doi:10.1111/j.1467-9892.2007.00569.x

USING LEAST SQUARES TO GENERATE FORECASTS IN REGRESSIONS WITH SERIAL CORRELATION

By Sergio G. Koreisha and Yue Fang

University of Oregon, University of Queensland and University of Oregon, China Europe International Business School

First Version received April 2006

Abstract. The topic of serial correlation in regression models has attracted a great deal of research in the last 50 years. Most of these studies have assumed that the structure of the error covariance matrix Ω was known or could be consistently estimated from the data. In this article, we describe a new procedure for generating forecasts for regression models with serial correlation based on ordinary least squares and on an approximate representation of the form of the autocorrelation. We prove that the predictors from this specification are asymtotically efficient under some regularity conditions. In addition, we show that there is not much to be gained in trying to identify the correct form of the serial correlation since efficient forecasts can be generated using autoregressive approximations of the autocorrelation. A large simulation study is also used to compare the finite sample predictive efficiencies of this new estimator vis-à-vis estimators based on ordinary least squares and generalized least squares.

Keywords. Autocorrelation; autoregressive moving average; least squares; model identification; prediction.

1. INTRODUCTION

It is well known that under some regularity conditions, ordinary least squares (OLS) yield unbiased, but inefficient estimates for parameters in regression models with serially correlated error structures, and that these OLS regression estimates usually have larger sampling variances than those obtained from procedures such as generalized least squares (GLS) that deal explicitly with the autocorrelation of the residuals. Furthermore, and of particular concern for us in this article, that for finite samples forecasts generated from such models can be seriously inefficient, not just because of issues associated with parameter estimation, but also because the error between the fitted and actual value in the last observation is apt to persist into the future.

Most estimation methods that deal explicitly with serial correlation such as generalized least squares or transfer functions presuppose that the structure of the covariance matrix can be correctly identified and estimated consistently from the data. The practical reality is, as reported by Thursby (1987), Koreisha and Pukkila (1987), among others, is quite different. For finite

0143-9782/08/03 555–580 JOURNAL OF TIME SERIES ANALYSIS Vol. 29, No. 3

samples identification of the serial correlation structure in regression models can be quite elusive.

In this article, we present an efficient, easy to implement forecasting procedure based solely on least squares which does not require a priori knowledge of covariance matrix for regression models with serial correlation. In Section 2, we describe the new two-step ordinary least squares (2SOLS) forecasting procedure, and prove that it is asymptotically efficient. Here we also show that as the sample size increases, predictive efficiency does not dependent on efficient estimates; the crucial factor in generating efficient forecast, as we shall demonstrate, is in dealing with the residual autocorrelation which we do by using an AR($\tilde{p}$) approximation. We show that for sufficiently large T (sample size) and $\tilde{p}$, the L-step ahead forecasting error associated with 2SOLS has the same limiting distribution as that of GLS. This implies that there is not much to be gained in trying to identify the correct form of the autocorrelation since efficient forecasts can be generated using AR($\tilde{p}$) approximations. In Section 3, we present results from an exhaustive simulation study for sample sizes ranging from 50 to 500 observations and covering a wide spectrum of ARMA (p,q) serial correlation structures. We compare the efficiency of forecasts generated by the 2SOLS method with other approaches based on OLS and GLS. In Section 4, we demonstrate the forecasting performance of the proposed new method using real economic data. Finally, in Section 5, we offer some concluding remarks.

2. TWO-STEP FORECASTING PROCEDURES BASED ON OLS

2.1. The procedure

Suppose the serial correlation of the regression model,

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{i,t} + a_t, \tag{1}$$

for which we are interested in generating forecasts follows a stationary and invertible ARMA process (Box and Jenkins, 1976),

$$\Phi(B)a_t = \Theta(B)v_t, \tag{2}$$

where $\Phi(B)$ and $\Theta(B)$ are finite polynomials of orders p and q respectively in the back shift operator B, such that $B^j w_t = w_{t-j}$, and $\{v_t\}$ is a white noise process with variance σ_v^2.

The two-step forecasting procedure consists in first obtaining OLS estimates for the residual series $\hat{a}_t$ of the regression model (1). Then, assuming that the form of the ARMA serial correlation can be approximated by an AR($\tilde{p}$) process (Box and Jenkins, 1976; Koreisha and Fang, 2001), an augmented regression model that includes $\tilde{p}$ additional first-step residual variables,

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{i,t} + \sum_{j=1}^{\tilde{p}} \rho_j \hat{a}_{t-j} + u_t \tag{3}$$

is estimated using OLS yielding new 2SOLS residuals and estimates for β_i, namely $\hat{\beta}_i^{2SOLS}$. Based on the thus derived $\hat{\beta}_i^{2SOLS}$ as well as $\hat{\rho}_j$ from (3), the L-step ahead forecasts at time T can be constructed by sequentially generating future estimates for $\hat{a}_{T+L}$ from

$$\hat{a}_{T+L} = \sum_{j=1}^{\tilde{p}} \hat{\rho}_j \hat{a}_{T+L-j},$$

thus yielding the forecasts for y_{T+L}, namely,

$$\hat{y}_{T+L}^{2SOLS} = \hat{\beta}_0^{2SOLS} + \sum_{i=1}^{k} \hat{\beta}_i^{2SOLS} x_{i,T+L} + \hat{a}_{T+L}. \tag{4}$$

The selection of an appropriately large $\tilde{p}$, as we shall demonstrate, is necessary to establish the efficiency of the forecasting method. Later in this section, we will present some guideline for selecting $\tilde{p}$ based on sample size.

It should also be noted that this 2SOLS procedure could be allowed to iterate until convergence based on some criterion such as mean square error (MSE) is established (see, for example, Kapetanios (2003) for discussions on whether such iterative procedures converge in small samples). In this study, we will only present the results associated with the first iteration. This is because in preliminary trials we did not observe significant gains in forecasting efficiencies after a few iterations.

2.2. Asymptotic properties

Although extensive research has been conducted on the properties of predictors from serially correlated regression models, there are very few studies that focus on predictors obtained when the form of the autocorrelation is misspecified. [In fact, the preponderance of these studies assume that the form of the residual correlation is known, and, because of estimation issues, that it is governed by an AR process, e.g. Yamamoto (1976), Bhansali (1978), and Baille (1979).] Of the few who have considered this problem like Fang and Koreisha (2004), practically all used predictors based on GLS estimation.

In this section, we will show that under classic regression regularity conditions the predictor based on the 2SOLS procedure is asymptotically efficient, i.e. for sufficiently large T, and an appropriately selected AR order $\tilde{p}$, the L-step ahead forecasting error of $\hat{y}_{T+L}^{2SOLS}$ obtained from (4) is essentially the same as the one based on GLS. In the process, we also show that the β_i^{2SOLS} estimates are consistent.

As in Amemiya (1985) and Judge et al. (1985) among others, we begin by making the following assumptions on the design matrix and residuals:

ASSUMPTION 1. $\{a_t\}$ are invertible and stationary. The coe cients of $\Theta(B)\Phi^{-1}(B)$ and $\Phi(B)\Theta^{-1}(B)$, $\{\vartheta_j\}$ and $\{\varphi_j\}$, are absolutely summable.

ASSUMPTION 2. $p\lim_{T\to\infty}(X'X/T)$ is finite and nonsigular.

ASSUMPTION 3. $p\lim_{T\to\infty}(X'A/T) = 0$.
where X is the $T \times k$ design matrix and A is an $T \times \tilde{p}$ matrix with the $(i,j)-$ entry $a_{i,j} = a_{T-i-j+1}$.

These assumptions ensure that the OLS estimates of β_i obtained from (1) are consistent. However, because some of the regressors (namely, the estimated residuals from (1)) are correlated with the error term in (3), to demonstrate that the β_i^{2SOLS} estimates are consistent, it is necessary to show that, as the lag $\tilde{p}$ in (3) goes to infinity, these correlations will approach zero in probability. Lemmas 1 and 2 below provide the basis for establishing our first theorem demonstrating the consistency of the β_i^{2SOLS} estimates.

Consider the autoregression of order $\tilde{p}$,

$$a_t = \sum_{j=1}^{\tilde{p}} \phi_j a_{t-j} + u_t.$$

Let U be an T-dimensional vector defined by

$$U = \left\{ \sum_{j=\tilde{p}+1}^{\infty} \phi_j a_{T-k-j} + v_{T-k} \right\}_{k=0}^{T-1}$$

and $\{\gamma_j\}$ is the covariance function of a_t.

