Journal of Statistical Computation and Simulation
Vol. 74, No. 9, September 2004, pp. 625–649

FORECASTING WITH SERIALLY CORRELATED REGRESSION MODELS

YUE FANG[a] and SERGIO G. KOREISHA[a,b,*]

[a]*University of Oregon, Eugene, OR 97403, USA and* [b]*University of Queensland, Brisbane, Qld 4072, Australia*

(Revised 5 November 2002; In final form 26 August 2003)

In this article we investigate the asymptotic and finite-sample properties of predictors of regression models with autocorrelated errors. We prove new theorems associated with the predictive efficiency of generalized least squares (GLS) and incorrectly structured GLS predictors. We also establish the form associated with their predictive mean squared errors as well as the magnitude of these errors relative to each other and to those generated from the ordinary least squares (OLS) predictor. A large simulation study is used to evaluate the finite-sample performance of forecasts generated from models using different corrections for the serial correlation.

Keywords: Asymptotic mean squared errors; Autoregressive disturbances; Generalized least squares; Incorrect generalized least squares; Predictive mean squared efficiency; Simulation

1 INTRODUCTION

Since Cochrane and Orcutt (1949) and Durbin and Watson (1950) developed an approximate transformation to deal with and test for autoregressive disturbances of order 1, we have witnessed a plethora of studies dealing with the problems of serial correlation in regression models. Chaudhury *et al.* (1999), among others, have documented the evolution of approximate/exact transformations to deal with more complex serial correlation structures, as well as the development of new, more powerful estimation methods that have occurred in the last 50 years. When disturbances exhibit serial correlation, least squares will yield unbiased, but inefficient estimators of parameters of the model, thus invalidating all tests of significance. In addition, these estimated regression coefficients will have larger sampling variances than other estimators such as generalized least squares (GLS) that deal explicitly with the autocorrelation in the residuals. Furthermore, forecasts generated from such models can be seriously inefficient, not only because of the inefficiency of the parameter estimators, but also because the error between the fitted and actual value in the last observation is apt to persist into the actual forecast interval.

* Corresponding author. E-mails: s.koreisha@business.uq.edu.au/sergiok@lcbmail.uoregon.edu.

ISSN 0094-9655 print; ISSN 1563-5163 online
DOI: 10.1080/00949650310001620112

With a few exceptions, most of the studies that have been conducted on this topic assume that the error covariance matrix Ω from the regression model,

$$Y = X\beta + \varepsilon, \tag{1}$$

where Y is a $(T \times 1)$ vector of observations on a dependent variable, X is a $(T \times k)$ design matrix, and ε is a random vector with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2\Omega = \{\gamma(i-j)\}_{i,j=1}^{T}$, is either known or could be estimated consistently from data. Until recently, of the very few studies that considered the properties of estimators, or of the forecasts when the structure of Ω was incorrectly identified, or when its parameters were inefficiently estimated, most dealt with or depended on asymptotic results (Goldberger, 1962; Amemiya, 1973; Engle, 1974). Koreisha and Fang (2001), on the other hand, compared the finite-sample efficiencies of ordinary least squares (OLS) and GLS *vis-à-vis* incorrect GLS (IGLS) estimators, *i.e.* estimators based on incorrectly identified Ω, and established theoretical efficiency bounds for IGLS relative to GLS and OLS. They found that GLS estimation based on autoregressive representations of autoregressive-moving average (ARMA) disturbances yielded more efficient estimates than OLS particularly when the order of the autoregression was set near $[\sqrt{T}/2]$, and that the differences in estimation efficiency between estimated IGLS and GLS were small.

In this article, we will augment the work of Koreisha and Fang (2001) by investigating the impact that estimated IGLS corrections may have on the forecasting performance of regression models with serial correlation. This work differs from others dealing with regression forecasts with autocorrelated disturbances such as Armstrong (1978) and Dielman (1985) in that it does not assume that the form of the autocorrelation is known (AR(1) in those cases) *a priori*. The main issue here is not determining which estimation procedure yields the best forecasts when Ω is known. Our goal is to show, both theoretically and with simulated data, that there exists a certain class of estimators based on incorrectly identified residual autocorrelation structures that for finite samples can yield as good, and more often, better forecasts than those generated from OLS or from GLS using the correct form of the residual autocorrelation structure.

The article is organized as follows: in Section 2, we describe the predictors used in this study to generate the forecasts and prove new theorems that extend existing results on prediction mean squared errors (PMSEs). In Section 3, we present results from an exhaustive Monte-Carlo study to contrast the finite-sample forecasting performance of models using GLS (correct and incorrect) with those using OLS estimators. In Section 4, we offer some concluding remarks.

2 FORECASTING EFFICIENCY

Estimates for forecasts generated at time T for h future observations, $Y^* = (y_{T+h}, y_{T+h-1}, \ldots, y_{T+1})'$, given by

$$Y^* = X^*\beta + \varepsilon^*, \tag{2}$$

where X^* is the design matrix of future exogenous variables with dimension $(h \times k)$ and ε^* is an $(h \times 1)$ vector of future correlated disturbances, can be derived, provided some assumptions are made about β, future values of the exogenous variables, and the variance–covariance matrix of the past and future disturbances, which can be depicted by the following partitioning scheme:

$$E\left[\begin{pmatrix} \varepsilon \\ \varepsilon^* \end{pmatrix}(\varepsilon'\varepsilon^{*\prime})\right] = \sigma^2\begin{bmatrix} \Omega & V \\ V' & \Omega^* \end{bmatrix} \equiv \sigma^2\Gamma. \tag{3}$$

2.1 Predictors with $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, and X^* Known

If β, Γ, and future values of the exogenous variables are known, then like in Goldberger (1962) the minimum mean squared predictor for Y^*, $Y^*_{(I)}$, can be written as the sum of a nonstochastic component based on future values of exogenous variables and of a stochastic component based on future disturbances, namely,

$$Y^*_{(I)} = X^*\beta + V'\Omega^{-1}(Y - X\beta), \tag{4}$$

and its PMSE is given by

$$\text{PMSE}(Y^*_{(I)}) = E[(Y^* - Y^*_{(I)})(Y^* - Y^*_{(I)})'] = \Omega^* - V'\Omega^{-1}V, \tag{5}$$

assuming that $\sigma^2 = 1$ (see also Harville, 1997, p. 456).

Ignoring altogether the correlation between past and future disturbances yields another set of predictors,

$$Y^*_{(II)} = X^*\beta, \tag{6}$$

with $\text{PMSE}(Y^*_{(II)}) = \Omega^*$, which is strictly greater than $\text{PMSE}(Y^*_{(I)})$ if $V \neq 0$, since Ω is positive definite.

It is also possible to derive similar expressions for predictors based on finite autoregressive representations (of order $\tilde{p}$) of the disturbance process. Since only the last $\tilde{p}$ observations will be involved in the construction of the forecasts, it will be easier to derive expressions for this set of predictors, $Y^*_{(III)}$, by focusing only on this subset of observations and re-expressing Eq. (1) as,

$$\mathcal{Y}_t = \mathcal{X}_t\beta + \mathcal{E}_t \tag{7}$$

and

$$\mathcal{E}_t = \Lambda\mathcal{E}_{t-1} + \mathcal{U}_t, \tag{8}$$

where $\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_{t-\tilde{p}+1})'$, $\mathcal{E}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\tilde{p}+1})'$ and $\mathcal{U}_t = (u_t, 0, \ldots, 0)'$ are all $\tilde{p}$-dimensional vectors and $\mathcal{X}_t$ and Λ are $\tilde{p} \times k$ and $\tilde{p} \times \tilde{p}$ matrices, respectively, defined as:

$$\mathcal{X}_t = \begin{pmatrix} x_{1,t} & x_{2,t} & \cdots & x_{k,t} \\ x_{1,t-1} & x_{2,t-1} & \cdots & x_{k,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,t-\tilde{p}+1} & x_{2,t-\tilde{p}+1} & \cdots & x_{k,t-\tilde{p}+1} \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_{\tilde{p}-1} & \pi_{\tilde{p}} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

As can be seen, only the first equation in this system has been modified: ε_t has been rewritten in terms of its AR($\tilde{p}$) representation, namely, $\varepsilon_t = \sum_{i=1}^{\tilde{p}} \pi_i \varepsilon_{t-i} + u_t$. The residual u_t from this equation will not necessarily be white noise, unless ε_t follows an AR(l) ($l \leq \tilde{p}$) process.

