma_\varepsilon^2)\,G\,(\bar{J}_T/\sigma_1^2 + E_T/\sigma_\varepsilon^2)]\}u$$

$\bar{J}_T = \iota_T\iota_T'/T$, $E_T = I_T - \bar{J}_T$, G is a bi-diagonal matrix with bi-diagonal elements all equal to one, and $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\varepsilon^2$. J^{11} denotes the part of the inverse of the information matrix relating to ρ. In fact,

$$\hat{J}^{11} = N^2T^2(T-1)/\det(\hat{J})\,4\hat{\sigma}_1^4\hat{\sigma}_\varepsilon^4$$

1. The reason of choosing $\sigma^2 = \gamma_0$ instead of choosing $\sigma^2 = \sigma_\varepsilon^2$ for the MA(1) case is that the former choice leads to a much simpler transformation and estimation method. For details, see BALTAGI and LI [1994].

where $\hat{\sigma}_\varepsilon^2 = \hat{u}'(I_N \otimes E_T)\hat{u}/N(T-1)$, $\hat{\sigma}_1^2 = \hat{u}'(I_N \otimes \bar{J}_T)\hat{u}/N$, and $\hat{u}$ are the maximum likelihood residuals under the null hypothesis.

$$\hat{J} = \begin{pmatrix} \hat{J}_{\rho\rho} & N(T-1)\,\hat{\sigma}_\varepsilon^2/\hat{\sigma}_1^4 & \dfrac{N(T-1)}{T}\,\hat{\sigma}_\varepsilon^2[1/\hat{\sigma}_1^4 - 1/\hat{\sigma}_\varepsilon^4] \\ & (NT^2/2\hat{\sigma}_1^4) & NT/2\hat{\sigma}_1^4 \\ & & \dfrac{N}{2}\left[\dfrac{1}{\hat{\sigma}_1^4} + \dfrac{T-1}{\hat{\sigma}_\varepsilon^4}\right] \end{pmatrix}$$

where $\hat{J}_{\rho\rho} = N[2a^2(T-1)^2 + 2a(2T-3) + (T-1)]$ with $a = [(\hat{\sigma}_\varepsilon^2 - \hat{\sigma}_1^2)/T\hat{\sigma}_1^2]$. Under the null hypothesis (H_0^a or H_0^b), LM is asymptotically distributed (for large N) as $N(0,1)$.

BALTAGI and LI [1995] also proposed a simple test statistic for testing a MA(1) process against an AR(1) process for the remainder error term in an error component model. This was denoted by BGT after Burke, GODFREY and TERMAYNE [1990]. The basic idea of the BGT test can be motivated as follows:

Let

$$Q_0 = \frac{\sum\sum u_{it}^2}{NT} = u'u/NT,$$

and

$$Q_s = \frac{\sum\sum u_{it}\,u_{i,t-s}}{N(T-s)} = u'\,(I_N \otimes G_s)u/N(T-s) \quad \text{for } s = 1, \ldots, S,$$

where $G_s = 1/2\,\text{Toeplitz}\,(\iota_s)$, ι_s is a vector of zeros except the $(s+1)$th element which is one. $s = 1, \ldots, S$ with $S \leq (T-1)$ and S is finite [2].

BALTAGI and LI [1995] showed that for large N one can distinguish the AR(1) process from the MA(1) process based on the information obtained from $Q_s - Q_{s+\ell}$, $s \geq 2$ and $\ell \geq 1$. To see this, note that $\text{plim}\,(Q_s - Q_{s+\ell}) = 0$ for the MA(1) process and $\text{plim}\,(Q_s - Q_{s+\ell}) = \sigma_\nu^2\,\rho^s(1-\rho^\ell) > 0$ for the AR(1) process. Hence, BALTAGI and LI [1995] suggest an asymptotic test for the MA(1) model against the AR(1) model based upon

$$\gamma = \sqrt{N/V}\,(Q_2 - Q_3) \tag{9}$$

where $V = 2\,\text{tr}\,\{[(\sigma_\mu^2 J_T + \sigma_\varepsilon^2\,V_\lambda)(G_2/(T-2) - G_3/(T-3))]^2\}$. Under some regularity conditions, γ is asymptotically distributed (for large N) as $N(0,1)$ under the null hypothesis of an MA(1) process. In order to calculate V, we note that for the MA(1) process, $\sigma_\nu^2 = \sigma_\varepsilon^2(1+\lambda^2)$ and