LEMMA 1. Under Assumptions 1–3, it can be shown that

(1)

$$p\lim_{T\to\infty} \frac{A'U}{T} = \begin{pmatrix} \sum_{j=\tilde{p}+1}^{\infty} \phi_j \gamma_{j-1} \\ \sum_{j=\tilde{p}+1}^{\infty} \phi_j \gamma_{j-2} \\ \vdots \\ \sum_{j=\tilde{p}+1}^{\infty} \phi_j \gamma_{j-\tilde{p}} \end{pmatrix};$$

(2) $p\lim_{T\to\infty} \frac{A'A}{T}$ is finite and nonsingular, where

$$p\lim_{T\to\infty}\frac{A'A}{T}=\begin{pmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{\tilde{p}-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{\tilde{p}-2}\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \gamma_{\tilde{p}-2} & \gamma_{\tilde{p}-3} & \cdots & \gamma_1\\ \gamma_{\tilde{p}-1} & \gamma_{\tilde{p}-2} & \cdots & \gamma_0\end{pmatrix};$$

(3)

$$p\lim_{T\to\infty}\frac{X'U}{T}=0.$$

Proof. For all $k = 0, 1, 2, \ldots, (\tilde{p} - 1)$, it can be shown that

$$T^{-1}\sum_{i=1}^{T}\left\{(a_{T-i-k})\left(\sum_{j=\tilde{p}+1}^{\infty}\phi_j a_{T-j-i+1}+v_{T-i+1}\right)\right\}$$
$$=\sum_{j=\tilde{p}+1}^{\infty}\phi_j(T^{-1}\sum_{i=1}^{T}a_{T-i-k}a_{T-i-j+1})+T^{-1}\sum_{i=1}^{T}a_{T-i-k}v_{T-i+1},$$

which has the limit $\sum_{j=\tilde{p}+1}^{\infty}\phi_j\gamma_{|j-k-1|}$. Hence, $T^{-1}A'U$ converges to the desired limit. Similarly, under Assumptions 1–3, the convergence results of $T^{-1}A'A$ and $T^{-1}X'U$ follow from direct matrix algebra manipulation. □

Since OLS residuals $\{\hat{a}_t\}$ have the same asymptotic distribution as $\{a_t\}$ [Judge et al., p. 172 (1985)], the covariance matrix of $\{\hat{a}_t\}$ approximates that of $\{a_t\}$ as $T \to \infty$. Hence, we have,

Lemma 2. The results in Lemma 1 hold for the autoregression of order $\tilde{p}$ based on OLS residuals $\{\hat{a}_t\}$,

$$\hat{a}_t=\sum_{j=1}^{\tilde{p}}\phi_j\hat{a}_{t-j}+w_t.$$

Now let β denote the vector of unknown regression parameters and $\hat{\beta}^{2SOLS}$ the corresponding vector of 2SOLS estimates.

Theorem 1. Under Assumptions 1–3,

$$p\lim_{T\to\infty}\hat{\beta}^{2SOLS}=\beta.$$

Proof. Consider the regression model

$$y_t=\beta_0+\sum_{i=1}^{k}\beta_i x_{i,t}+\sum_{j=1}^{\tilde{p}}\phi_j a_{t-j}+u_t. \tag{5}$$

Let $\beta^* \equiv (\beta_0, \beta_1, \beta_2, \ldots, \beta_k, \phi_1, \phi_2, \ldots, \phi_{\tilde{p}})'$ be the OLS coefficients of the regressors and $\hat{\beta}^*$ the corresponding estimates, then

$$\hat{\beta}^* - \beta^* = \left(\frac{\Xi'\Xi}{T}\right)^{-1} \frac{\Xi' U}{T}, \tag{6}$$

where Ξ = (X,A), and U and A as defined in Lemma 1.

Note that the first term on the right-hand side of (6)

$$\Xi'\Xi = \begin{pmatrix} X' \\ A' \end{pmatrix} (X \quad A) = \begin{pmatrix} X'X & X'A \\ A'X & A'A \end{pmatrix}.$$

Therefore,

$$T(\Xi'\Xi)^{-1} = T\begin{pmatrix} (X'X)^{-1} + (X'X)^{-1}(X'A)B^{-1}(A'X)(X'X)^{-1} & -(X'X)^{-1}(X'A)B^{-1} \\ -B^{-1}(A'X)(X'X)^{-1} & B^{-1} \end{pmatrix},$$

where $B = A'A - (A'X)(X'X)^{-1}(X'A)$.

From the general assumptions and the results of Lemma 1, it can be verified that

$$TB^{-1} = T(A'A)^{-1} + o_p(1),$$

$$T[(X'X)^{-1} + (X'X)^{-1}(X'A)B^{-1}(A'X)(X'X)^{-1}] = T(X'X)^{-1} + o_p(1),$$

and

$$T(X'X)^{-1}(X'A)B^{-1} = o_p(1).$$

Thus,

$$T(\Xi'\Xi)^{-1} = T\begin{pmatrix} (X'X)^{-1} & 0 \\ 0 & (A'A)^{-1} \end{pmatrix} + o_p(1),$$

which has a finite and non-singular limit.

The second term on the right-hand side of (6),

$$\frac{\Xi'U}{T} = \frac{\begin{pmatrix} X'U \\ A'U \end{pmatrix}}{T}.$$

Let $\beta^*_{(k)} \equiv (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)'$ be first k-entries of β^* and $\hat{\beta}^*_{(k)}$ the corresponding estimates. Note that $(\hat{\beta}^*_{(k)} - \beta^*_{(k)})$, is an k-dimensional vector consisting of the first k-entries of the $(k + \tilde{p})$-dimensional vector $(\hat{\beta}^* - \beta^*)$. Since $T^{-1}A'A = O_p(1)$ and $T^{-1}A'U = o_p(1)$ by Lemma 1, we have

$$\hat{\beta}^*_{(k)} - \beta^*_{(k)} = T(X'X)^{-1}\frac{X'U}{T} + o_p(1).$$

The first term on the right-hand side of the above equation, $T(X'X)^{-1}\frac{X'U}{T}$, converges to zero in probability by Assumption 2 and the results of Lemma 1. Therefore, $\hat{\beta}^{2SOLS} = \beta + o_p(1)$ by Lemma 2. □

Now expanding the mean square error of the forecasts based on our 2SOLS procedure,

$$\left(y_{T+L} - \hat{y}^{2SOLS}_{T+L}\right)^2 = [x'_{T+L}(\beta - \hat{\beta}^{2SOLS}) + (a_{T+L} - \hat{a}_{T+L})]^2$$

yields,

$$(x'_{T+L}(\beta - \hat{\beta}^{2SOLS}))^2 + (a_{T+L} - \hat{a}_{T+L})^2 + 2(x'_{T+L}(\beta - \hat{\beta}^{2SOLS}))(a_{T+L} - \hat{a}_{T+L}).$$

From Theorem 1, $(\beta - \hat{\beta}^{2SOLS})$ and hence $x'_{T+L}(\beta - \hat{\beta}^{2SOLS})$ converge to zero in probability as T goes to infinity. Consequently, to show that these forecasts are asymptotically efficient, it will be necessary to just demonstrate that the term relating to the difference between actual future residuals and our estimated AR($\tilde{p}$) approximation, $(a_{T+L} - \hat{a}_{T+L})$, converges to the minimum mean square L-step prediction error for Y_{T+L}. Below we derive this result for sufficiently large T and $\tilde{p}$.

Lemma 3. Given $\epsilon > 0$, one can find $T_0(\epsilon)$ and $\tilde{p}_0(\epsilon)$ such that for $T > T_0(\epsilon)$ and $\tilde{p} > \tilde{p}_0(\epsilon)$,

$$|E(\hat{a}_{T+L} - a_{T+L})^2 - \sigma_L^2| < \epsilon,$$

where σ_L^2 is the minimum mean square L-step prediction error based on $\{a_T, a_{T-1}, \ldots\}$, which is independent of T and given by $\sigma_v^2 \sum_{j=0}^{L-1} \vartheta_j^2$.

The proof is given in the Appendix.

Therefore,

Theorem 2. Under Assumptions 1–3, given $L > 0$ and $\delta > 0$, for any $\epsilon > 0$, one can find $T_0(\epsilon)$ and $\tilde{p}_0(\epsilon)$ such that for $T > T_0(\epsilon)$ and $\tilde{p} > \tilde{p}_0(\epsilon)$,

$$P[|(\hat{y}^{2SOLS}_{T+L} - y_{T+L})^2 - \sigma_L^2| > \delta) < \epsilon,$$

where σ_L^2 is defined in Lemma 3.

It should be noted that the 2SOLS procedure does not, in general, yield efficient estimates of the regression parameters β's. (As in Zyskind (1967), it can be shown that 2SOLS will be efficient if and only if $X = Q\Gamma$, where Q contains the characteristic vectors of Ω, and Γ is a non-singular matrix.) Hence, alternative methods based on GLS (see Koreisha and Fang, 2001) should be used if model estimation and inference are the main focus of the study. However, forecasting efficiency, as we just demonstrated, does not depend on the use of efficient estimators. The key to obtaining efficient forecasts lies in dealing with the serial correlation. In fact, from Theorem 2, we see that the 2SOLS predictor $\hat{y}_{T+L}^{2SOLS}$ is efficient for all L, provided that T and $\tilde{p}$ are sufficiently large. In Section 3, using a large simulation study, we will demonstrate that even for sample sizes generally available to model builders, $\hat{y}_{T+L}^{2SOLS}$ can yield comparable forecasts to those generated from GLS using the correct form of the residual autocorrelation structure. We will also decompose the predictive mean square errors into their component parts, namely the portion attributable to the estimation of the regression parameters and the portion associated with the serial correlation to gain additional insights on relative causes of the forecast inefficiencies.

2.3. Choosing the lag order $\tilde{p}$

The rationale for using an AR($\tilde{p}$) approximation is based on the fact that any stationary and invertible ARMA(p,q) model can be expressed as an infinite autoregression, $a_t = \Pi(B)a_t + v_t$, where $\Pi(B) = \sum_{i=1}^{\infty} \pi_i B^i$. Hence the serially correlated regression model (3) and (4) can be rewritten as

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i x_{i,t} + \Pi(B)a_t + v_t. \tag{7}$$

To operationalize the proposed 2SOLS procedure it is necessary to select the value of $\tilde{p}$ as an input. Since the coefficients in Π(B) in (7) may be effectively zero beyond some finite lag in the sense of Hannan (1970), the infinite autoregressive representation may be approximated by an AR($\tilde{p}$) process whose order $\tilde{p}$ depends on the number of observations, and the rate for which the AR coefficients converge to zero, i.e. $\tilde{p} = p(T)$. For time-series models typically $\tilde{p} = O(T^{\lambda})$ for some positive λ (Berk, 1974; Bhansali, 1978) or $\tilde{p} = O\{(\ln T)^{\alpha}\}$ for $\alpha > 0$ (Hannan et al., 1980; Saikkonen, 1986). It is well-known that one needs to allow the order of the AR process, $\tilde{p}$, to go to infinity as $T \to \infty$ to obtain efficient ARMA estimates (Wahlberg, 1989). Not much, however, is known about such types of convergence rates in the context of regressions with serially correlated ARMA residuals nor in finite sample cases.