Successive substitution of h past values of ε_t in Eq. (8) yields

$$\mathcal{E}_t = \sum_{j=0}^{h-1} \Lambda^j \mathcal{U}_{t-j} + \Lambda^h \mathcal{E}_{t-h}. \tag{9}$$

Thus, substituting Eq. (9) into Eq. (7) for $t = T + h$ gives rise to the set of equations,

$$\begin{aligned}\mathcal{Y}_{T+h} &= \mathcal{X}_{T+h}\beta + \mathcal{E}_{T+h} = \mathcal{X}_{T+h}\beta + \sum_{j=0}^{h-1} \Lambda^j \mathcal{U}_{T+h-j} + \Lambda^h \mathcal{E}_T \\ &= \sum_{j=0}^{h-1} \Lambda^j \mathcal{U}_{T+h-j} + \mathcal{X}_{T+h}\beta + \Lambda^h(\mathcal{Y}_T - \mathcal{X}_T\beta),\end{aligned}$$

which characterize the values of $Y^*_{(III)}$, namely

$$Y^*_{(III)} = X^*\beta + \tau'(\mathcal{Y}_T - \mathcal{X}_T\beta), \tag{10}$$

where τ' contains the first h rows of Λ^h. Note that the exclusion of the first term in Eq. (10) from $Y^*_{(III)}$ is what primarily differentiates this predictor from $Y^*_{(I)}$. Its influence is presumed to be relatively small because for sufficiently large $\tilde{p}$, the AR($\tilde{p}$) approximation should adequately depict stationary processes generating the serial correlation (see, for example, Berk, 1974; Bhansali, 1978).

If $\varrho' \equiv (\tau', \mathcal{O}')$, where $\mathcal{O}'$ is the null matrix of dimension $h \times (T - \tilde{p})$, then, $Y^*_{(III)}$ can be rewritten as

$$Y^*_{(III)} = X^*\beta + \varrho'(Y - X\beta), \tag{11}$$

since the second term depends on all the available observations up to time T (*i.e.* X and Y).

From Eqs. (1)–(3) it can be shown that

$$\text{PMSE}(Y^*_{(III)}) = \Omega^* + \varrho'\Omega\varrho - V'\varrho - \varrho'V. \tag{12}$$

We summarize these results in the form of a theorem.

THEOREM 1 *If* β, Γ *and future values of the exogenous variables are known, then the predictor based on* AR($\tilde{p}$) *correction of the serial correlation is given by*

$$Y^*_{(III)} = X^*\beta + \varrho'(Y - X\beta)$$

and

$$\text{PMSE}(Y^*_{(III)}) = \Omega^* + \varrho'\Omega\varrho - V'\varrho - \varrho'V.$$

Note that $Y^*_{(III)}$ will be less efficient than $Y^*_{(I)}$, unless ε_t follows an AR(l) process with $l \le \tilde{p}$. When ε_t follows an AR process, $\varrho' = V'\Omega^{-1}$ (Baillie, 1979), and consequently, PMSE($Y^*_{(III)}$) = PMSE($Y^*_{(I)}$) by Eqs. (5) and (12). Hence, we establish the following corollary.

COROLLARY 1 *If β, Γ and future values of the exogenous variables are known, then*

$$\mathrm{PMSE}(y^*_{(III)}(T+h)) \ge \mathrm{PMSE}(y^*_{(I)}(T+h)),$$

where $y^*_{(III)}(T+h)$ and $y^*_{(I)}(T+h)$ are predictors of y_{T+h} at time T based on $Y^*_{(III)}$ and $Y^*_{(I)}$, respectively. The equality holds if ε_t follows an AR(l) *process, where $l \le \tilde{p}$.*

The relative magnitude of the predictive efficiencies between PMSE($Y^*_{(II)}$) and PMSE($Y^*_{(III)}$) depends on the correlation structure of the disturbance as well as $\tilde{p}$ and h. The following corollary gives the necessary and sufficient condition for PMSE($y^*_{(III)}(T+h)$) to be less than or equal to PMSE($y^*_{(II)}(T+h)$), where $y^*_{(II)}(T+h)$ is the predictor of y_{T+h} at time T based on $Y^*_{(II)}$.

COROLLARY 2 *If β, Γ and future values of the exogenous variables are known, then*

$$\mathrm{PMSE}(y^*_{(III)}(T+h)) \le \mathrm{PMSE}(y^*_{(II)}(T+h))$$

if and only if $\min_i(b_i/w_i) \ge 1/2$ ($w_i \ne 0$), *where b_i is the ith element of the vector $B = \Upsilon'^{-1}\tilde{V}$, w_i is the ith element of the vector $W = \Upsilon\eta$, and $\tilde{V}' = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+\tilde{p}-1))$; η' is the first row of Λ^h; and Υ is a nonsingular unitary matrix such that $\Upsilon'\Upsilon = \{\gamma(i-j)\}^{\tilde{p}}_{i,j=1}$.*

Proof See Appendix.

Noting that $\varrho'\Omega\varrho$ and $V'\varrho$ are functions of π_is and the autocorrelation function of ε, $\gamma(\cdot)$, we can re-express Corollary 2 in a more convenient form for computational purposes.

COROLLARY 3 *If β, Γ and future values of the exogenous variables are known, then*

$$\mathrm{PMSE}(y^*_{(III)}(T+h)) \le \mathrm{PMSE}(y^*_{(II)}(T+h))$$

if and only if $\Delta \equiv \sum_{i=1}^{\tilde{p}}\sum_{j=1}^{\tilde{p}} \gamma(i-j)\pi_i^{(h)}\pi_j^{(h)} - 2\sum_{i=1}^{\tilde{p}} \gamma(h+i-1)\pi_i^{(h)} \le 0$, where $\pi_l^{(h)}$ is the lth element of the first row of Λ^h.

Proof See Appendix.

2.2 Predictors with Estimated Parameters

Of greater practical value are predictors generated from estimated values of β and Λ. Such predictors and their properties have been investigated by Goldberger (1962), Yamamoto (1979), and Baillie (1979) among others, and can be developed further using some of the results associated with $Y^*_{(m)}$, $m = \{I, II, III\}$ discussed in Section 2.1.

Assuming first that Γ is known, it is possible to develop a general class of linear predictors based on different estimates for β. The theorem below specifies the form of these predictors and their corresponding PMSEs.

THEOREM 2 *If Γ and future values of the exogenous variables are known, the predictor*

$$\hat{Y}^* = X^*\hat{\beta} + \mathcal{C}'(Y - X\hat{\beta}), \tag{13}$$

where $\mathcal{C}'$ is an $(h \times T)$ matrix and $\hat{\beta} = (X'\Xi^{-1}X)^{-1}X'\Xi^{-1}Y$, has a prediction mean squared error given by

$$\text{PMSE}(\hat{Y}^*) = \Omega^* + D\Omega D' - DV - V'D', \tag{14}$$

where $D = X^*A + \mathcal{C}'(I - XA)$ with $A = (X'\Xi^{-1}X)^{-1}X'\Xi^{-1}$.

Proof See Appendix.

As can be easily seen, setting $\mathcal{C} = 0$ and $\Xi = I$ defines the OLS predictor $\hat{Y}^*_{\text{OLS}}$. Its prediction mean squared error is given by

$$\text{PMSE}(\hat{Y}^*_{\text{OLS}}) = \Omega^* + X^*(X'X)^{-1}X'X(X'X)^{-1} - X^*(X'X)^{-1}X'V - V'X(X'X)^{-1}X^{*\prime}. \tag{15}$$

Moreover, if $\Xi = \Omega$ and $\mathcal{C}' = V'\Omega^{-1}$, then $\hat{Y}^*$ becomes the GLS predictor $\hat{Y}^*_{\text{GLS}}$, which Goldberger (1962) proved was the best linear unbiased predictor (BLUP) for Y^*; and furthermore,

$$\begin{aligned}\text{PMSE}(\hat{Y}^*_{\text{GLS}}) = \Omega^* &+ X^*(X\Omega^{-1}X^*)^{-1}X^{*\prime} - V'(\Omega^{-1} - \Omega^{-1}X(X\Omega^{-1}X^*)^{-1}X'\Omega^{-1})V \\ &- X^*(X\Omega^{-1}X^*)^{-1}X'\Omega^{-1}V - V'\Omega^{-1}X(X\Omega^{-1}X^*)^{-1}X^{*\prime}.\end{aligned}$$

(see also Judge *et al.*, 1985, p. 316) We will refer to $\hat{Y}^*_{\text{AR}(\tilde{p})}$ as the predictor based on the finite autoregressive representation of the serial correlation structure if $\mathcal{C}' = \varrho'$ and $\Xi = \Sigma$, where Σ is the covariance matrix of the $\tilde{p}$th order autoregressive representation of the disturbance.