2. Let $a = (a_1, a_2, \ldots, a_n)'$ denote an arbitrary $n \times 1$ vector, then Toeplitz (a) is an $n \times n$ symmetric matrix generated from the $n \times 1$ vector a with the diagonal elements all equal to a_1, second diagonal elements equal to a_2, etc.

$\sigma_\varepsilon^2 V_\lambda = \sigma_\nu^2 I_T + \sigma_\varepsilon^2 \lambda G$. Therefore we do not need to estimate λ in order to compute the test statistic γ, all we need to get are some consistent estimators for σ_ν^2, $\lambda\sigma_\varepsilon^2$ and σ_μ^2. These are obtained as follows:

$$\hat{\sigma}_\nu^2 = \hat{Q}_0 - \hat{Q}_2$$

$$\lambda\hat{\sigma}_\varepsilon^2 = \hat{Q}_0 - \hat{Q}_1$$

$$\hat{\sigma}_\mu^2 = \hat{Q}_2$$

where $\hat{Q}_s$ are obtained from Q_s by replacing u_{it} by the OLS residual $\hat{u}_{it}$. Substituting these consistent estimators into V we get $\hat{V}$, and the test statistic γ becomes

$$\hat{\gamma} = \sqrt{N/\hat{V}}\,(\hat{Q}_2 - \hat{Q}_3) \tag{10}$$

where

$$(\hat{Q}_2 - \hat{Q}_3) = \sum_{i=1}^{N}\sum_{t=3}^{T} \hat{u}_{it}\hat{u}_{i,t-2}/N(T-2) - \sum_{i=1}^{N}\sum_{t=4}^{T} \hat{u}_{it}\hat{u}_{i,t-3}/N\,(T-3),$$

and

$$\hat{V} = 2\mathrm{tr}\,\{[(\hat{\sigma}_\mu^2\, J_T + \hat{\sigma}_\nu^2 I_T + \sigma_\varepsilon^2\,\hat{\lambda}G)(G_2/(T-2) + G_3/(T-3))]^2\}$$

$\hat{\gamma}$ is asymptotically distributed (for large N) as $N(0,1)$ under the null hypothesis and is referred to as the BGT test.

Using these two test statistics, one can obtain a pretest estimator of β given by:

$$\hat{\beta}_{\mathrm{pre}} = \hat{\beta}_{\mathrm{CGLS}} \text{ if } |\,LM\,| \leq c_\alpha;$$
$$\hat{\beta}_{\mathrm{pre}} = \hat{\beta}_{\mathrm{CLSM}} \text{ if } |\,LM\,| > c_\alpha \text{ and } |\,\mathrm{BGT}\,| \leq c_\alpha;$$
$$\hat{\beta}_{\mathrm{pre}} = \hat{\beta}_{\mathrm{CLSA}} \text{ if } |\,LM\,| > c_\alpha \text{ and } |\,\mathrm{BGT}\,| > c_\alpha;$$

where the CGLS, GLSM and GLSA estimators of β were defined above, and c_α is the α-level critical value for a standard normal random variable [3].