Through extensive experimentation Poskitt and Salau (1995) in demonstrating asymptotic equivalency of the Koreisha and Pukkila (1990) generalized least squares estimation procedure for univariate and vector ARMA processes vis-à-vis Gaussian estimation procedures, have shown that for sample sizes similar to the ones used in our study, $\tilde{p}$ must be less than or equal to $(\ln T)^{1.8}$. Pukkila et al. (1990) have also provided some basis for fixing $\tilde{p}$ at $kT^{1/2}$, where k ranged from 0.5 to 1.5.

There are other methods that can be used for determining the lag length $\tilde{p}$. Koreisha and Pukkila (1987), for instance, ‘experimented with several order determination criteria to determine the lag length $\tilde{p}$, [but] found them to be unsatisfactory. For example, BIC (Schwartz, 1987) often chose an AR(3) structure to fit an MA(1) model with $\theta = 0.9$.' It may also be possible that cross-validation methods (Burman, 1989; Shao, 1993) could be useful in selecting the lag order $\tilde{p}$, but such methods can be computationally complex and expensive, thus making their use not always practical (Shao, 1993; Racine, 1997).

3. FINITE SAMPLE PROPERTIES

3.1. The simulation design

In this section, we contrast the forecast performance of 2SOLS procedure based on an AR($\tilde{p}$) approximation of the residual autocorrelation with the predictive performance of models obtained from OLS; estimated GLS (EGLS), i.e. correctly specified Ω with estimated parameters; as well as estimated GLS-AR($\tilde{p}$) representations of the serial correlation (EARGLS) which Koreisha and Fang (2004) demonstrated yield efficient forecasts. Here, we adopt the same, simple data-driven procedure for selection the autoregressive order ($\tilde{p}$) proposed in Fang and Koreisha (2004), by allowing $\tilde{p}$ to increase as the sample size increases, i.e. we set $\tilde{p} = [\sqrt{T}/2]$.

Using the SAS random number subroutine RANNOR, we generated, for sample sizes of 50 to 500 observations, 1,000 realizations for each of a variety of stationary and invertible Gaussian ARMA(p,q) structures with varying parameter values as the residuals of a regression model with one exogenous variable generated by an AR(1) process. The parameter values for the residual ARMA structures were chosen to not only conform with other previously published studies of Engle (1974), Pukkila et al. (1990), Zinde-Walsh and Galbraith (1991), and Fang and Koreisha (2004), but also to provide a representative set of examples of possible autocorrelated error structures in regression models.

Then, we created the regression model

$$y_t = 2.0 + 0.5x_t + a_t, \tag{8}$$

where the generating process for the exogenous variable x_t followed an AR(1) process, $(1 - \varphi B)x_t = w_t$, with $w_t \sim IN(0,1)$, and $E(a_t,w_s) = 0$, for all t and s, and $\varphi = \{0.0,0.5,0.9\}$. Only one set of random numbers was generated for each of the AR(1) model structures of the exogenous variable used in (8).

Breusch (1980) has shown that for a fixed regressor the distribution of $(\hat{\beta}_{EGLS} - \beta)/\sigma$ does not depend on β and σ^2. In addition, the result also holds if the covariance matrix is misspecified (Koreisha and Fang, 2001). This implies that in simulation studies, only one point in the parameter space (β,σ^2) needs to be considered for estimated EGLS and EARGLS. This also holds for 2SOLS if $\tilde{p}$ and T are sufficiently large.

Ten additional observations were generated for each sample size to evaluate the forecast performance of all methods. The relative predictive efficiencies among estimation methods based on predictive mean squared error (PMSE),

$$\hat{\xi}_{i/j}(T+L) \equiv \frac{\sum(\hat{y}^{(i)}_{T+L} - y_{T+L})^2}{\sum(\hat{y}^{(j)}_{T+L} - y_{T+L})^2},$$
$$i,j = \{\text{2SOLS, OLS, EGLS, EARGLS}\}, \quad \text{and } i \neq j, \tag{9}$$

where $\hat{y}^{(m)}_{T+L}$ represents the forecasted value based on method m for time T + L, y_{T+L} is the actual generated value (true model) at T + L, was calculated for four forecast horizons, L = {1,5,10}. A ratio less than 1 indicates that forecasts obtained from method i in (9) are more efficient than those generated from method j.

3.2. Simulation results

Tables I and II contrast selected predictive relative mean squared error efficiencies among 2SOLS and 3 other procedures: OLS, GLS based on the correct residual model structures but with estimated ARMA coefficients (EGLS and to be considered as the performance benchmark); and EARGLS based on an approximating AR($\tilde{p}$) correction with lag $\tilde{p}$, as with 2SOLS, set equal to the closest integer part of $\sqrt{T}/2$ (denoted as EARGLS). (We also experimented with other autoregressive corrective order (multiples of $\sqrt{T}$), but like Fang and Koreisha (2004), found that for sample sizes considered in the study, $\tilde{p}$ set at $[\sqrt{T}/2]$ yielded the best forecasts.) SAS procedures (PROC ARIMA and PROC AUTOREG) were used to obtain the GLS estimates from which the forecasts were generated. To provide an idea of the magnitude of the actual PMSE we also included in these tables the actual estimates for PMSE(2SOLS). For the sake of brevity and to avoid a great deal of repetitiveness, these tables do not include all permutations of sample sizes and parameterizations of the serial correlation structure. Additional results can be obtained from the authors website: http://darkwing.uoregon.edu/~sergiok/JTSA08.

In examining the results from the tables we see that the predictive efficiencies of estimated GLS (correct and approximating) procedures and of the 2SOLS approach are higher than those obtained from OLS for short and medium term horizons ($L \leq 5$) for practically all model structures and parameterizations. Of the few cases in which OLS generated forecasts with lower mean square error than 2SOLS, none were by more than 5% when $L \leq 5$. In fact, the differences in efficiencies in most, if not all these cases (particularly in comparison to EGLS and EARGLS, cannot be distinguished from sampling variation.

For these horizons the degree of improvement in the relative predictive efficiency of 2SOLS (as well as EGLS and EARGLS) vis-à-vis OLS depends on the structure of the serial correlation. The improvement in predictive efficiency of 2SOLS over OLS, for instance, as expected, ranges from comparable for error processes which are close to a white noise, such as the AR(1) and MA(1) parameterizations with $\phi_1 = \theta_1 = 0.5$ and the ARMA(1,1) structure with

TABLE I

Relative Predictive Efficiencies Associated with AR(1), MA(1), and ARMA(1,1) Error Processes

		$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
			Efficiency $\hat{\xi}$				Efficiency $\hat{\xi}$				Efficiency $\hat{\xi}$		
		PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS
AR(1)													
ϕ_1	L												
(−0.90)	1	1.123	1.006	0.997	0.231	1.104	1.050	1.061	0.197	1.112	1.005	1.003	0.210
T = 100	5	3.678	1.086	1.089	0.668	3.875	1.023	0.980	0.703	3.659	1.032	1.025	0.695
	10	4.728	1.048	1.050	0.879	4.739	1.131	1.136	0.882	4.730	1.092	1.013	0.908
(−0.50)	1	1.184	1.043	1.047	0.900	1.083	1.063	0.982	0.725	1.080	1.011	1.005	0.795
T = 100	5	1.386	1.018	1.006	1.003	1.358	1.072	1.054	1.022	1.368	1.062	1.015	1.017
	10	1.378	1.001	0.989	0.975	1.359	1.026	1.026	1.035	1.353	1.020	1.038	1.001
(0.50)	1	1.137	1.057	1.052	0.744	1.177	1.117	1.035	0.922	1.117	1.084	0.994	0.776
T = 50	5	1.472	1.029	0.998	0.987	1.470	1.002	1.001	0.955	1.436	0.992	0.997	0.981
	10	1.482	0.996	1.001	1.000	1.540	0.974	1.026	1.083	1.390	1.047	1.020	1.005
T = 100	1	1.096	1.070	1.073	0.796	1.021	1.091	1.010	0.772	1.059	1.018	1.005	0.799
	5	1.417	1.096	1.085	0.982	1.440	0.995	1.021	0.990	1.430	1.002	1.054	1.018
	10	1.446	1.083	1.062	1.044	1.445	1.010	0.986	1.028	1.428	1.061	1.032	1.003
T = 200	1	1.022	1.001	1.002	0.761	1.023	1.024	1.031	0.770	1.015	1.040	1.009	0.739
	5	1.339	1.028	1.024	0.962	1.365	1.100	1.051	1.043	1.410	1.010	1.010	1.028
	10	1.377	1.029	1.025	0.963	1.378	1.040	1.042	1.024	1.380	1.005	0.998	1.002
MA(1)													
θ_1	L												
(0.90)	1	1.118	1.071	1.054	0.613	1.214	1.075	1.067	0.698	1.221	1.102	1.096	0.661
T = 100	5	1.819	1.007	1.008	0.963	1.812	1.038	1.040	0.990	1.845	1.013	0.999	1.033
	10	1.859	1.027	1.024	0.989	1.956	1.011	1.002	1.065	1.853	1.055	1.032	0.999
(0.50)	1	1.075	1.083	1.046	0.902	1.085	1.010	0.969	0.782	1.087	1.092	1.049	0.847
T = 100	5	1.258	1.000	0.966	0.904	1.318	1.003	0.998	0.993	1.286	1.003	0.997	1.002
	10	1.277	1.009	0.973	0.950	1.376	0.989	0.976	0.991	1.271	1.005	1.014	0.998