If Γ has to be estimated along with β, then estimates of the presample disturbances will be functions of estimated parameters. Consequently, results such as those in Theorem 2 will not reflect all the uncertainty in the predictions. Under these circumstances, the derivations for PMSEs become more involved, requiring asymptotic approximations.

Assuming that the serial correlation followed an AR(p) process, Baillie (1979) considered the case for which β and matrices in Eq. (3) were estimated simultaneously using the maximum likelihood procedures. Applying a result in Pierce (1971) dealing with asymptotic properties of maximum likelihood estimators, he was able to obtain the asymptotic PMSE (APMSE) for Goldberger's (1962) predictor based on estimated β as well as Γ. Yamamoto (1979) derived similar results for several other predictors assuming that the disturbances followed an AR(1) process.

Now consider the predictor

$$\hat{\hat{Y}}^* \equiv X^*\hat{\hat{\beta}} + \hat{\mathcal{C}}'(Y - X\hat{\hat{\beta}}), \tag{16}$$

where $\hat{\hat{\beta}} = (X'\hat{\Xi}^{-1}X)^{-1}X'\hat{\Xi}^{-1}Y$, $\hat{\Xi}$ and $\hat{\mathcal{C}}$ are Ξ and $\mathcal{C}$ with elements replaced by their estimates, respectively. Note that

$$\hat{\hat{Y}}^* - Y^* = X^*\hat{\hat{\beta}} + \hat{\mathcal{C}}'(Y - X\hat{\hat{\beta}}) - Y^*,$$

which can also be expressed as a sum of three components, namely,

$$\hat{\hat{Y}}^* - Y^* = X^*(\hat{\hat{\beta}} - \beta) + (\hat{\mathcal{C}}' - \mathcal{C}')(Y - X\hat{\hat{\beta}}) + (\mathcal{C}'\varepsilon - \varepsilon^*). \tag{17}$$

The first two are essentially estimation errors associated with β and $\mathcal{C}'$, respectively, while the third captures the uncertainty associated with future disturbances. Under some mild assumptions, the estimator $\hat{\hat{\beta}}$ can be shown to be consistent, so the autocovariance function of ε can be estimated using residuals

$$\hat{\varepsilon}_t = y_t - w_t\hat{\hat{\beta}}, \quad t = 1, 2, \ldots, T,$$

where w_t is the tth column of the matrix X. Thus,

THEOREM 3 *If X is nonstochastic and satisfies the assumptions*:

(a) *the regressors* (*columns of X*) *are linearly independent*,
(b) $\lim_{T\to\infty}(T^{-1}X'X)$ *finite and nonsingular*, *and*
(c) *if* $p\lim_{T\to\infty} T^{-1}X'\hat{\Xi}^{-1}\hat{\Omega}^{-1}\hat{\Xi}^{-1}X$ *is finite and nonsingular and* $p\lim_{T\to\infty} T^{-1}\times X'\hat{\Xi}^{-1}\hat{\Omega}^{-1}\Upsilon_0\varepsilon = 0$, *where* Υ_0 *is a nonsingular unitary matrix such that* $\Upsilon_0{}'\Upsilon_0 = \{\gamma(i-j)\}_{i,j=1}^{T}$, *then*
  (1) $\hat{\hat{\beta}} - \beta = O(T^{-1/2})$ *and*
  (2) $\hat{\gamma}_{\hat{\varepsilon}}(l) - \gamma(l) = O(T^{-1})$, *where* $\hat{\gamma}_{\hat{\varepsilon}}(l) = T^{-1}\sum_{t=1}^{T-l}\hat{\varepsilon}_t\hat{\varepsilon}_{t+l}$.

Proof See Appendix.

Similar results to Theorem 3 also hold even if the regressor matrix X is stochastic, but the underlying assumptions in such a case are more complicated and restrictive. See, for example, Schmidt (1976) for discussions on general issues regarding stochastic regressors.

If β is estimated consistently, then, as the sample size increases, the first component of Eq. (17) converges to zero. Similarly since $\mathcal{C}$ is a function of Γ, if the parameters of the disturbance process can be estimated consistently from data, then the second term also converges to zero as the sample size increases. This, however, will not be the case for the third term. Its variance, depending on $\mathcal{C}$, can be shown to be identical to PMSE($Y^*_{(m)}$), $m = I, II$, or III. Thus, based on the fact that the PMSE($\hat{\hat{Y}}^*$) is primarily dominated by $(\mathcal{C}'\varepsilon - \varepsilon^*)$, it is possible to derive an asymptotic expression for the PMSE of $\hat{Y}^*$, which we state below as a new theorem.

THEOREM 4 *If future values of the exogenous variables are known and both β and Γ* (*and hence also $\mathcal{C}$*) *are estimated consistently*, *then as $T \to \infty$*,

$$\text{APMSE}(\hat{\hat{Y}}^*) = \Omega^* + \mathcal{C}'\Omega\mathcal{C} - V'\mathcal{C} - \mathcal{C}'V.$$

Proof See Appendix.

In addition, let $\hat{\hat{Y}}^*_{\text{EGLS}}$ and $\hat{\hat{Y}}^*_{\text{EAR}(\tilde{p})}$ be, respectively, the GLS predictor using the correct serial correlation structure and the IGLS predictor with AR($\tilde{p}$) correction generated from Eq. (16) using estimated β and Γ. Then we have, COROLLARY 4 *If future values of the exogenous variables are known and both β and Γ (and hence also $\mathcal{C}$) are estimated consistently, then*

- $\mathrm{APMSE}(\hat{\hat{Y}}^*_{\mathrm{EGLS}}) = \mathrm{PMSE}(Y^*_{(I)}) = \Omega^* - V'\Omega^{-1}V$,
- $\mathrm{APMSE}(\hat{\hat{Y}}^*_{\mathrm{OLS}}) = \mathrm{PMSE}(Y^*_{(II)}) = \Omega^*$, *and*
- $\mathrm{APMSE}(\hat{\hat{Y}}_{\mathrm{EAR}(\tilde{p})}) = \mathrm{PMSE}(Y^*_{(III)}) = \Omega^* + \varrho'\Omega\varrho - V'\varrho - \varrho'V$.

Furthermore,

$$\mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{EAR}(\tilde{p})}(T+h)) \geq \mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{EGLS}}(T+h)),$$

and if $\Delta \equiv \sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{p}} \gamma(i-j)\pi_i^{(h)}\pi_j^{(h)} - 2\sum_{i=1}^{\tilde{p}} \gamma(h+i-1)\pi_i^{(h)} \leq 0$ ($\pi_l^{(h)}$ *is defined in Corollary* 3), *then*

$$\mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{EAR}(\tilde{p})}(T+h)) \leq \mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{OLS}}(T+h)),$$

where $\hat{\hat{y}}^*_{\mathrm{EAR}(\tilde{p})}(T+h)$, $\hat{\hat{y}}^*_{\mathrm{EGLS}}(T+h)$ *and* $\hat{\hat{y}}^*_{\mathrm{OLS}}(T+h)$ *are predictors of* y_{T+h} *at time T based on* $\hat{\hat{Y}}^*_{\mathrm{EAR}(\tilde{p})}$, $\hat{\hat{Y}}^*_{\mathrm{EGLS}}$ *and* $\hat{\hat{Y}}^*_{\mathrm{OLS}}$, *respectively.*

We present below two examples operationalizing the condition on Δ for which $\mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{EAR}(\tilde{p})}(T+h))$ will be less than $\mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{OLS}}(T+h))$. Finite-sample results for many other structures will be presented in Section 3.

Example 1 Assume that ε_t follows an ARMA(1, 1) process, $\varepsilon_t - \phi_1\varepsilon_{t-1} = a_t - \theta_1 a_{t-1}$ with $\sigma_a^2 = 1$. Applying Corollary 4,

$$\mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{EAR}(\tilde{p})}(T+h)) \leq \mathrm{APMSE}(\hat{\hat{y}}^*_{\mathrm{OLS}}(T+h)) \quad \text{if and only if } \Delta \leq 0.$$

Since $\pi_1 = \gamma(1))/(\gamma(0)$ and from Corollary 3, if $\tilde{p} = 1$ and $h = 1$, then

$$\Delta_{\tilde{p}=1,h=1} = \gamma(0)\pi_1^2 - 2\gamma(1)\pi_1 = -\frac{\gamma^2(1)}{\gamma(0)} \leq 0,$$

where $\gamma(0) = (1+\theta_1^2 - 2\phi_1\theta_1)/(1-\phi_1^2)$ and $\gamma(1) = (\phi_1 - \theta_1)(1-\phi_1\theta_1)/(1-\phi_1^2)$. Note that $\Delta_{\tilde{p}=1,h=1}$ is always less than zero unless $\phi_1 = \theta_1$ which is effectively the parameterization of white noise error process. We also note that the magnitude of $\Delta_{\tilde{p}=1,h=1}$ depends on the values of ϕ_1 and θ_1, as shown in Figure 1(a).