3. It is worth emphasizing the pre-test estimator's performance depends on the level of significance c_α of the test.

3 The Monte Carlo Results

First, we describe the Monte Carlo design. We use the following simple regression equation:

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

where $\alpha = 5$ and $\beta = 0.5$, and a similar data generating process as that in NERLOVE [1971], namely $x_{it} = 0.1\,t + 0.5\,x_{i,t-1} + w_{it}$ where w_{it} is uniformly distributed on the interval $[-0.5, 0.5]$. The initial values x_{i0} were chosen as $(5 + 10\,w_{i0})$. $u_{it} = \mu_i + \nu_{it}$ with $\mu_i \sim IIN(0,\ \sigma_\mu^2)$, $\nu_{it} = \rho\nu_{i,t-1} + \varepsilon_{it}$ or $\nu_{it} = \varepsilon_{it} + \lambda\varepsilon_{i,t-1}$ with $\varepsilon_{it} \sim IIN(0,\ \sigma_\varepsilon^2)$. We fix the total variance across experiments to be $\sigma_\mu^2 + \sigma_\nu^2 = 20$ [4]. Next we vary $\eta = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$ to take values (0.2, 0.5, 0.8), respectively. The values of the serial correlation parameters ρ and λ are varied over the following values (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9). We use two different pairs of sample sizes $(N, T) = (50,\ 6)$ and (100, 6) and 2000 replications are performed.

The estimators considered are: (i) the GLS estimator assuming that ν_{it} follows a MA(1) process, denoted by GLSM; (ii) the GLS estimator of β assuming that ν_{it} follows an AR(1) process and using $\hat{\rho}$ given by (7) as an estimator for ρ. This is denoted by GLSA; (iii) the GLS estimator of β assuming that ν_{it} follows an AR(1) process and using $\tilde{\rho}$ given below (6) as an estimator for ρ [5]. This is denoted by GLSA2; (iv) the conventional one-way error components GLS estimator of β assuming no serial correlation for the remainder disturbances, denoted by CGLS; (v) the within estimator of β which treats the μ_i's as fixed effects, denoted by Within; (vi) the OLS estimator, (vii) the pre-test estimator $\hat{\beta}_{\mathrm{PRE}}$ of β which can be CGLS, GLSA or GLSM depending on the results of the preliminary tests. This is denote by PRE. (viii) The GMM estimator based on $E(u_{it}\,x_{it}) = 0$. This estimator ignores the error component structure and estimates a general variance-covariance matrix across time based on $\sum_{i=1}^{N} \hat{u}_i\hat{u}_i'/N$ where $\hat{u}_i$ denote the OLS residuals for the i-th individual. (ix) the true GLS estimator $\hat{\beta}_{\mathrm{TGLS}}$. We then compute the relative efficiency of any estimator $\bar{\beta}$ of β by $\mathrm{MSE}(\tilde{\beta})/\mathrm{MSE}(\hat{\beta}_{\mathrm{TGLS}})$ where $\mathrm{MSE}(\tilde{\beta}) = \sum_{r=1}^{2000} (\tilde{\beta}_r - \beta)^2/2000$ and $\tilde{\beta}_r$ denotes an estimator of β in the r-th replication. These relative MSE's

4. The results of the experiments are invariant to the choice of the true values of the regression parameters α and β (which are fixed at 5 and 0.5) and the remainder variance (which is fixed at 20), see BREUSCH [1980].

5. Note that $\hat{\rho}$ for GLSA is based upon OLS resuduals whereas $\tilde{\rho}$ for GLSA2 is based on within residuals. The former estimator of ρ is consistent when N is large whereas the latter is consistent only if T is large.

are denoted by RGLSM, RGLSA, RGLSA2, RCGLS, RW, ROLS, RPRE and RGMM, respectively.

Tables 1 and 2 below report our findings for the case of $(N,T) = (50,6)$, and Figures 1 and 2 plot these results for $\eta = 0.2$ from the above.