TABLE I
CONTINUED

		$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
		PMSE (2SOLS)	Efficiency $\hat{\xi}$			PMSE (2SOLS)	Efficiency $\hat{\xi}$			PMSE (2SOLS)	Efficiency $\hat{\xi}$		
			$\frac{2SOLS}{EGLS}$	$\frac{2SOLS}{EARGLS}$	$\frac{2SOLS}{OLS}$		$\frac{2SOLS}{EGLS}$	$\frac{2SOLS}{EARGLS}$	$\frac{2SOLS}{OLS}$		$\frac{2SOLS}{EGLS}$	$\frac{2SOLS}{EARGLS}$	$\frac{2SOLS}{OLS}$
(−0.50)	1	1.135	1.120	1.071	0.931	1.147	1.133	1.070	0.853	1.140	1.063	1.060	0.997
T = 50	5	1.327	1.015	0.986	0.991	1.279	1.022	1.014	0.934	1.286	1.002	1.000	0.995
	10	1.335	1.026	0.990	1.002	1.328	1.058	1.054	0.995	1.288	1.009	1.010	1.009
T = 100	1	1.086	1.017	1.022	0.830	1.081	0.988	1.016	0.807	1.095	1.008	1.001	0.863
	5	1.298	0.993	0.972	0.961	1.334	1.064	1.041	0.977	1.311	1.058	1.045	1.017
	10	1.295	0.956	0.964	0.938	1.315	1.040	1.037	0.969	1.309	1.060	1.024	1.002
T = 200	1	1.022	1.028	0.996	0.766	1.023	1.006	1.009	0.763	1.024	1.031	1.004	0.800
	5	1.258	1.054	0.998	0.940	1.276	1.056	1.017	1.011	1.306	1.052	1.001	1.026
	10	1.276	1.027	0.998	0.955	1.277	1.008	1.013	0.923	1.277	1.001	1.040	1.002
ARMA(1,1) (ϕ_1,θ_1)	L												
(0.8,0.5)	1	1.079	1.067	1.077	0.951	1.067	1.069	1.076	1.009	1.058	1.013	1.010	1.020
T = 100	5	1.083	0.999	1.003	0.957	1.097	0.987	0.989	0.968	1.075	0.994	0.996	1.004
	10	1.081	0.996	0.999	1.015	1.115	0.990	0.988	0.989	1.094	1.013	1.001	0.987
(−0.8,−0.7)	1	1.080	1.079	1.083	1.028	1.046	1.034	1.012	0.958	1.046	1.008	1.006	0.997
T = 100	5	1.067	0.991	0.995	0.970	1.075	1.033	0.994	0.979	1.060	1.005	1.012	0.998
	10	1.105	1.005	0.994	0.977	1.072	1.027	0.989	0.982	1.109	1.010	1.001	1.005
(−0.8,0.7)	1	1.255	1.022	1.016	0.166	1.181	1.090	1.023	0.174	1.313	1.005	1.005	0.174
T = 50	5	7.081	1.008	0.990	0.915	6.670	1.033	0.987	0.882	6.661	0.999	0.988	0.937
	10	7.342	1.030	1.003	0.959	7.459	1.004	0.999	1.008	7.326	1.016	1.008	1.001
T = 100	1	1.270	1.133	1.090	0.165	1.148	1.055	1.060	0.152	1.313	1.026	1.016	0.174
	5	6.484	0.991	0.962	0.875	6.537	1.021	1.013	0.832	6.661	1.059	1.050	0.937
	10	7.186	1.018	1.006	1.058	7.205	1.044	0.998	0.945	7.329	0.996	0.988	1.001
T = 200	1	1.066	1.028	1.009	0.148	1.071	1.076	1.068	0.149	1.215	1.079	1.047	0.167
	5	6.554	1.035	1.040	0.879	6.076	1.075	1.043	0.893	6.485	1.017	1.010	0.904
	10	7.155	1.023	1.004	0.924	7.152	1.023	1.014	1.006	7.197	1.018	1.003	0.999

Notes: PMSE: prediction mean squared error. The order $\tilde{p}$ of the AR correction for EARGLS and 2SOLS is set at $[\sqrt{T}/2]$ for all cases.

TABLE II

RELATIVE PREDICTIVE EFFICIENCIES ASSOCIATED WITH AR(2), MA(2), ARMA(1,2), AND ARMA(2,1) ERROR PROCESSES

		$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
			Efficiency $\hat{\xi}$				Efficiency $\hat{\xi}$				Efficiency $\hat{\xi}$		
		PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS
AR(2) (ϕ_1,ϕ_2)	L												
(1.42,−0.73) T = 100	1	1.165	1.074	1.082	0.179	1.165	1.106	1.101	0.172	1.101	1.013	0.999	0.169
	5	5.900	1.010	1.008	0.855	5.905	1.046	1.044	0.754	6.000	1.051	1.034	0.876
	10	7.362	1.003	1.002	1.002	6.901	1.015	1.021	0.899	6.993	1.049	1.015	0.998
(1.8,−0.9) T = 50	1	1.433	1.247	1.117	0.028	1.630	1.469	1.296	0.033	1.685	1.532	1.340	0.033
	5	29.589	1.068	1.040	0.514	30.916	1.112	1.083	0.525	31.239	1.120	1.089	0.465
	10	43.134	1.041	1.022	0.760	43.901	1.034	1.040	0.751	45.844	1.106	1.086	0.708
T = 100	1	1.258	1.099	1.067	0.024	1.534	1.398	1.342	0.029	1.479	1.348	1.294	0.028
	5	28.043	1.050	1.022	0.470	28.959	1.083	1.053	0.475	28.982	1.085	1.054	0.469
	10	40.912	1.017	1.002	0.739	41.607	1.034	1.020	0.743	42.646	1.061	1.046	0.741
T = 200	1	1.091	1.062	1.034	0.021	1.095	1.067	1.039	0.021	1.152	1.124	1.094	0.022
	5	24.794	1.018	1.000	0.470	25.621	1.052	1.034	0.486	25.794	1.060	1.041	0.473
	10	34.477	1.003	1.001	0.679	35.074	1.020	1.018	0.679	35.583	1.037	1.034	0.681
MA(2) (θ_1,θ_2)	L												
(1.42,−0.73) T = 50	1	1.483	1.140	1.096	0.387	1.415	1.079	1.077	0.376	1.207	1.111	1.101	0.337
	5	3.851	0.993	0.990	0.993	3.634	1.083	1.029	1.018	3.687	1.012	0.999	1.053
	10	4.171	1.017	1.013	1.077	3.622	1.051	1.029	1.030	3.610	1.002	0.989	1.006
T = 100	1	1.232	1.135	1.044	0.331	1.310	1.129	1.105	0.383	1.103	1.062	1.010	0.307
	5	3.580	1.039	1.039	0.973	3.598	1.019	1.030	1.010	3.571	1.086	1.041	1.031
	10	3.628	1.028	1.021	0.979	3.638	1.032	1.035	1.024	3.638	1.040	1.025	1.005
T = 200	1	1.082	1.056	1.046	0.303	1.139	0.995	0.966	0.317	1.069	1.035	1.028	0.313
	5	3.684	1.062	1.062	1.033	3.694	1.135	1.094	1.007	3.697	1.057	1.003	1.032
	10	3.521	0.999	0.993	1.004	3.523	1.015	1.013	0.963	3.522	1.009	1.025	1.008

TABLE II
CONTINUED

		$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
		PMSE (2SOLS)	Efficiency $\hat{\xi}$ 2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	Efficiency $\hat{\xi}$ 2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS	PMSE (2SOLS)	Efficiency $\hat{\xi}$ 2SOLS/EGLS	2SOLS/EARGLS	2SOLS/OLS
(1.6,−0.64)	1	1.492	1.099	1.057	0.360	1.328	1.073	1.010	0.315	1.161	1.076	1.010	0.289
T = 100	5	4.102	0.998	0.988	0.961	4.029	0.999	0.980	0.968	4.005	1.003	0.999	1.037
	10	4.168	0.999	1.002	0.983	4.113	0.969	0.998	0.993	4.091	1.024	1.011	1.006
ARMA(1,2) (ϕ_1,θ_1,θ_2)	L												
(−0.8;1.4,−0.6)	1	1.569	1.120	1.049	0.069	1.659	1.104	1.062	0.071	1.768	1.103	1.101	0.074
T = 50	5	19.944	1.004	0.968	0.960	19.057	1.042	0.969	0.872	19.021	1.001	0.984	0.915
	10	21.470	1.030	1.031	1.041	22.563	1.000	1.001	1.083	21.519	1.001	1.005	1.005
T = 100	1	1.436	1.095	1.039	0.061	1.337	1.084	1.025	0.058	1.387	1.089	1.090	0.069
	5	18.310	1.083	1.012	0.777	18.507	1.144	1.079	0.763	18.334	1.059	0.995	0.870
	10	21.035	1.021	1.004	0.896	21.192	0.998	1.004	0.913	21.185	1.070	0.984	1.007
T = 200	1	1.138	1.056	0.961	0.050	1.088	1.007	1.018	0.051	1.227	1.049	1.059	0.057
	5	18.385	0.995	0.997	0.823	17.420	1.067	1.052	0.737	18.708	1.028	1.033	0.868
	10	21.135	1.039	1.037	1.025	21.142	1.045	1.039	0.934	21.257	0.996	0.996	1.011
ARMA(2,1) (ϕ_1,ϕ_2,θ_1)	L												
(−0.5,−0.9;0.6)	1	1.445	1.082	1.087	0.143	1.270	1.016	1.023	0.143	1.504	1.076	1.091	0.169
T = 50	5	4.416	1.101	1.040	0.457	4.369	1.037	1.044	0.494	4.410	1.102	1.067	0.455
	10	7.196	1.055	0.996	0.753	6.893	1.001	1.010	0.773	6.970	1.019	0.971	0.752
T = 100	1	1.167	1.074	1.072	0.125	1.199	1.085	1.086	0.121	1.176	1.120	1.075	0.120
	5	3.940	1.032	0.981	0.419	4.045	1.024	1.025	0.375	4.006	1.015	0.995	0.432
	10	6.768	1.110	1.033	0.743	6.736	0.983	0.981	0.719	6.684	1.073	1.043	0.686
T = 200	1	1.096	1.092	1.100	0.113	1.012	1.005	1.008	0.096	1.056	1.065	1.016	0.118
	5	3.780	1.054	0.980	0.410	3.816	1.000	1.001	0.365	3.876	1.001	1.003	0.434
	10	6.259	1.086	1.013	0.712	6.262	0.957	0.964	0.689	7.254	1.021	1.013	0.782

Notes: PMSE: prediction mean squared error. The order $\tilde{p}$ of the AR correction for EARGLS and 2SOLS is set at $[\sqrt{T}/2]$ for all cases.