If $\tilde{p} = 1$ but $h = 2$, then

$$\Delta_{\tilde{p}=1,h=2} = \gamma(0)(\pi_1^{(2)})^2 - 2\gamma(2)\pi_1^{(2)} = -\frac{\gamma^2(1)}{\gamma^3(0)}(\phi_1^2\gamma^2(0) - \theta_1^2),$$

which is also nonpositive unless $|\theta_1/\phi_1| > \gamma(0)$. When $|\theta_1/\phi_1| > \gamma(0)$, ε_t can be viewed approximately as an MA(1) process. In such a case, two-step-ahead forecasts based on the AR(1) correction are not efficient (see Fig. 1(b)).

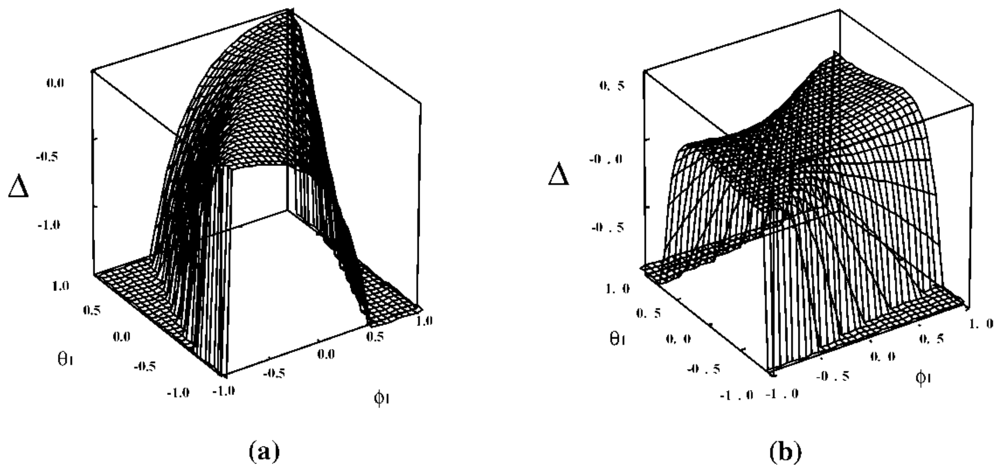

FIGURE 1 Δ for ARMA(1, 1) error processes: (a) $\tilde{p} = 1$ and $h = 1$; (b) $\tilde{p} = 1$ and $h = 2$.

Example 2 Suppose that ε_t follows an MA(2) process, $\varepsilon_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ with $\sigma_a^2 = 1$. Again, if $h = 1$ and $\tilde{p} = 1$, using results in Corollary 3, we have

$$\Delta_{\tilde{p}=1,h=1} = \gamma(0)\pi_1^2 - 2\gamma(1)\pi_1 = -\frac{\gamma^2(1)}{\gamma(0)} \leq 0,$$

where $\gamma(0) = (1 + \theta_1^2 + \theta_2^2)$ and $\gamma(1) = -\theta_1(1 - \theta_2)$. For MA(2) processes, $\Delta_{\tilde{p}=1,h=1}$ is strictly less than zero unless $\theta_1 = 0$. As we increase $\tilde{p}$ from 1 to 2,

$$\begin{aligned}\Delta_{\tilde{p}=2,h=1} &= \gamma(0)[\pi_1^2 + \pi_2^2] + 2\gamma(1)\pi_1\pi_2 - 2[\gamma(1)\pi_1 + \gamma(2)\pi_2] \\ &= \gamma^{-2}(0)[\gamma^2(1)(2\gamma(2) - \gamma(0)) - \gamma^2(2)\gamma(0)].\end{aligned}$$

Since $2\gamma(2) - \gamma(0) = -(1 + \theta_2)^2 - \theta_1^2 \leq 0$, $\Delta_{\tilde{p}=2,h=1}$ is less than or equal to zero. The absolute values of both $\Delta_{\tilde{p}=2,h=1}$ and $\Delta_{\tilde{p}=1,h=1}$ depend on the values of θ_1 and θ_2 (see Fig. 2).

Long-term forecasts play a vital role in the development of plans and formulation of strategies in all organizations. Selection of the appropriate set of numbers to be used is often based on performance measures applied to figures generated by different forecasting methods. The following theorem, however, shows that there is not much to be gained in terms of long-term predictive efficiency by modeling the correlation structure of the disturbances.

THEOREM 5 *If future values of the exogenous variables are known and both β and Γ are estimated consistently, then*

$$\lim_{h\to\infty} \text{APMSE}(\hat{\hat{y}}^*_{\text{EGLS}}(T+h)) = \lim_{h\to\infty} \text{APMSE}(\hat{\hat{y}}^*_{\text{OLS}}(T+h)) = \lim_{h\to\infty} \text{APMSE}(\hat{\hat{y}}^*_{\text{EAR}(\tilde{p})}(T+h)),$$

where $\hat{\hat{y}}^*_{\text{EGLS}}(T+h)$, $\hat{\hat{y}}^*_{\text{OLS}}(T+h)$, and $\hat{\hat{y}}^*_{\text{EAR}(\tilde{p})}(T+h)$ are predictors of y_{T+h} at time T based on $\hat{\hat{Y}}^*_{\text{EGLS}}$, $\hat{\hat{Y}}^*_{\text{OLS}}$, and *$\hat{\hat{Y}}^*_{\text{EAR}(\tilde{p})}$, respectively.*

Proof See Appendix.

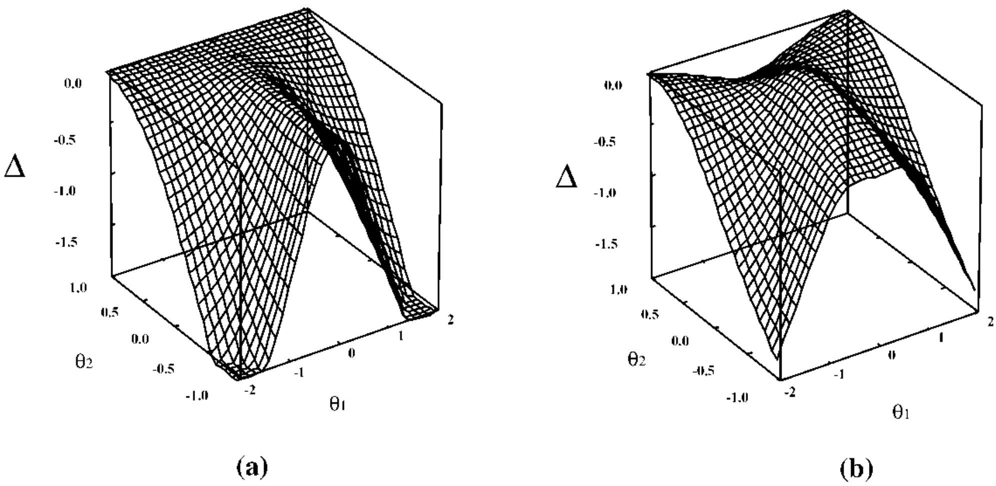

FIGURE 2 Δ for MA(2) error processes (a) $\tilde{p} = 1$ and $h = 1$; (b) $\tilde{p} = 2$ and $h = 1$.

3 SIMULATION RESULTS

In this section, we will contrast the forecast performance of IGLS methods applied on serially correlated regression models relative to those obtained from OLS and GLS procedures. All comparisons will be made on the basis of *estimated* parameter values. Here we will assume that the serially correlated disturbances ε in Eq. (1) follow an ARMA process:

$$\Phi(B)\varepsilon_t = \Theta(B)a_t, \tag{18}$$

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 $\{a_t\}$ is a Gaussian white noise process with variance σ_a^2. We say that process (18) is both stationary and invertible if the roots of the characteristic equations $\Phi(B) = 0$ and $\Theta(B) = 0$ are outside the unit circle. Stationary and invertible ARMA(p, q) models can also be expressed as either infinite autoregressions, $\Pi(B)\varepsilon_t = a_t$ or infinite moving averages, $\varepsilon_t = \psi(B)a_t$. In practice, however, such representations may be approximated by processes of relatively low order because the coefficients in $\Pi(B)$ and $\Psi(B)$ may be effectively zero beyond some finite lag.