TABLE 1

$(N,T) = (50,6)$ ***Relative MSE with respect to true GLS. True model is MA(1); 2000 replications.***

					$\eta = 0.2$					
λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
RPRE	1.012	1.010	1.005	1.004	1.009	1.014	1.021	1.019	1.007	1.018
RGLSM	1.010	1.007	1.005	1.004	1.007	1.013	1.016	1.009	1.000	1.000
RGLSA	1.009	1.010	1.018	1.043	1.097	1.183	1.289	1.379	1.406	1.437
RGLSA2	1.053	1.067	1.084	1.108	1.145	1.198	1.266	1.353	1.362	1.386
RCGLS	1.018	1.047	1.093	1.155	1.234	1.332	1.442	1.547	1.594	1.625
RW	1.073	1.106	1.156	1.221	1.304	1.404	1.514	1.612	1.638	1.674
RGMM	1.775	1.843	1.875	1.989	2.149	2.187	2.165	2.355	2.286	2.434
ROLS	1.198	1.233	1.286	1.360	1.458	1.587	1.754	1.973	2.355	2.319
					$\eta = 0.5$					
RPRE	1.007	1.012	1.013	1.017	1.025	1.036	1.036	1.036	1.041	1.044
RGLSM	1.011	1.011	1.013	1.017	1.027	1.036	1.032	1.025	1.022	1.024
RGLSA	1.013	1.017	1.031	1.064	1.128	1.221	1.325	1.415	1.454	1.479
RGLSA2	1.042	1.057	1.076	1.104	1.144	1.201	1.277	1.358	1.406	1.479
RCGLS	1.011	1.039	1.084	1.146	1.226	1.327	1.446	1.563	1.633	1.668
RW	1.057	1.085	1.130	1.193	1.274	1.376	1.496	1.605	1.620	1.699
RGMM	1.656	1.812	1.789	1.786	1.988	2.010	2.279	2.350	2.309	2.477
ROLS	1.726	1.739	1.788	1.880	2.028	2.255	2.607	3.207	4.678	4.469
					$\eta = 0.8$					
RPRE	1.026	1.036	1.037	1.044	1.058	1.073	1.078	1.082	1.096	1.104
RGLSM	1.034	1.034	1.037	1.044	1.055	1.061	1.058	1.061	1.071	1.075
RGLSA	1.013	1.020	1.037	1.075	1.139	1.222	1.310	1.391	1.438	1.466
RGLSA2	1.038	1.050	1.065	1.089	1.126	1.182	1.255	1.336	1.392	1.418
RCGLS	1.008	1.032	1.072	1.128	1.203	1.298	1.410	1.523	1.600	1.633
RW	1.017	1.042	1.083	1.139	1.215	1.310	1.422	1.533	1.607	1.641
RGMM	1.625	1.701	1.770	1.762	2.022	2.102	2.115	2.205	2.385	2.451
ROLS	4.285	4.198	4.219	4.369	4.690	5.272	6.302	8.313	13.936	12.992

TABLE 2

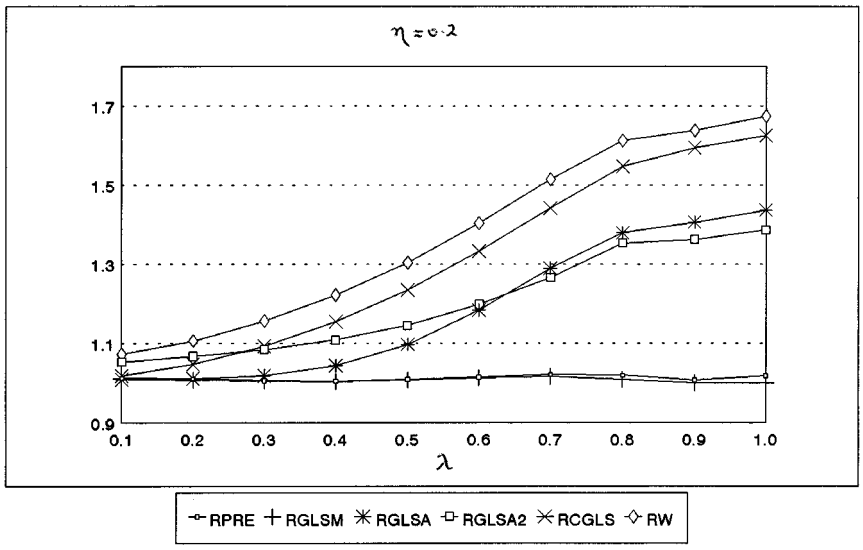
$(N,T) = (50,6)$ ***Relative MSE with respect to true GLS. True model is AR(1); 2000 replications.***