$\phi_1 = 0.8$ and $\theta_1 = 0.7$ (Table I), to two to ten times for error processes which have strong autocorrelations, such as some of the mixed ARMA(p, q) parameterizations in Table II.

In general, for short to medium forecast horizons ($L \leq 5$), the predictive efficiency of the 2SOLS method is comparable to that of EGLS and EARGLS with $\tilde{p} = [\sqrt{T}/2]$ and does not seem to depend on φ. For the majority of error structures, we see that $\hat{\xi}_{\frac{2SOLS}{EGLS}}$ and $\hat{\xi}_{\frac{2SOLS}{EARGLS}}$ are much less than 1.08. The most notable exception to this observation occurs for the nearly nonstationary AR(2) serial correlation structure with $(\phi_1, \phi_2) = (1.8, -0.9)$ when $L = 1$ and $T \leq 100$ observations (Table II). In this case increasing the sample size improves the predictive performance of the 2SOLS procedure. When $T = 500$ for $(\phi_1, \phi_2) = (1.8, -0.9)$ the predictive efficiency of 2SOLS relative to EARGLS, i.e. $\hat{\xi}_{\frac{2SOLS}{EARGLS}}(T+1)$, decreases to 0.997, 1.002 and 1.017 for $\varphi = 0, 0.5$, and 0.9 respectively (derived from Table III). The corresponding relative efficiencies, $\hat{\xi}_{\frac{2SOLS}{EGLS}}(T+1)$ are respectively 1.023, 1.028 and 1.043. As the forecast horizon increases even for this serial correlation structure the differences in the PMSEs become almost trivial, i.e. for $\varphi = 0,0.5$ and 0.9, $\hat{\xi}_{\frac{2SOLS}{EARGLS}}(T+5)$ are respectively 0.988, 1.016 and 1.011, and respectively 1.006, 1.035 and 1.030 for $\hat{\xi}_{\frac{2SOLS}{EGLS}}(T+5)$. It is also interesting to note that for this AR(2) parameterization the predictive efficiency of 2SOLS is more than an order of magnitude higher than that of OLS for all values of φ.

As the forecast horizon increases the differences in predictive efficiencies decrease among all methods. As L reaches 10, $\hat{\xi}_{i/j}$ nearly always approaches 1 regardless of the value of φ or sample size. For example, consider the case of the ARMA(1,1) error structure with $(\phi_1, \theta_1) = (-0.8, 0.7)$ in Table I when $\varphi = 0.5$ and $T = 200$; $\hat{\xi}_{\frac{2SOLS}{OLS}}$ changes from 0.149 when $L = 1$, to 0.444 when $L = 2$ (not shown in the table), to 0.893 when $L = 5$, and approaches 1 when $L = 10$.

To gain additional insights on the relative causes of the forecast inefficiencies we have decomposed the predictive mean square errors associated with 2SOLS, EGLS and EARGLS into their component parts, i.e.

$$\begin{aligned}
\mathrm{MSE}_i &= E[(\beta_0 + \beta_1 x_t + a_t) - (\hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{a}_t)]^2 = E\{[(\beta_0 + \beta_1 x_t) \\
&\quad - (\hat{\beta}_0 + \hat{\beta}_1 x_t)] + (a_t - \hat{a}_t)\}^2 \\
&= E[(\beta_0 + \beta_1 x_t) - (\hat{\beta}_0 + \hat{\beta}_1 x_t)]^2 + E(a_t - \hat{a}_t)^2 + 2E\{[(\beta_0 + \beta_1 x_t) \\
&\quad - (\hat{\beta}_0 + \hat{\beta}_1 x_t)](a_t - \hat{a}_t)\} \\
&\equiv \mathrm{MSEReg}_i + \mathrm{MSEAuto}_i + \mathrm{MSECross}_i,
\end{aligned}$$

where i = {2SOLS, EGLS, EARGLS}.

Table III contrasts the decomposed predictive mean square errors for some selected autocorrelation structures among the various estimation methods for sample sizes of 100 and 500 observations. As can be seen across all structures and estimation methods the dominant component of the predictive mean square error is the portion attributable to the serial correlation. For any given sample size, the magnitude of $\mathrm{MSE\text{-}Auto_i}$ increases as the forecast horizon increases

TABLE III

Decomposed Predictive Mean Square Errors For Some Selected Autocorrelation Structures

	$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
	MSE-Total	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross
Panel A. T = 100												
					AR(2) (1.8,–0.9)							
L = 1												
EGLS	1.145	0.971	2.133	–1.959	1.098	0.947	1.679	–1.528	1.097	0.976	1.706	–1.584
EARGLS	1.179	0.975	2.163	–1.959	1.143	0.951	1.719	–1.527	1.143	0.979	1.746	–1.582
2SOLS	1.258	1.056	2.274	–2.071	1.534	1.176	2.088	–1.730	1.479	1.361	1.821	–1.703
L = 5												
EGLS	26.705	0.905	24.015	1.786	26.750	0.950	24.805	0.995	26.721	0.965	24.843	0.913
EARGLS	27.426	0.955	25.431	1.040	27.498	0.956	25.503	1.038	27.501	0.973	25.568	0.960
2SOLS	28.043	1.029	26.010	1.005	28.959	1.246	25.654	2.059	28.982	1.956	26.189	0.837
L = 10												
EGLS	40.242	0.944	37.912	1.386	40.220	0.947	37.894	1.380	40.212	0.972	37.865	1.376
EARGLS	40.835	0.948	38.373	1.513	40.804	0.951	38.338	1.515	40.783	0.980	38.308	1.495
2SOLS	40.912	1.043	38.622	1.246	41.607	1.565	38.731	1.311	42.646	1.373	39.236	2.036
					ARMA(1,1) (0.8,–0.7)							
L = 1												
EGLS	1.082	0.596	1.426	–0.940	1.082	0.600	1.430	–0.948	1.085	0.632	1.462	–1.009
EARGLS	1.105	0.607	1.438	–0.940	1.106	0.611	1.440	–0.944	1.108	0.640	1.468	–0.999
2SOLS	1.171	0.669	1.532	–1.030	1.256	0.750	1.619	–1.112	1.277	1.051	1.828	–1.602
L = 5												
EGLS	6.597	0.605	6.248	–0.257	6.591	0.605	6.249	–0.263	6.581	0.623	6.250	–0.292
EARGLS	6.821	0.616	6.413	–0.208	6.820	0.618	6.415	–0.213	6.818	0.635	6.419	–0.236
2SOLS	6.862	0.670	6.468	–0.276	6.949	0.716	6.493	–0.260	7.190	1.046	6.548	–0.404
L = 10												
EGLS	8.381	0.597	7.780	0.004	8.384	0.599	7.781	0.004	8.431	0.625	7.784	0.021
EARGLS	8.563	0.609	7.913	0.041	8.558	0.611	7.908	0.039	8.588	0.640	7.909	0.040
2SOLS	8.580	0.674	7.973	–0.067	8.713	0.761	7.983	–0.031	9.127	1.135	8.002	–0.010

TABLE III

CONTINUED

	$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
	MSE-Total	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross
	ARMA(2,1) (1.4,–0.6;–0.8)											
L = 1												
EGLS	1.110	0.722	1.541	−1.153	1.110	0.718	1.544	−1.153	1.111	0.734	1.562	−1.185
EARGLS	1.167	0.741	1.577	−1.151	1.169	0.742	1.578	−1.151	1.170	0.753	1.590	−1.173
2SOLS	1.385	0.802	1.814	−1.230	1.445	0.934	1.652	−1.141	1.418	1.100	1.903	−1.584
L = 5												
EGLS	21.132	0.728	19.788	0.616	21.120	0.725	19.783	0.612	21.142	0.734	19.805	0.603
EARGLS	21.971	0.747	20.483	0.741	21.948	0.748	20.467	0.733	21.930	0.753	20.450	0.727
2SOLS	22.031	0.794	20.559	0.677	22.301	0.970	20.550	0.780	22.894	1.630	20.604	0.659
L = 10												
EGLS	22.437	0.722	21.594	0.120	22.439	0.717	21.597	0.124	22.496	0.727	21.598	0.170
EARGLS	23.386	0.742	22.495	0.149	23.378	0.741	22.496	0.140	23.409	0.750	22.498	0.161
2SOLS	23.213	0.806	22.430	−0.023	23.513	0.845	22.301	0.366	24.583	1.768	22.502	0.312
Panel B. T = 500												
	AR(2) (1.8,–0.9)											
L = 1												
EGLS	1.023	0.217	1.215	−0.409	1.024	0.218	1.219	−0.412	1.026	0.223	1.232	−0.429
EARGLS	1.050	0.217	1.245	−0.412	1.051	0.218	1.249	−0.416	1.052	0.224	1.264	−0.435
2SOLS	1.047	0.221	1.243	−0.417	1.053	0.310	1.500	−0.756	1.070	0.524	1.634	−1.088
L = 5												
EGLS	23.497	0.217	23.240	0.040	23.487	0.217	23.232	0.037	23.489	0.222	23.248	0.019
EARGLS	23.935	0.218	23.662	0.055	23.926	0.218	23.657	0.051	23.917	0.223	23.669	0.025
2SOLS	23.638	0.222	23.401	0.015	24.299	0.314	23.926	0.059	24.191	0.594	23.860	−0.263
L = 10												
EGLS	34.666	0.219	34.195	0.252	34.688	0.218	34.197	0.272	34.750	0.221	34.215	0.314
EARGLS	35.115	0.219	34.623	0.273	35.134	0.219	34.621	0.294	35.196	0.222	34.640	0.334
2SOLS	34.685	0.224	34.243	0.219	34.747	0.330	34.232	0.186	34.858	0.537	34.088	0.233