We will focus solely on AR($\tilde{p}$) GLS corrections for disturbances generated by mixed ARMA(p, q) processes since they are probably the most widely used ones by practitioners, at least based on textbook coverage and implementation in statistical packages.

First, we generated, for sample sizes of 50, 100, and 200 observations, 1000 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. (To minimize initial value effects, we generated $(T + 100)$ observations, and only the latter T observations were used in simulations.) The parameter values for the residual ARMA structures were chosen to not only conform with other previously published studies such as Engle (1974), Pukkila *et al.* (1990), Zinde-Walsh and Galbraith (1991), and Koreisha and Fang (2001), but also to provide a representative set of examples of possible autocorrelated error structures in regression models. All residual structures were simulated using the SAS/ETS ARIMA Procedure.

Then we created the regression model,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \tag{19}$$

where the generating process for the exogenous variable x_t followed an AR(1) process, $(1 - \varphi B)x_t = v_t$, with $v_t \sim IN(0, 1)$, and $E(\varepsilon_t, v_s) = 0$, $\forall t \neq s$, and $\varphi = \{0.0, 0.5, 1.0\}$. For each sample size, only one set of random numbers was generated for each of the AR(1) model structures of the exogenous variable. Breusch (1980) has shown that for a fixed regressor the distribution of $(\hat{\hat{\beta}}_{\text{EGLS}} - \beta)/\sigma$ does not depend on β and σ^2. In addition, the result also holds if the covariance matrix is misspecified. This implies that in simulation studies, only one point in the parameter space (β, σ^2) needs to be considered for estimated IGLS (EIGLS). For illustrative purposes, we take $\beta_0 = 2.0$, $\beta_1 = 0.5$, and $\sigma^2 = 1$ in our simulations.

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

$$\hat{\xi}_{i/j}(T+h) \equiv \frac{E(\hat{\hat{y}}^*_{(i)}(T+h) - y_{T+h})^2}{E(\hat{\hat{y}}^*_{(j)}(T+h) - y_{T+h})^2}, \quad i, j = \{\text{OLS, EGLS, EIGLS}\} \text{ and } i \neq j, \tag{20}$$

where $\hat{\hat{y}}^*_{(m)}(T+h)$ represents the forecasted value based on method m at the time $T+h$, y_{T+h} is the actual generated value at $T+h$, was calculated for four forecast horizons, $h = \{1, 2, 5, 10\}$. A ratio less than 1 indicates that forecasts obtained from method i in Eq. (20) are more efficient than those generated from method j.

Tables I–V contrast selected predictive relative mean squared errors efficiencies among OLS and 4 GLS procedures: the GLS based on the correct residual model structures but with estimated ARMA coefficients (denoted as EGLS and to be considered as the performance benchmark); the GLS based on AR(1) correction with an estimated AR coefficient (denoted as EAR(1)); and two other EAR($\tilde{p}$) estimates with lags $\tilde{p}$ equal to the closest integer part of $\sqrt{T}/2$ and $\sqrt{T}$, respectively. Estimates for the AR parameters used in EAR($\tilde{p}$) correction were obtained using unconditional least squares, also referred to as nonlinear least squares (Spitzer, 1979). To provide an idea of the magnitude of the actual PMSE, we also included in these tables the actual estimates for PMSE($\hat{\hat{y}}^*_{\text{EAR}(\tilde{p})}(T+h)$). For the sake of brevity and to avoid a great deal of repetitiveness, these tables do not include all permutations of sample sizes and autoregressive corrective orders associated with each simulated parameterization of the serial correlation structure.

Examining the results of Tables I–V, we see that regardless of sample size for practically all model structures and parameterizations, the predictive efficiency of estimated GLS, including those based on incorrectly identified error structures (AR($\tilde{p}$)), is higher than OLS for short and medium term horizons ($h \leq 5$). The degree of improvement in relative predictive efficiencies, however, depends on the structure of the serial correlation. For pure MA(q) error processes (Tab. I and IV), as expected, noticeable improvement in the forecasts occurs primarily in the first q horizons. For other error structures, the improvement in the relative predictive efficiency can range from nearly zero to more than an order of magnitude. In addition, these results do not appear to be dependent on the stationarity (or lack thereof) of the exogenous variable.

The differences in predictive efficiencies between EGLS and EAR($\tilde{p}$), with few exceptions, is not very large. In fact when the error structure is modeled as an AR($\tilde{p}$) process with

TABLE I Relative Predictive Efficiencies Associated with AR(1) and MA(1) Error Processes ($\tilde{p} = [\sqrt{T}/2]$).

			$\varphi = 0.0$			$\varphi = 0.5$			$\varphi = 1.0$		
ϕ_1	T	h	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$
AR(1)											
0.9	100	1	1.187	1.059	0.233	1.074	1.030	0.239	1.087	1.083	0.201
		2	2.080	1.059	0.389	1.941	1.036	0.396	1.873	1.030	0.348
		5	3.975	1.087	0.719	3.651	1.062	0.749	4.292	1.042	0.743
		10	5.211	1.040	0.853	5.648	1.007	0.992	5.663	1.005	0.996
−0.9	100	1	1.126	1.009	0.232	1.041	0.990	0.186	0.999	0.990	0.180
		2	2.030	1.048	0.424	1.822	0.995	0.337	1.901	1.008	0.345
		5	3.377	0.997	0.614	3.955	1.044	0.717	3.721	1.039	0.691
		10	4.504	0.998	0.837	4.171	0.995	0.776	4.996	0.990	0.951
−0.5	100	1	1.131	0.996	0.860	1.103	1.082	0.739	1.086	0.990	0.865
		2	1.338	0.994	0.989	1.241	1.000	0.869	1.159	1.012	0.915
		5	1.378	1.012	0.997	1.289	1.017	0.970	1.249	0.998	0.951
		10	1.393	1.001	0.986	1.325	1.000	1.009	1.253	1.031	0.972
0.5	50	1	1.081	1.005	0.707	1.137	1.079	0.891	1.264	1.094	0.748
		2	1.354	1.066	0.828	1.433	1.045	0.992	1.478	1.059	0.851
		5	1.475	1.031	0.989	1.469	1.001	0.955	1.754	1.043	0.921
		10	1.480	0.995	0.999	1.501	0.949	1.056	1.911	1.006	0.966
	100	1	1.021	0.997	0.741	1.011	1.080	0.764	1.051	0.993	0.718
		2	1.245	1.001	0.876	1.344	1.094	0.922	1.343	0.996	0.911
		5	1.306	1.010	0.905	1.411	0.975	0.970	1.374	1.017	0.922
		10	1.361	1.019	0.983	1.466	1.024	1.043	1.403	1.012	0.949
	200	1	1.020	0.999	0.759	0.992	0.993	0.747	0.999	0.995	0.702
		2	1.233	1.004	0.913	1.263	1.017	0.881	1.186	1.006	0.958
		5	1.307	1.003	0.939	1.299	1.047	0.992	1.349	1.028	0.939
		10	1.344	1.004	0.940	1.323	0.998	0.983	1.370	1.010	1.010

MA(1)											
θ_1											
0.9	100	1	1.061	1.016	0.581	1.138	1.008	0.654	1.103	1.051	0.591
		2	1.792	0.997	0.965	1.664	0.992	0.891	1.734	0.998	0.950
		5	1.805	0.999	0.956	1.743	0.999	0.952	1.790	1.011	0.958
		10	1.815	1.003	0.966	1.952	1.009	1.063	1.843	1.046	0.988
−0.9	100	1	1.037	1.029	0.535	1.126	1.070	0.665	1.071	1.005	0.568
		2	1.701	1.043	0.836	1.910	1.069	0.994	1.755	0.996	0.922
		5	1.878	1.057	0.970	1.913	1.070	0.998	1.865	1.033	0.971
		10	1.988	1.088	0.999	1.863	1.028	1.015	1.901	1.043	0.993
0.5	100	1	1.028	1.035	0.862	1.120	1.043	0.807	1.015	1.042	0.851
		2	1.254	1.003	0.935	1.319	1.015	0.973	1.255	1.047	0.975
		5	1.302	1.035	0.935	1.320	1.005	0.995	1.327	1.020	0.975
		10	1.312	1.037	0.976	1.410	1.013	1.016	1.382	1.024	1.024
−0.5	50	1	1.060	1.046	0.870	1.072	1.059	0.798	1.127	1.035	0.690
		2	1.257	1.009	0.962	1.243	1.016	0.940	1.514	1.006	0.951
		5	1.346	1.030	1.005	1.261	1.008	0.920	1.599	1.040	0.976
		10	1.348	1.036	1.011	1.260	1.004	0.944	1.643	0.991	1.002
	100	1	1.063	0.995	0.812	1.064	0.973	0.795	1.076	1.036	0.791
		2	1.241	1.061	0.925	1.243	1.009	0.914	1.224	1.003	0.887
		5	1.335	1.021	0.988	1.281	1.022	0.938	1.234	1.006	0.949
		10	1.343	0.992	0.972	1.268	1.003	0.934	1.309	1.026	0.949
	200	1	1.026	1.032	0.769	1.014	0.997	0.756	1.075	1.078	0.916
		2	1.202	1.042	0.895	1.245	1.000	0.909	1.206	0.998	0.951
		5	1.260	1.056	0.941	1.255	1.039	0.994	1.232	1.017	0.984
		10	1.279	1.029	0.957	1.260	0.994	0.911	1.261	1.006	1.006

Note: PMSE, prediction mean squared error.