					$\eta = 0.2$					
ρ	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
RPRE	1.010	1.013	1.020	1.022	1.026	1.030	1.027	1.033	1.024	1.032
RGLSM	1.015	1.012	1.014	1.021	1.037	1.062	1.101	1.160	1.244	1.341
RGLSA	1.010	1.010	1.012	1.014	1.016	1.018	1.019	1.020	1.017	1.015
RGLSA2	1.041	1.055	1.074	1.098	1.129	1.166	1.212	1.269	1.335	1.413
RCGLS	1.006	1.019	1.050	1.101	1.170	1.260	1.373	1.511	1.679	1.877
RW	1.055	1.073	1.107	1.159	1.226	1.312	1.417	1.546	1.703	1.893

RGMM	1.715	1.781	1.818	1.946	2.051	2.014	2.033	2.025	2.127	2.085
ROLS	1.182	1.202	1.248	1.324	1.437	1.602	1.847	2.241	2.974	4.938
					$\eta = 0.5$					
RPRE	1.005	1.008	1.013	1.018	1.023	1.026	1.040	1.043	1.036	1.065
RGLSM	1.012	1.011	1.012	1.017	1.027	1.047	1.081	1.132	1.196	1.282
RGLSA	1.011	1.012	1.014	1.015	1.016	1.015	1.014	1.012	1.011	1.021
RGLSA2	1.030	1.042	1.058	1.078	1.101	1.130	1.165	1.212	1.275	1.362
RCGLS	1.000	1.011	1.037	1.079	1.136	1.209	1.302	1.418	1.569	1.765
RW	1.046	1.056	1.082	1.112	1.177	1.248	1.337	1.448	1.593	1.780
RGMM	1.606	1.760	1.728	1.747	1.928	2.108	2.179	2.178	2.251	2.442
ROLS	1.749	1.728	1.746	1.805	1.914	2.094	2.389	2.903	3.944	7.010
					$\eta = 0.8$					
RPRE	1.010	1.026	1.033	1.036	1.041	1.059	1.070	1.067	1.092	1.179
RGLSM	1.035	1.034	1.035	1.040	1.053	1.077	1.115	1.168	1.241	1.355
RGLSA	1.011	1.013	1.016	1.019	1.022	1.024	1.022	1.020	1.027	1.057
RGLSA2	1.028	1.039	1.054	1.072	1.097	1.129	1.171	1.228	1.302	1.401
RCGLS	1.000	1.008	1.033	1.074	1.131	1.208	1.309	1.440	1.612	1.837
RW	1.007	1.017	1.043	1.084	1.142	1.220	1.320	1.451	1.621	1.837
RGMM	1.586	1.660	1.715	1.824	1.963	2.001	2.029	2.162	2.201	2.359
ROLS	4.474	4.289	4.213	4.249	4.420	4.772	5.412	6.603	9.159	17.221