TABLE III

CONTINUED

	$\varphi = 0.0$				$\varphi = 0.5$				$\varphi = 0.9$			
	MSE-Total	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross	**MSE-Total**	MSE-Reg	MSE-Auto	MSE-Cross
					ARMA(1,1) (0.8,–0.7)							
L = 1												
EGLS	1.028	0.153	1.162	–0.287	1.028	0.153	1.163	–0.289	1.028	0.159	1.174	–0.305
EARGLS	1.048	0.153	1.180	–0.286	1.048	0.154	1.182	–0.288	1.049	0.160	1.194	–0.305
2SOLS	1.050	0.158	1.180	–0.289	1.079	0.172	1.224	–0.318	1.064	0.242	1.272	–0.449
L = 5												
EGLS	6.372	0.153	6.323	–0.105	6.371	0.153	6.323	–0.106	6.366	0.159	6.327	–0.119
EARGLS	6.549	0.154	6.477	–0.083	6.547	0.154	6.477	–0.084	6.539	0.160	6.478	–0.099
2SOLS	6.501	0.159	6.446	–0.104	6.526	0.177	6.456	–0.107	6.549	0.268	6.483	–0.202
L = 10												
EGLS	6.926	0.154	6.802	–0.030	6.934	0.154	6.802	–0.022	6.946	0.158	6.801	–0.013
EARGLS	7.122	0.155	6.981	–0.014	7.130	0.155	6.981	–0.006	7.145	0.159	6.984	0.002
2SOLS	7.045	0.160	6.942	–0.057	7.049	0.174	6.939	–0.064	7.063	0.252	6.953	–0.142
					ARMA(2,1) (1.4,–0.6;–0.8)							
L = 1												
EGLS	1.029	0.175	1.182	–0.328	1.028	0.175	1.182	–0.328	1.028	0.177	1.186	–0.334
EARGLS	1.050	0.175	1.204	–0.329	1.050	0.175	1.205	–0.330	1.050	0.178	1.209	–0.337
2SOLS	1.094	0.179	1.256	–0.341	1.094	0.226	1.279	–0.411	1.099	0.329	1.482	–0.712
L = 5												
EGLS	19.161	0.175	18.941	0.046	19.159	0.175	18.940	0.044	19.154	0.177	18.940	0.038
EARGLS	19.691	0.175	19.450	0.065	19.688	0.175	19.449	0.063	19.678	0.177	19.447	0.053
2SOLS	19.471	0.180	19.227	0.065	19.621	0.232	19.316	0.073	19.569	0.373	19.259	–0.064
L = 10												
EGLS	19.641	0.175	19.452	0.013	19.647	0.175	19.453	0.019	19.662	0.176	19.455	0.031
EARGLS	20.039	0.176	19.847	0.017	20.046	0.176	19.847	0.023	20.061	0.177	19.850	0.034
2SOLS	20.079	0.182	19.942	–0.045	19.948	0.227	19.807	–0.085	20.074	0.340	19.914	–0.180

Note: The order $\tilde{p}$ of the AR correction for EARGLS and 2SOLS is set at $[\sqrt{T}/2]$ for all cases.

whereas, as expected, the magnitude of MSE-Reg$_i$ remains relatively constant as the forecast horizon increases. It is also important to note that the cross term component (MSE-Cross$_i$) plays a significant role in the adjustment (negative) of the total predictive mean square error for short forecasting horizons. For long horizons such as $L = 10$, its role is minor. MSE-Auto$_i$ overshadows the other components. In fact, the ratio of the MSE-Auto$_i$ to MSE-Total$_i$ typically exceeds 92% across all estimation procedures. The stochastic nature of the exogenous variable does not appear to have an impact on the relative magnitude of the predictive error components. Finally, as anticipated by Theorem 2, MSE-Reg$_i$ decreases remarkably fast as T increases from 100 to 500 observations, e.g. for the ARMA(2,1) parametrization when $L = 1$ and $\varphi = 0$, MSE-Reg$_{2SOLS}$ changes from 0.802 to 0.179, MSE-Reg$_{EARGLS}$ changes from 0.741 to 0.175, and MSE-Reg$_{EGLS}$ changes from 0.722 to 0.175. The corresponding changes for $\varphi = 0.9$ are respectively, from 1.100 to 0.329, from 0.753 to 0.178, and 0.734 to 0.177.

4. EMPIRICAL EXAMPLES

We also compared the forecasting performance of the 2SOLS procedure with other methods using real economic and business data. The data used to construct the models below can be found at the specified sources and more conveniently at the authors' website (http://darkwing.uoregon.edu/~sergiok/JTSA08).

EXAMPLE 1. Interest Rates, Aggregate Demand, and Liquid Assets.

Pindyck and Rubinfeld (1998) constructed a model to explain and forecast the movement of monthly interest rates, $R3_t$ (3-month US Treasury bill rate, percent per year) based on industrial production, IP_t (Federal Reserve Board Index, $1987 = 100$); the rate of growth of nominal money supply, $GM2_t$ ($(M2_t - M2_{t-1})/M2_{t-1}$); and the lagged rate of wholesale price inflation, GPW_{t-1} ($GPW_t = (PW_t - PW_{t-1})/PW_{t-1}$, where PW is the producer price index for all commodities). Noting that the residuals from the OLS regression (from January 1960 to August 1995) relating these variables were serially correlated, Pindyck and Rubinfeld estimated a combined regression-time-series model using an ARMA(8,2) process for the residuals. Because they did not tabulate the actual forecasts generated from their model [the forecasts were only reported in graphical form (p. 600)], we had to re-estimate their model to generate predictions. Our estimates differ slightly from those reported by Pindyck and Rubinfeld. This may be due to the fact we used a more modern transfer function program with maximum likelihood procedures (Scientific Computing Associates) than they had available at the time they estimated the serially correlated regression model.

Table IV contrasts the forecasts generated by our re-estimated transfer function model with those obtained from 2SOLS ($\tilde{p} = 10$) and OLS for the same 6 out-of-sample periods used by Pindyck and Rubinfeld. As can be seen for all six horizons the 2SOLS ($\tilde{p} = 10$) forecasts tracked the actual series much more closely than all forecasts generated by the other methods. The predictive efficiency (based on average MSE) of the 2SOLS ($\tilde{p} = 10$) forecasts is more than 50% higher (better) than that of the forecasts generated by the transfer function model with an ARMA(8,2) serial correlation correction, and more than an order of magnitude higher than of the OLS forecasts.

Example 2. New York & London IBM Stock Prices.

Since the London Stock Exchange closes earlier than the New York Stock Exchange DeLurgio (1998) attempted to predict the closing price of IBM stock in New York, $\mathrm{NYP_t}$, using the closing price in London, $\mathrm{LONP_t}$. After differencing the data covering the period from 12/31/93 to 12/5/94 (242 observations) to remove trends and carrying out appropriate prewhitening and pretreatment of the series, he discovered that the 'direction of the causality [was] in the opposite direction than [originally] hypothesized.' An analysis of the cross correlations between these series indicated that New York price changes preceded London price changes by one period. Noting that the residuals from the simple model relating the changes in prices were autocorrelated, he estimated the following transfer function model,

$$\nabla \mathrm{LONP}_t = \underset{(0.0172)}{0.6205}\nabla \mathrm{NYP}_{t-1} + (1 - \underset{(0.0559)}{0.5067}B)a_t,$$

where ∇ is the difference operator.

We replicated his results using the entire set of observations, and then re-estimated the model leaving out 10 observations to compare the forecasts generated by the transfer function model with those obtained from 2SOLS and OLS. Table V contains the one-step-ahead as well as multi-step-ahead forecasts generated by the three approaches. (Because the transfer function model was formulated in terms of differences, forecasts for future London prices require using either predicted or actual London prices for the immediately preceding period. The one-step ahead forecasts were generated using actual data while the multi-step ahead forecasts were obtained from predicted values.) The predictive efficiency (based on average MSE) of the 2SOLS ($\tilde{p} = 8$) one-step-ahead forecasts is almost 50% higher (better) than that of the one-step-ahead forecasts generated by the transfer function model, and the predictive efficiency of the 2SOLS ($\tilde{p} = 8$) multi-step-ahead forecasts is more than an order of magnitude higher than of corresponding transfer function forecasts. Although the average mean square error of the one-step-ahead predictions based on OLS appears to be comparable to that of 2SOLS, the average mean square error of the OLS multi-step-ahead forecasts is more than 50% larger (worst) than the average generated from the 2SOLS ($\tilde{p} = 8$) forecasts. The transfer function forecasts were quite accurate for

TABLE IV

Comparison of the Models and Forecasts Generated for Monthly Interest Rates (From Sep. 95 to Feb. 96)