TABLE II Relative Predictive Efficiencies Associated with ARMA(1, 1) Error Processes.

				$\varphi = 0.0$			$\varphi = 0.5$			$\varphi = 1.0$		
(ϕ_1, θ_1)	T	$\tilde{p}$	h	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$
(0.8, −0.7)	100	5	1	1.127	1.071	0.159	1.068	1.024	0.165	1.059	1.007	0.144
			2	3.731	1.073	0.478	3.599	1.053	0.539	3.861	1.090	0.491
			5	7.472	0.995	0.982	7.882	1.094	0.912	7.424	1.057	0.923
			10	8.535	1.010	1.018	8.744	1.094	0.996	8.814	1.098	0.979
(0.8, 0.7)	100	5	1	1.002	0.991	0.884	0.992	0.994	0.938	1.050	1.052	0.976
			2	1.028	1.003	0.990	1.022	1.001	0.952	1.051	0.997	0.998
			5	1.080	0.996	0.954	1.109	0.998	0.979	1.051	1.005	1.011
			10	1.082	0.997	1.016	1.128	1.002	1.001	1.108	0.996	1.001
(−0.8, −0.7)	100	5	1	0.997	0.996	0.949	1.034	1.022	0.947	1.020	1.005	0.942
			2	1.056	1.002	0.974	1.070	1.046	0.977	1.059	1.030	0.957
			5	1.072	0.995	0.975	1.081	1.038	0.985	1.089	1.027	0.995
			10	1.112	1.011	0.983	1.084	1.038	0.993	1.098	1.005	0.998
(−0.8, 0.7)	50	1	1	1.458	1.187	0.193	1.662	1.535	0.244	1.604	1.273	0.205
			2	3.708	1.086	0.532	3.766	1.068	0.562	3.561	1.041	0.575
			5	7.976	1.136	1.031	6.898	1.068	0.912	6.845	1.075	0.955
			10	8.592	1.206	1.123	8.135	1.095	1.099	8.265	1.222	1.114
		4	1	1.235	1.006	0.163	1.154	1.066	0.170	1.286	1.021	0.164
			2	3.679	1.078	0.528	3.615	1.026	0.540	3.629	0.959	0.586
			5	7.151	1.018	0.924	6.761	1.047	0.894	6.487	1.019	0.905
			10	7.317	1.027	0.956	7.469	1.005	1.009	6.658	0.984	0.898
		7	1	1.265	1.030	0.167	1.264	1.167	0.186	1.495	1.187	0.191
			2	3.928	1.151	0.563	3.693	1.048	0.551	3.965	1.048	0.641
			5	6.964	0.991	0.900	7.264	1.124	0.960	6.891	1.082	0.961
			10	8.577	1.203	1.121	7.626	1.027	1.030	8.499	1.256	1.146

100	1	1	1.623	1.448	0.211	1.496	1.375	0.197	1.255	1.201	0.185
		2	4.059	1.261	0.543	3.642	1.105	0.510	3.236	1.139	0.468
		5	6.748	1.031	0.911	6.964	1.088	0.886	6.506	1.131	0.901
		10	6.989	0.990	1.049	7.481	1.084	0.981	7.882	1.182	1.145
	5	1	1.165	1.039	0.151	1.083	0.995	0.143	1.088	1.041	0.160
		2	3.280	1.019	0.439	3.349	1.016	0.469	3.295	1.058	0.477
		5	6.743	1.031	0.910	6.452	1.008	0.821	6.608	1.047	0.915
		10	7.146	1.012	1.032	7.221	1.046	0.947	6.717	1.007	0.976
	10	1	1.194	1.065	0.155	1.065	0.979	0.149	1.231	1.178	0.181
		2	4.023	1.249	0.538	3.207	0.973	0.499	3.435	1.103	0.497
		5	7.423	1.135	1.002	6.735	1.052	0.857	6.695	1.061	0.927
		10	8.474	1.200	1.247	7.997	1.158	1.049	7.211	1.082	1.048
200	1	1	1.593	1.536	0.222	1.370	1.377	0.190	1.451	1.386	0.181
		2	4.104	1.268	0.577	3.394	1.187	0.474	3.580	1.175	0.478
		5	6.974	1.101	0.935	6.878	1.217	1.011	6.386	1.126	0.858
		10	8.207	1.173	1.059	7.217	1.032	1.016	7.713	1.201	1.002
	7	1	1.057	1.019	0.147	1.003	1.008	0.139	1.099	1.050	0.137
		2	3.205	0.990	0.451	3.006	1.052	0.420	3.194	1.048	0.426
		5	6.299	0.994	0.845	5.827	1.031	0.857	6.593	1.059	0.886
		10	7.126	1.019	0.920	7.052	1.008	0.992	6.688	1.041	0.869
	14	1	1.074	1.036	0.149	1.105	1.110	0.153	1.17	1.117	0.146
		2	3.342	1.033	0.470	3.211	1.123	0.448	3.703	1.215	0.494
		5	7.047	1.112	0.945	6.213	1.099	0.913	6.396	1.128	0.859
		10	8.009	1.145	1.034	8.315	1.189	1.170	8.278	1.289	1.075

Note: PMSE, prediction mean squared error.

TABLE III Relative Predictive Efficiencies Associated with AR(2) Error Processes.

(ϕ_1, ϕ_2)	T	$\tilde{p}$	h	$\varphi = 0.0$ PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	$\varphi = 0.5$ PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	$\varphi = 1.0$ PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$
(1.42, −0.73)	100	5	1	1.077	0.993	0.165	1.058	1.005	0.156	1.106	0.996	0.170
			2	3.419	1.048	0.461	3.221	1.022	0.428	3.270	1.023	0.441
			5	5.852	1.002	0.848	5.657	1.002	0.722	5.812	1.000	0.843
			10	7.350	1.001	1.001	6.757	0.994	0.880	6.897	1.076	0.965
(1.8, −0.9)	100	5	1	1.205	0.986	0.022	1.206	0.998	0.023	1.188	1.058	0.024
			2	5.768	0.999	0.105	6.002	1.000	0.107	5.021	1.004	0.100
			5	41.08	0.966	0.675	38.37	1.048	0.692	36.33	1.073	0.687
			10	72.11	0.895	1.146	66.41	0.944	1.100	59.21	0.938	1.092
(1.6, −0.64)	50	1	1	1.817	1.817	0.052	1.735	1.605	0.055	1.881	1.630	0.053
			2	5.721	1.472	0.167	5.503	1.321	0.172	5.709	1.333	0.130
			5	21.06	1.063	0.552	20.21	1.050	0.469	21.75	1.023	0.434
			10	43.76	0.997	1.030	42.91	1.070	1.094	47.66	1.020	0.933
		4	1	1.087	1.087	0.031	1.087	1.006	0.034	1.147	0.994	0.033
			2	4.163	1.071	0.121	4.414	1.059	0.138	3.993	0.933	0.091
			5	19.70	0.995	0.516	20.03	1.040	0.465	21.78	1.025	0.435
			10	42.05	0.998	0.989	41.85	1.043	1.067	46.98	1.005	0.920
		7	1	1.327	1.327	0.038	1.333	1.233	0.042	1.309	1.134	0.037
			2	4.976	1.280	0.145	4.987	1.197	0.156	4.973	1.161	0.113
			5	22.31	1.126	0.585	20.13	1.046	0.467	21.99	1.034	0.439
			10	43.88	0.999	1.032	43.07	1.074	1.098	47.08	1.007	0.922