FIGURE 1

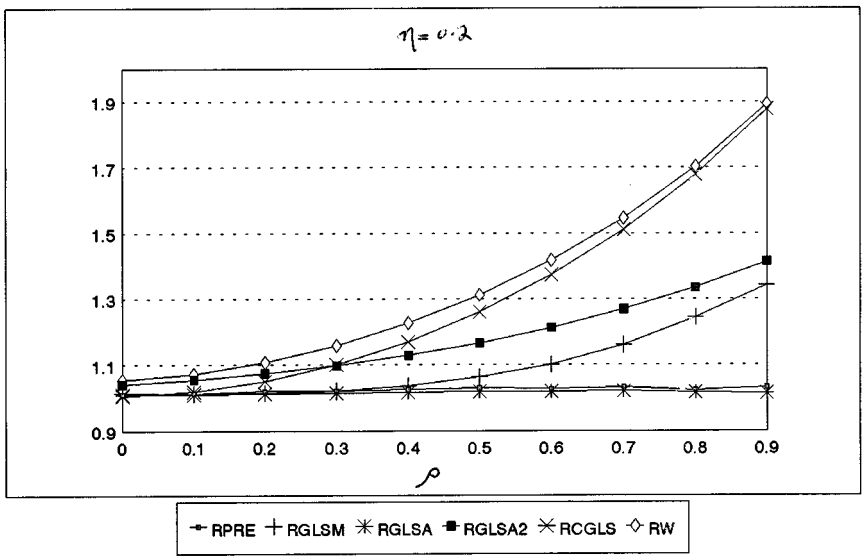
Relative MSE (True Model: MA(1)) Case I: (N,T) = (50,6).



From the above results, we observe the following: (i) OLS performs the worst, and its performance deteriorates as the variance component ratio η increases or the serial correlation coefficient (ρ or λ) increases. The GMM

FIGURE 2

Relative MSE (True Model: AR(1)) Case I: (N,T) = (50,6).



estimator performs better than OLS but its performance is still worse than the other error component type estimators [6]. (ii) CGLS is the best only when $\rho = \lambda = 0.$. However, both CGLS and the within estimator deteriorate in performance as the serial correlation coefficient (ρ or λ) increases. This is not surprising given that both estimators ignore the serial correlation in the remainder term. (iii) As expected, when ν_{it} follows a MA(1) process, GLSM gives an all around best performance for all values of λ. This is followed closely by the preliminary test estimator PRE. GLSA and GLSA2 which assume an AR(1) process perform badly as λ gets large, especially for $\lambda \geq 0.5$ (iv) However, when ν_{it} follows an AR(1) process, GLSA gives an all around best performance for all values of ρ This is followed closely by the preliminary test estimator PRE. GLSM which assumes an MA(1) process performs badly as ρ increases, especially for $\rho > 0.6$. GLSA2, which uses $\tilde{\rho}$, performs badly compared to GLSA especially when ρ is > 0.4. In fact, GLSM performs better than GLSA2 even though ν_{it} follows an AR(1) process. Therefore, for typical panels where $N \gg T$, if the true model is AR(1) it is better to use GLSA rather than GLSA2 since the former provides a consistent estimator of ρ which relies on large N rather than large T to achieve consistency. Table 3 gives the MSE of $\tilde{\rho}$ defined below

6. OLS and GMM were not included in the figures due to their relatively worse performance.

TABLE 3

MSE of various estimates of ρ. True model is AR(1); 2000 replications.

					$(N,T)=(50,6)$					
					$\eta=0.2$					
ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\rho}$	.0101	.0102	.0117	.0138	.0146	.0164	.0185	.0215	.0217	.0243
$\tilde{\rho}$	.0441	0.566	.0718	.0886	.1110	.1340	.1647	.2018	.2462	.2916
					$\eta=0.5$					
$\hat{\rho}$	.0103	.0119	.0139	.0142	.0165	.0189	.0235	.0239	.0262	.0300
$\tilde{\rho}$	.0442	.0569	.0715	.0881	.1102	.1347	.1664	.2022	.2432	.2925
					$\eta=0.8$					
$\hat{\rho}$	.0160	.0161	.0195	.0229	.0253	.0292	.0340	.0364	.0380	.0515
$\tilde{\rho}$	.0441	.0566	.0718	.0886	.1111	.1339	.1647	.2019	.2462	.2916
					$(N,T)=(100,6)$					
					$\eta=0.2$					
ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\rho}$	.0048	.0054	.0060	.0063	.0068	.0082	.0097	.0113	.0130	.0130
$\tilde{\rho}$	.0419	.0542	.0686	.0859	.1013	.1303	.1598	.1960	.2378	.2873
					$\eta=0.5$					
$\hat{\rho}$	.0054	.0060	.0066	.0071	.0081	.0099	.0103	.0140	.0160	.0166
$\tilde{\rho}$	.0428	.0540	.0684	.0852	.1054	.1316	.1612	.1958	.2372	.2879
					$\eta=0.8$					
$\hat{\rho}$	.0074	.0087	.0094	.0105	.0126	.0159	.0279	.0220	.0250	.0297
$\tilde{\rho}$	.0421	.0545	.0684	.0863	.1052	.1311	.1604	.1959	.2363	.2867