Variable	Value	STD Error	T-value
Panel A. OLS $R3_t = \beta_0 + \beta_1 IP_t + \beta_2 GM2_t + \beta_3 GPW_{t-1} + a_t$			
CONST	1.1731	0.5478	2.141
IP_t	0.0484	5.46E-03	8.850
$GM2_t$	142.8995	35.7825	3.994
GPW_{t-1}	104.4443	17.3183	6.031
Panel B. 2SOLS(10) $R3_t = \beta_0 + \beta_1 IP_t + \beta_2 GM2_t + \beta_3 GPW_{t-1} + \sum_{j=1}^{\tilde{p}} \gamma_j \hat{a}_{t-j} + a_t$			
CONST	2.1844	0.2234	9.780
IP_t	0.0438	2.12E-03	20.638
$GM2_t$	57.3354	15.0601	3.807
GPW_{t-1}	58.0009	6.5693	8.830
$\hat{a}_{t-1}$	0.6917	0.0472	14.639
$\hat{a}_{t-2}$	0.1150	0.0547	2.101
$\hat{a}_{t-3}$	0.0699	0.0549	1.274
$\hat{a}_{t-4}$	−0.0682	0.0552	−1.235
$\hat{a}_{t-5}$	0.1733	0.0550	3.151
$\hat{a}_{t-6}$	−0.0910	0.0549	−1.658
$\hat{a}_{t-7}$	−0.0165	0.0548	−0.301
$\hat{a}_{t-8}$	0.1316	0.0545	2.414
$\hat{a}_{t-9}$	0.1153	0.0548	2.105
$\hat{a}_{t-10}$	−0.1309	0.0451	−2.902
Panel C. Pindyck and Rubinfeld transfer function* $R3_t = \beta_0 + \beta_1 IP_t + \beta_2 GM2_t + \beta_3 GPW_{t-1} + (1 - \theta_1 B - \theta_2 B^2)/(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_8 B^8) a_t$			
CONST	−1.8423	9.7921	−0.188
IP_t	0.1722	0.0400	4.305
$GM2_t$	−30.9312	7.9909	−3.871
GPW_{t-1}	5.7963	2.0811	2.785
θ_1	0.8201	0.1982	4.138
θ_2	−0.4459	0.1395	−3.196
ϕ_1	2.1646	0.1961	11.038
ϕ_2	−2.1382	0.3242	−6.595
ϕ_3	1.3960	0.2665	5.238
ϕ_4	−0.6855	0.1910	−3.589
ϕ_5	0.5112	0.1668	3.065
ϕ_6	−0.6379	0.1568	−4.068
ϕ_7	0.6131	0.1351	4.538
ϕ_8	−0.2254	0.0694	−3.248

Panel D. Predictions

			Pindyck-Rubinfeld		2SOLS(10)		OLS	
l	Period	Actual	Forecast	MSE	Forecast	MSE	Forecast	MSE
1	Sep. 95	0.0528	0.0550	4.84E-06	0.0519	7.63E-07	0.0746	4.77E-04
2	Oct. 95	0.0528	0.0556	7.84E-06	0.0527	1.81E-08	0.0744	4.67E-04
3	Nov. 95	0.0536	0.0563	7.29E-06	0.0529	5.54E-07	0.0761	5.06E-04
4	Dec. 95	0.0514	0.0563	2.40E-05	0.0543	8.41E-06	0.0776	6.85E-04
5	Jan. 96	0.0500	0.0562	3.84E-05	0.0547	2.21E-05	0.0782	7.96E-04
6	Feb. 96	0.0483	0.0600	1.37E-04	0.0561	6.09E-05	0.0825	1.17E-03
Ave MSE				3.66E-05		1.55E-05		6.83E-04

Notes: *Based on their estimates Pindyck and Rubinfeld noted that the out-of-sample 6 month forecasts "track[ed] the actual series until the last 2 months, when it diverge[d] significantly."

TABLE V

COMPARISON OF THE MODELS AND FORECASTS GENERATED FOR DAILY IBM STOCK PRICES (BETWEEN 12/31/93 AND 12/5/94)

Variable	Value	STD Error	T-value
	Panel A. OLS $\nabla \mathrm{LONP_t} = \beta_0 + \beta_1 \nabla \mathrm{NYP_{t-1}} + \mathrm{a_t}$		
CONST	−0.0137	0.0243	−0.565
$\nabla \mathrm{NYP_{t-1}}$	0.6216	0.0219	28.390
	Panel B. 2SOLS(8) $\nabla LONP_t = \beta_0 + \beta_1 \nabla NYP_{t-1} + \sum_{j=1}^{\tilde{p}} \gamma_j \hat{a}_{t-j} + a_t$		
CONST	−0.0046	0.0223	−0.208
$\nabla \mathrm{NYP_{t-1}}$	0.6288	0.0200	31.403
$\hat{a}_{t-1}$	−0.4919	0.0690	−7.126
$\hat{a}_{t-2}$	−0.3275	0.0769	−4.258
$\hat{a}_{t-3}$	−0.1520	0.0802	−1.894
$\hat{a}_{t-4}$	−0.0804	0.0816	−0.985
$\hat{a}_{t-5}$	0.0707	0.0824	0.858
$\hat{a}_{t-6}$	−0.0859	0.0828	−1.037
$\hat{a}_{t-7}$	−0.0154	0.0787	−0.196
$\hat{a}_{t-8}$	−0.0029	0.0698	−0.041
	Panel C. DeLurgio's transfer function $\nabla \mathrm{LONP_t} = \mathrm{w_0} \nabla \mathrm{NYP_{t-1}} + (1 - \theta_1 \mathrm{B})\mathrm{a_t}$		
$\mathrm{w_0}$	0.6193	0.0178	34.792
θ_1	0.5065	0.0576	8.793

Panel D. One-step ahead predictions

		DeLurgio		2SOLS(8)		OLS	
Period	Actual	Forecast	MSE	Forecast	MSE	Forecast	MSE
11/22/94	46.5	46.1388	0.1305	46.0632	0.1908	45.8822	0.3817
11/23/94	44.25	44.2034	0.0022	44.6387	0.1511	44.5438	0.0863
11/24/94	44.438	44.3582	0.0064	44.3394	0.0097	44.3917	0.0021
11/25/94	44.939	44.5905	0.1215	44.7014	0.0565	44.6573	0.0793
11/28/94	45.438	44.8227	0.3786	45.0698	0.1355	45.1583	0.0782
11/29/94	45.125	44.7453	0.1442	45.4611	0.1130	45.3466	0.0491
11/30/94	45.25	44.7453	0.2547	45.1035	0.0215	45.1113	0.0193
12/1/94	45.063	44.8227	0.0577	45.2956	0.0541	45.3140	0.0630
12/2/94	44.438	44.0486	0.1516	44.2893	0.0221	44.2723	0.0275
12/5/94	45.75	45.2098	0.2918	45.5981	0.0231	45.5897	0.0257
Ave MSE			0.1539		0.0777		0.0812

Panel E. Multi-step ahead predictions

			DeLurgio		2SOLS(8)		OLS	
l	Period	Actual	Forecast	MSE	Forecast	MSE	Forecast	MSE
1	11/22/94	46.5	46.1388	0.1305	46.0632	0.1908	45.8822	0.3817
2	11/23/94	44.25	43.8422	0.1663	44.2020	0.0023	43.9260	0.1050
3	11/24/94	44.438	43.9504	0.2378	44.2914	0.0215	44.0677	0.1372
4	11/25/94	44.939	44.1029	0.6991	44.5547	0.1477	44.2870	0.4251
5	11/28/94	45.438	43.9866	2.1066	44.6856	0.5662	44.5063	0.8680
6	11/29/94	45.125	43.2939	3.3529	44.7087	0.1733	44.4149	0.5042
7	11/30/94	45.25	42.9142	5.4560	44.6872	0.3168	44.4012	0.7205
8	12/1/94	45.063	42.4869	6.6363	44.7328	0.1091	44.4651	0.3575
9	12/2/94	44.438	41.4725	8.7942	43.9591	0.2294	43.6744	0.5831
10	12/5/94	45.75	42.2443	12.2899	45.1192	0.3980	44.8261	0.8536
Ave MSE				3.9869		0.2155		0.4936

the first three forecast horizons, but for the remaining periods it consistently underestimated the actual values by a substantial amount. The absolute percent deviation from actual associated with the transfer function multi-step-ahead forecasts for L = 5 and 10 were respectively 3.194% and 7.663%. Comparable values for the 2SOLS($\tilde{p} = 8$) forecasts were 1.658% and 1.379%, respectively.

It should be noted that although the 2SOLS forecasts appear to be quite good, as DeLurgio observed, "whether one can make money in the stock market using th[ese] relationship[s] is speculative."

5. CONCLUSION

In this article, we proposed a new procedure for generating forecasts for regression models with serial correlation based on ordinary least squares and on the fact that the autocorrelation can be adequately represented by a finite AR process. We showed that predictors based on our approach are efficient for all L, provided T and $\tilde{p}$ are sufficient large. Moreover, from a large simulation study we found that for finite samples the predictive efficiency of our two-step linear approach using an AR($\tilde{p}$) approximation with $\tilde{p} = [\sqrt{T}/2]$ for the serial correlation is higher than that of OLS for short and medium horizons. In addition, we showed that the predictive efficiency of the 2SOLS method is very comparable to that of GLS based on AR($\tilde{p}$) corrections with $\tilde{p} = [\sqrt{T}/2]$ and EGLS, i.e. with known but estimated Ω. This suggests that for predictive purposes there is not much to be gained in trying to identifying the correct order and form of the serial correlation or in using more efficient estimation methods such as generalized least squares or maximum likelihood procedures which often require inversion of large matrices. We have shown that it is more important to tackle the serial correlation than to obtain the most efficient parameter estimates. For longer horizons, OLS yields forecasts that are as efficient as those generated by 2SOLS and GLS approaches.