100	1	1	1.793	1.739	0.055	1.822	1.737	0.059	1.723	1.737	0.048
		2	5.747	1.529	0.207	5.661	1.423	0.220	5.610	1.504	0.151
		5	21.08	1.134	0.593	20.37	1.050	0.594	20.76	1.090	0.520
		10	43.58	1.085	1.083	42.22	1.018	1.121	45.34	1.192	1.137
	5	1	1.075	1.043	0.033	1.057	1.008	0.034	1.090	1.099	0.030
		2	4.063	1.081	0.146	3.979	1.000	0.155	3.922	1.052	0.105
		5	18.41	0.991	0.518	19.33	0.996	0.564	20.62	1.083	0.516
		10	39.51	0.983	0.982	41.12	0.991	1.092	40.93	1.076	1.026
	10	1	1.177	1.142	0.036	1.153	1.099	0.037	1.119	1.128	0.031
		2	4.319	1.149	0.155	4.091	1.028	0.159	4.565	1.224	0.123
		5	21.94	1.180	0.617	19.86	1.023	0.579	21.77	1.143	0.545
		10	43.36	1.079	1.078	42.27	1.019	1.123	43.22	1.136	1.084
200	1	1	1.664	1.602	0.047	1.602	1.485	0.048	1.502	1.560	0.047
		2	5.262	1.354	0.146	5.184	1.336	0.152	5.117	1.423	0.156
		5	20.77	1.175	0.517	20.23	1.197	0.592	20.25	1.185	0.559
		10	39.55	1.046	1.010	40.25	1.238	1.011	44.18	1.148	1.168
	7	1	1.031	0.992	0.029	1.064	0.986	0.032	1.006	1.045	0.031
		2	3.873	0.997	0.108	3.903	1.006	0.114	3.714	1.033	0.113
		5	17.69	1.001	0.441	17.85	1.056	0.522	17.19	1.006	0.475
		10	37.47	0.991	0.957	35.11	1.080	0.882	38.43	0.999	1.016
	14	1	1.110	1.068	0.032	1.120	1.038	0.034	1.082	1.124	0.034
		2	3.946	1.015	0.110	4.177	1.077	0.122	3.814	1.061	0.116
		5	17.58	0.994	0.438	18.64	1.103	0.545	17.59	1.029	0.486
		10	37.45	0.991	0.957	38.65	1.189	0.971	38.50	1.001	1.018

Note: PMSE, prediction mean squared error.

TABLE IV Relative Predictive Efficiencies Associated with MA(2) Error Processes.

				$\varphi = 0.0$			$\varphi = 0.5$			$\varphi = 1.0$		
(θ_1, θ_2)	T	$\tilde{p}$	h	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	PMSE $EAR(\tilde{p})$	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$
(1.8, −0.9)	100	5	1	1.445	1.096	0.276	1.436	1.099	0.257	1.387	1.038	0.258
			2	4.751	1.041	0.913	4.694	1.088	0.895	4.881	1.020	0.934
			5	5.281	0.996	0.940	5.015	1.037	0.924	5.092	1.027	0.968
			10	5.285	0.990	0.998	5.087	1.031	1.004	5.141	0.997	0.953
(1.6, −0.64)	100	5	1	2.824	1.041	0.341	2.631	1.063	0.312	2.733	1.026	0.333
			2	7.992	1.012	0.903	7.505	1.026	0.897	7.484	1.008	0.932
			5	8.301	1.010	0.973	8.222	1.019	0.987	7.677	1.003	0.912
			10	8.320	0.997	0.981	8.246	0.972	0.995	7.808	0.994	0.980
(1.42, −0.73)	50	1	1	2.078	1.597	0.542	2.100	1.602	0.557	2.253	1.686	0.558
			2	4.147	1.104	1.052	3.487	1.042	0.954	3.264	1.079	0.896
			5	4.351	1.122	1.122	3.672	1.094	1.028	3.843	1.051	0.981
			10	4.390	1.070	1.133	3.757	1.090	1.069	4.006	1.134	1.059
		4	1	1.354	1.041	0.353	1.314	1.002	0.349	1.331	0.996	0.330
			2	3.847	1.025	0.976	3.421	1.023	0.936	3.239	1.071	0.889
			5	3.889	1.003	1.003	3.530	1.052	0.989	3.860	1.055	0.985
			10	4.119	1.004	1.063	3.520	1.021	1.001	3.708	1.050	0.980
		7	1	1.364	1.048	0.356	1.369	1.044	0.363	1.493	1.118	0.370
			2	3.861	1.028	0.979	3.831	1.145	1.048	3.591	1.188	0.986
			5	4.050	1.044	1.044	4.287	1.277	1.201	4.385	1.199	1.119
			10	4.196	1.023	1.083	4.347	1.261	1.236	4.402	1.246	1.164

100	1	1	1.854	1.707	0.498	1.939	1.672	0.567	2.092	1.919	0.613
		2	3.449	1.168	0.949	3.586	1.170	0.997	3.084	1.020	0.908
		5	3.608	1.047	0.981	3.619	1.025	1.015	3.787	1.135	1.042
		10	3.817	1.082	1.030	3.664	1.039	1.031	3.830	1.112	1.053
	5	1	1.181	1.087	0.317	1.185	1.022	0.346	1.143	1.049	0.335
		2	3.079	1.043	0.847	3.302	1.078	0.918	3.082	1.019	0.907
		5	3.446	1.000	0.937	3.494	0.990	0.980	3.622	1.086	0.996
		10	3.554	1.007	0.959	3.516	0.997	0.989	3.639	1.057	1.001
	10	1	1.184	1.090	0.318	1.272	1.097	0.372	1.169	1.072	0.342
		2	3.480	1.179	0.958	3.657	1.194	1.017	3.260	1.078	0.960
		5	3.738	1.085	1.016	3.746	1.061	1.051	3.675	1.102	1.011
		10	4.088	1.158	1.103	3.842	1.090	1.081	3.728	1.082	1.025
200	1	1	1.815	1.771	0.508	1.850	1.616	0.515	2.163	2.050	0.678
		2	3.223	1.127	0.945	3.558	1.181	0.957	3.196	1.057	0.920
		5	3.568	1.029	1.001	3.661	1.125	0.998	3.682	1.192	0.999
		10	3.711	1.053	1.058	3.629	1.046	0.992	3.877	1.120	1.048
	7	1	1.035	1.010	0.290	1.179	1.030	0.328	1.144	1.084	0.359
		2	3.034	1.060	0.890	3.231	1.073	0.869	3.058	1.011	0.880
		5	3.469	1.000	0.973	3.377	1.038	0.920	3.308	1.071	0.898
		10	3.547	1.006	1.012	3.479	1.003	0.951	3.600	1.040	0.973
	14	1	1.084	1.058	0.304	1.180	1.031	0.329	1.166	1.105	0.365
		2	3.076	1.075	0.902	3.473	1.153	0.934	3.342	1.105	0.962
		5	3.495	1.008	0.980	3.769	1.158	1.027	3.791	1.227	1.029
		10	3.721	1.056	1.061	3.831	1.104	1.047	3.829	1.106	1.035

Note: PMSE, prediction mean squared error.

TABLE V Relative Predictive Efficiencies Associated with ARMA(1, 2) and ARMA(2, 1) Error Processes ($\tilde{p} = [\sqrt{T}/2]$).

	T	h	$\varphi = 0.0$			$\varphi = 0.5$			$\varphi = 1.0$		
			PMSE *EAR*($\tilde{p}$)	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	*PMSE* *EAR*($\tilde{p}$)	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$	*PMSE* *EAR*($\tilde{p}$)	$\hat{\xi}_{EAR(\tilde{p})/EGLS}$	$\hat{\xi}_{EAR(\tilde{p})/OLS}$
ARMA(1, 2) (ϕ_1, θ_1, θ_2)											
(0.6, −0.5, −0.9)	100	1	1.354	1.040	0.229	1.395	1.013	0.224	1.403	1.043	0.240
		2	4.118	1.028	0.724	4.177	1.016	0.632	3.911	0.998	0.663
		5	4.864	1.062	0.810	4.855	1.010	0.743	4.561	1.027	0.733
		10	5.857	1.011	1.002	5.695	1.093	0.969	5.912	1.097	0.974
(−0.8, 1.4, −0.6)	50	1	1.496	1.068	0.066	1.562	1.039	0.067	1.552	0.969	0.065
		2	7.346	1.062	0.342	7.537	1.049	0.341	6.862	1.014	0.368
		5	20.61	1.038	0.992	19.68	1.076	0.900	19.63	1.067	0.846
		10	20.82	0.999	1.009	22.54	1.000	1.082	24.80	1.064	1.054
	100	1	1.301	0.992	0.055	1.305	1.075	0.057	1.273	0.999	0.064
		2	7.192	1.029	0.317	6.004	1.050	0.278	6.801	1.009	0.354
		5	18.09	1.070	0.768	17.16	1.061	0.707	18.07	0.998	0.841
		10	20.95	1.017	0.892	21.10	0.994	0.909	20.81	1.052	0.945
	200	1	1.184	1.098	0.052	1.069	0.989	0.050	1.159	0.991	0.054
		2	5.894	1.047	0.268	6.000	1.002	0.276	6.465	1.070	0.312
		5	18.43	0.998	0.825	16.57	1.015	0.701	17.61	0.981	0.754
		10	20.37	1.001	0.988	20.34	1.005	0.898	20.91	1.087	0.923