(6) and $\hat{\rho}$ defined in (7), when the true model is AR(1) and $(N,T)=(50,6)$ and (100,6) respectively. It is clear, that $\hat{\rho}$ performs better in MSE than $\tilde{\rho}$ for all values of ρ and η. This confirms the asymptotic results of large N and small T and shows that for $T=6$, the MSE performance of $\tilde{\rho}$ can be relatively bad compared to $\hat{\rho}$ especially for large values of ρ. When the true model is MA(1), GLSA performs better than GLSA2 for $\lambda<0.5$ and worse for $\lambda\geq 0.5$, for all values of η [7]. These rankings are robust to doubling N from 50 to 100, keeping T fixed at 6, as shown in Figures 3 and 4 for $\eta=0.2$ [8].

In conclusion, we note that the correct GLS procedure is always the best, but the researcher does not have perfect foresight on which one it is : GLSA for an AR(1) process, or GLSM for an MA(1) process. In this case, the pre-test estimator is a viable alternative given that its performance is a close second to correct GLS whether the true serial correlation process is AR(1) or MA(1). This is clearly demonstrated by Figures 1 to 4.

7. One exception, for $\eta=0.2$, this rule has to be modified to $\lambda>0.6$.

8. Similar results for $(N,T)=(100,6)$ are obtained for $\eta=0.5$ and $\eta=0.8$ and these graphs are available upon request from the authors.

FIGURE 3

Relative MSE (True Model: MA(1)) Case II: (N,T) = (100,6).

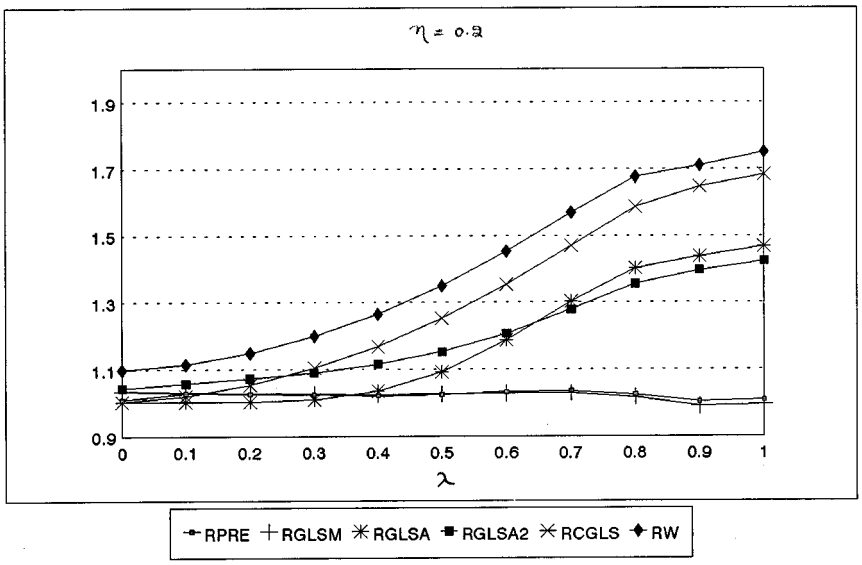
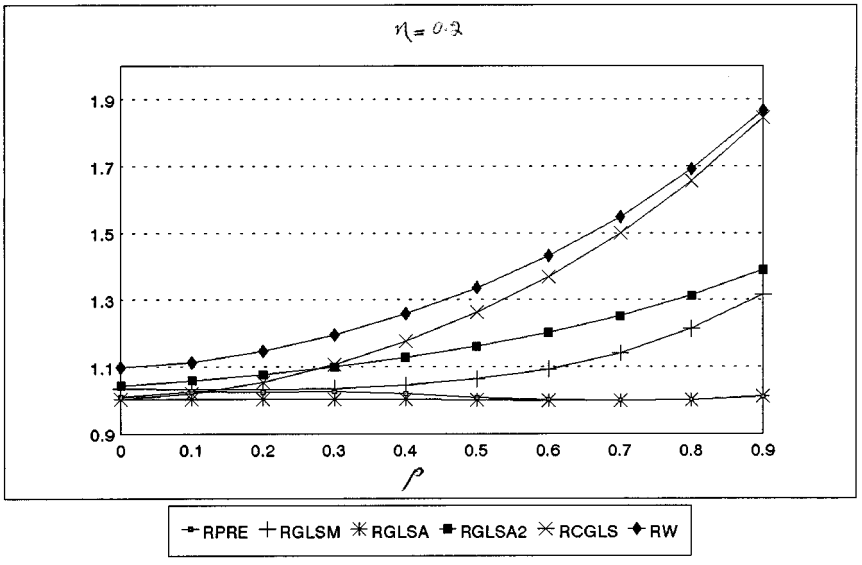