APPENDIX

Proof of Lemma 3. Using the spectral representation of a_t, we have

$$a_{T+L} = \int_{-\pi}^{\pi} e^{i(T+L)\lambda} dZ(\lambda).$$

Consider the predictor

$$\tilde{a}_{T+L} = \sum_{j=0}^{\infty} c_j a_{T-j},$$

and its spectral representation

$$\tilde{a}_{T+L} = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}T\lambda} C(\mathrm{e}^{-\mathrm{i}\lambda}) \mathrm{d}Z(\lambda),$$

where $C(z) = \sum_{j=0}^{\infty} c_j z^j$.

Using the fact that the $\{\mathrm{d}Z(\lambda)\}$ is orthogonal to $G(z) \equiv \Pi(z)\Theta^{-1}(z)$, it can be shown that

$$\sigma_L^2 \equiv E(\tilde{a}_{T+L} - a_{T+L})^2 = (2\pi)^{-1}\sigma_v^2 \int_{-\pi}^{\pi} |\mathrm{e}^{\mathrm{i}L\lambda} G(\mathrm{e}^{-\mathrm{i}\lambda}) - C(\mathrm{e}^{-\mathrm{i}\lambda}) G(\mathrm{e}^{-\mathrm{i}\lambda})|^2 \mathrm{d}\lambda. \tag{A1}$$

Note that $G(\mathrm{e}^{-\mathrm{i}\lambda})$ and $C(\mathrm{e}^{-\mathrm{i}\lambda})G(\mathrm{e}^{-\mathrm{i}\lambda})$ are backward transforms. Therefore, σ_L^2 is minimized by choosing $C(\mathrm{e}^{-\mathrm{i}\lambda})$ so that $C(\mathrm{e}^{-\mathrm{i}\lambda})G(\mathrm{e}^{-\mathrm{i}\lambda})$ is the backward part of $\mathrm{e}^{\mathrm{i}L\lambda}G(\mathrm{e}^{-\mathrm{i}\lambda})$. Using the same approach as in Priestley (1981), we decompose $\mathrm{e}^{\mathrm{i}L\lambda}G(\mathrm{e}^{-\mathrm{i}\lambda})$ as the sum of the backward and forward transforms

$$\mathrm{e}^{\mathrm{i}L\lambda} G(\mathrm{e}^{-\mathrm{i}\lambda}) = G_+(\mathrm{e}^{-\mathrm{i}\lambda}) + G_-(\mathrm{e}^{-\mathrm{i}\lambda}),$$

where $G_+(z) = [z^{-L}G(z)]_+$ and $G_-(z) = [z^{-L}G(z)]_-$. Substituting the above decomposition into (A1) and using the orthogonality properties of spectral functions, we have

$$\sigma_L^2 = (2\pi)^{-1}\sigma_v^2 \left\{ \int_{-\pi}^{\pi} |G_-(\mathrm{e}^{-\mathrm{i}\lambda})|^2 \mathrm{d}(\lambda) + \int_{-\pi}^{\pi} |G_+(\mathrm{e}^{-\mathrm{i}\lambda}) - C(\mathrm{e}^{-\mathrm{i}\lambda}) G(\mathrm{e}^{-\mathrm{i}\lambda})|^2 \mathrm{d}(\lambda) \right\}, \tag{A2}$$

which is minimized by choosing

$$C(\mathrm{e}^{-\mathrm{i}\lambda}) = \frac{G_+(\mathrm{e}^{-\mathrm{i}\lambda})}{G(\mathrm{e}^{-\mathrm{i}\lambda})}.$$

The second term in (A2) vanishes with the above choice of $C(\mathrm{e}^{-\mathrm{i}\lambda})$. Thus, the minimum L-step mean square prediction error is

$$\sigma_L^2 = (2\pi)^{-1}\sigma_v^2 \int_{-\pi}^{\pi} |G_-(\mathrm{e}^{-\mathrm{i}\lambda})|^2 \mathrm{d}(\lambda).$$

Noting that $\tilde{a}_{T+L}$ can also be written as an MA(∞) process

$$\tilde{a}_{T+L} = \sum_{j=0}^{\infty} \vartheta_{j+L} v_{t-j},$$

or equivalently,

$$\tilde{a}_{T+L} = [z^{-L}G(z)]_+ / G(z) a_t,$$

yields

$$E(\tilde{a}_{T+L} - a_{T+L})^2 = \sigma_v^2 \sum_{j=0}^{L-1} \vartheta_j^2.$$

Since one can always find an AR process of finite order, say, AR($\tilde{p}$) such that

$$|C_{\tilde{p}}(\mathrm{e}^{-\mathrm{i}}\lambda) - G_+(\mathrm{e}^{-\mathrm{i}}\lambda)/G(\mathrm{e}^{-\mathrm{i}}\lambda)| < \epsilon$$

for all $\lambda \in [-\pi,\pi]$ [see, for example, Fuller (1996)], it follows that for any given $\epsilon > 0$, one can also find $T_0(\epsilon)$ and $\tilde{p}_0(\epsilon)$ such that for $T > T_0(\epsilon)$ and $\tilde{p} > \tilde{p}_0(\epsilon)$,

$$|E(\tilde{a}_{T+L} - a_{T+L})^2 - \sigma_L^2| < \epsilon.$$

Finally, the above result holds if one replaces $\tilde{a}_{T+L}$ by $\hat{a}_{T+L}$ since the OLS residuals obtained from (1) converges in probability to a_t. □

NOTE

Corresponding author: Sergio G. Koreisha, University of Oregon, Eugene, OR 97403, USA. E-mail: sergiok@lcbmail.uoregon.edu

REFERENCES

AMEMIYA, T. (1985) Advanced Econometrics. Cambridge: Harvard University Press.

BAILLE, R. (1979) The asymptotic mean square error of multistep prediction from the regression model with autoregressive errors. Journal of the American Statistical Association 74, 175–84.

BHANSALI, R. J. (1978) Linear prediction by autoregressive model fitting in the time domain. Annals of Statistics 6, 224–31.

BERK, K. (1974) Consistent autoregressive spectral estimates. Annals of Statistics 2, 489–502.

BOX, G. and JENKINS, G. (1976) Time Series Analysis: Forecasting and Control (2nd edn). San Francisco: Holden-Day.

BREUSCH, T. (1980) Useful invariance results for generalized regression models. Journal of Econometrics 13, 327–40.

BURMAN, P. (1989) A comparative study of ordinary cross-validation, v-hold cross-validation, and the repeated learning-testing methods. Biometrika 76, 503–14.

DELURGIO, S. (1998) Forecasting Principles and Applications. Boston: Irwin/McGraw-Hill.

ENGLE, R. (1974) Specification of the disturbance for efficient estimation. Econometrica 42, 135–46.

FANG, Y. and KOREISHA, S. (2004) Forecasting with serially correlated regression models. Journal of Statistical Computation and Simulation 74, 625–49.

FULLER, W. A. (1996) Introduction to Statistical Time Series. New York: John Wiley & Sons Inc.

HANNAN, E. J. (1970) Multiple Time Series. New York: John Wiley & Sons Inc.

HANNAN, E. J., DUNSMUIR, A. and DEISTLER, M. (1980) Estimation of vector ARMAX models. Journal of Multivariate Analysis 10, 275–95.

JUDGE, G. J., GRIFFITHS, W. E., HILL, R. C., LÜTKEPOHL and LEE, T. (1985) The Theory and Practice of Econometrics. New York: John Wiley & Sons Inc.

KAPETANIOS, G. (2003) A note on an iterative least-squares estimation method for ARMA and VARMA models. Economics Letters 79, 305–12.

KOREISHA, S. and FANG, Y. (2001) Generalized least squares with misspecified serial correlation structures. Journal of the Royal Statistical Society, Series B 63, 515–31.

KOREISHA, S. and PUKKILA, S. (1987) Identification of nonzero elements in the polynomial matrices of VARMA processes. Journal of the Royal Statistical Society, Series B 49, 112–26.

KOREISHA, S. and PUKKILA, S. (1990). A generalized least squares approach for estimation of autoregressive moving average models. Journal of Time Series Analysis 11, 139–51.

PINDYCK, R. and RUBINFELD, D. (1998) Econometric Models and Economic Forecasts (4th edn). Boston: Irwin/McGraw-Hill.

PRIESTLEY, M. B. (1981) Spectral Analysis and Time series. Volume 2: Multivariate Series, Prediction and Control. New York: Academic Press.

POSKITT, D. and SALAU, M. (1995). On the relationship between generalized least squares and Gaussian estimation of vector ARMA models. Journal of Time Series Analysis 16, 617–45.

PUKKILA, T., KOREISHA, S. and KALLINEN, A. (1990) The identification of ARMA models. Biometrika 77, 537–49.

RACINE, J. (1997) Feasible cross-validatory model selection for general stationary processes. Journal of Applied Econometrics 12, 169–79.

SAIKKONEN, P. (1986) Asymptotic properties of some preliminary estimators for autoregressive moving average time series models. Journal of Time Series Analysis 7, 133–55.

Scientific Computing Associates (2000) The SCA Statistical System. Villa Park, IL: Scientific Computing Associates Corporation.

Shao, J. (1993) Linear model selection by cross-validation. Journal of American Statistical Association 88, 486–94.

Thursby, J. (1987) OLS or GLS in the presence of specification error? Journal of Econometrics 35, 359–74.

Wahlberg, B. (1989) Estimation of autoregressive moving-average models via high-order autoregressive approximations. Journal of Time series Analysis 10, 283–99.

Yamamoto, T. (1976) Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. Applied Statistics 25, 123–27.

Zinde-Walsh, V. and Galbraith, J. (1991) Estimation of a linear regression model with stationary ARMA(p,q) errors. Journal of Econometrics 47, 333–57.

Zyskind, G. (1967) On Canonical forms, non-negative covariance matrices and best and simple linear least squares estimators in linear models. Annuals of Mathematical Statistics 38, 1092–109.