FIGURE 4

Relative MSE (True Model: AR(1)) Case II: (N,T) = (100,6).



• References

BALTAGI, B.H., LI, Q. (1991). – "A Transformation that will Circumvent the Problem of Autocorrelation in an Error Component model", *Journal of Econometrics*, 48, pp. 385-393.

BALTAGI, B.H., LI, Q. (1994). – "Estimating Error Component Models with General MA(q) Disturbances, *Econometric Theory*, 10, pp. 396-408.

BALTAGI, B. H., LI, Q. (1995). – "Testing AR(1) Against MA(1) Disturbances in an Error Component Model", *Journal of Econometrics*, 68, pp. 133-151.

BREUSCH T. S. (1980). – "Useful Invariance Results for Generalized Regression Models", *Journal of Econometrics*, 13, pp. 327-340.

BURKE, S. P., GODFREY L. G., TERMAYNE A. R. (1990). – "Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model: An Alternative Procedure", *Review of Economic Studies*, 57, pp. 135-145.

FULLER, W. A., BATTESE G. E. (1974). – "Estimation of Linear Models with Cross-Error Structure", *Journal of Econometrics*, 2, pp. 67-78.

GRIFFITHS, W. E., BEESLEY P. A. A. (1984). – "The Small-Sample Properties of Some Preliminary Test Estimators in a Linear Model with Autocorrelated Errors", *Journal of Econometrics*, 25, pp. 49-61.

HSIAO, C. (1986). – "*Analysis of panel data*", (Cambridge University Press, Cambridge).

JUDGE G. G., GRIFFITHS, W. E. HILL R. C., LUTKEPOHL H., LEE T. C. (1985). – *The theory and practice of econometrics*, (Wiley, New York).

KING M. L., GILES D. E. A. (1984). – "Autocorrelation prestesting in the Linear Model: Estimation, Testing and Prediction", *Journal of Econometrics*, 25, pp. 35-48.

KING, M. L., MCALEER, M. (1987). – "Further Results on Testing AR(1) against MA(1) Disturbances in the Linear Regression Model", *Review of Economic Studies*, 54, pp. 649-663.

LI, Q. (1992). – "Estimating a Linear Regression Model with General MA(q) Disturbances", *Working paper*, University of Guelph, Ontario, Canada.

MÁTYÁS, L., SEVESTRE P., Editors (1996). – The Econometrics of Panel Data: Handbook of Theory and Applications (Kluwer Academic Publishers, Dordrecht).

NERLOVE, M. (1971). – "Further Evidence of the Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections", *Econometrica*, 39, pp. 359-382.