ARMA(2, 1) (ϕ_1, θ_2, θ_1)											
(1.4, −0.6, −0.8)	100	1	1.213	1.042	0.063	1.166	1.074	0.058	1.117	1.063	0.059
		2	6.906	0.968	0.320	6.634	0.988	0.268	7.127	1.051	0.316
		5	25.26	1.004	0.908	21.37	0.990	0.768	24.17	1.004	0.936
		10	27.07	0.999	1.002	27.18	0.983	0.986	25.68	1.008	0.999
(−0.5, −0.9, 0.6)	50	1	1.329	0.996	0.132	1.242	0.993	0.139	1.277	0.997	0.117
		2	2.563	0.995	0.262	2.380	1.092	0.240	2.561	1.051	0.242
		5	4.248	1.060	0.440	4.187	0.994	0.473	4.444	1.021	0.420
		10	7.222	1.058	0.756	6.825	0.991	0.765	7.048	0.996	0.720
	100	1	1.089	1.002	0.116	1.104	0.999	0.112	1.068	1.022	0.116
		2	2.321	1.055	0.244	2.220	1.002	0.206	2.306	0.997	0.250
		5	4.018	1.053	0.427	3.947	0.999	0.366	3.890	1.003	0.442
		10	6.552	1.075	0.719	6.864	1.002	0.733	6.557	1.002	0.700
	200	1	0.996	0.993	0.103	1.004	0.997	0.096	1.096	1.061	0.118
		2	2.275	0.993	0.241	2.015	1.000	0.201	2.285	1.018	0.264
		5	3.857	1.076	0.419	3.814	1.000	0.365	3.768	0.998	0.444
		10	6.181	1.073	0.703	6.493	0.993	0.714	6.189	1.002	0.713

Note: PMSE, prediction mean squared error.

$\tilde{p} = [\sqrt{T}/2]$, the forecast performance of this correction is quite comparable to that of EGLS regardless of the actual simulated serial correlation structure, sample size, and stationarity condition of the exogenous variable. When $\tilde{p} = [\sqrt{T}/2]$, it is very infrequent that we observe values for $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{EGLS}}(T + h)$ to be greater than 1.06. (The greatest observed inefficiency of almost 10% occurred when the error structure followed an MA(2) process with a nearly non-invertible parameterization (1.8, − 0.9).) For the vast majority of the error structures we see that $1.01 \leq \hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{EGLS}}(T + h) \leq 1.05$ when $\tilde{p} = [\sqrt{T}/2]$. In many cases, the differences in efficiencies cannot be distinguished from sampling variation.

Moreover, it should also be pointed out that when $\tilde{p} = [\sqrt{T}/2]$, $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{OLS}}(T + h)$ is less than 1 (often considerably less than 1) for all but *basically* one case. For the nearly nonstationary AR(2) parameterization (1.8, −0.9) with 100 observations (Tab. III), $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{OLS}}(T + 10)$ equals 1.146, 1.1, and 1.092 for $\varphi = 0$, 0.5, and 1.0, respectively. With an increase in sample size, however, this seemingly small OLS superiority *vis-à-vis* EAR($\tilde{p}$) vanishes. For sample sizes of 200 and 400 observations (not shown in the table), for example, for $\pi = 0$, the corresponding values drop to 1.034 and 1.004; for $\varphi = 0.5$ (1.0), the respective figures are 1.049 and 1.009 (1.042 and 0.972). When the sample size is small, there are a few EAR($\tilde{p}$)/OLS efficiency ratios with values slightly above one for other error structures, but these too revert to values less than one as the sample size increases, as is illustrated by the ARMA(1, 1) parameterization (−0.8, 0.7) in Table II for $h = 10$. When T increases from 100 to 200 observations, the corresponding values for $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{OLS}}(T + 10)$ change from 1.032 to 0.920.

Results from this parameterization in particular also serve to further solidify the claim that for finite samples the order of the autoregressive correction should be set at $\tilde{p} = [\sqrt{T}/2]$ (Note the variation in $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{OLS}}(T + 10)$ when $\tilde{p}$ changes from 1 to $\sqrt{T}$.). For much shorter forecast horizons, however, it appears that even an EAR(1) correction, *i.e.* the most-often taught method for alleviating autocorrelation in regression models, can yield more accurate forecasts than ignoring the problem altogether.

Nevertheless, consistent with Theorem 4 we observe that as h increases, the differences in predictive efficiencies decrease among all methods. This can be more easily seen by examining the results in Table V, which contains the relative predictive efficiencies of ARMA(1, 2) and ARMA(2, 1) parameterizations. For the ARMA(1, 2) parameterization (−0.8, 1.4, −0.6) with $\varphi = 0$, $T = 200$, for instance, $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{EGLS}}$ and $\hat{\xi}_{\mathrm{EAR}(\tilde{p})/\mathrm{OLS}}$ change from 1.098 and 0.052, respectively, when $h = 1$, to 1.001 and 0.988 when $h = 10$.

4 CONCLUSION

In this article we have examined the relative forecast efficiency of GLS and IGLS predictors for regression models with serial correlation *vis-à-vis* each other and OLS predictors. We have derived new theorems associated with these predictors and established the form of the predictive mean squared errors as well as their magnitude relative to each other. From a large simulation study we have also found that for finite samples, EGLS corrections including those based on incorrectly identified disturbance structures yield more efficient short and medium-term forecasts than OLS. Furthermore, when the order of autoregressive corrections is set at $[\sqrt{T}/2]$ the differences in forecast efficiency between EAR($\tilde{p}$) and EGLS is very small. This suggests that there is not much to be gained in trying to identify the correct order of OLS residuals when generating short- to medium-term forecasts. Model builders upon detecting serial correlation in the residuals generated from an OLS regression should simply reestimate the regression equation using GLS with an AR($\sqrt{T}/2$) correction. Automatic procedures for performing this estimation without even providing initial estimates for the autoregressive coefficients can be found in most widely used statistical packages such as SAS (Autoreg) and Splus (ARIMA). On the other hand, for longer horizons, it appears that OLS yields forecasts that are just as efficient as EGLS.

References

Amemiya, T. (1973). Generalized least squares with an estimated autocovariance matrix. *Econometrica*, **41**, 723–732.

Armstrong, J. (1978). *Long-Range Forecasting: from Crystal Ball to Computer*. Wiley, New York.

Baillie, R. (1979). The asymptotic mean squared error of multistep prediction from the regression model with autoregressive errors. *Journal of American Statistical Association*, **74**, 175–184.

Berk, K. (1974). Consistent autoregressive spectral estimates. *The Annals Statistics*, **2**, 489–502.

Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals Statistics*, **6**, 224–231.

Breusch, T. (1980). Useful invariance results for generalized regression models. *Journal of Econometrics*, **13**, 327–340.

Chaudhury, A., Hubata, R. and Louis, R. (1999). Understanding time-series regression estimators. *The American Statistician*, **53**, 342–348.

Cochrane, D. and Orcutt, G. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of American Statistical Association*, **44**, 32–61.

Dielman, T. (1985). Regression forecasts when disturbances are autocorrelated. *Journal of Forecasting*, **4**, 263–271.

Durbin, J. and Watson, G. (1950). Testing for serial correction in least squares regression. *Biometrika*, **37**, 409–428.

Engle, R. (1974). Specification of the disturbance for efficient estimation. *Econometrica*, **42**, 135–146.

Fuller, W. (1996). *Introduction to Statistical Time Series*. John Wiley & Sons, Inc., New York.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of American Statistical Association*, **57**, 369–375.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Harville, D. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

Judge, G., Griffiths, W., Hill, R., Lütkepohl, H. and Lee, T. C. (1985). *The Theory and Practice of Econometrics*. 2nd ed., Wiley, New York.

Koreisha, S. and Fang, Y. (2001). Generalized least squares with misspecified serial correlation structures. *Journal of Royal Statistical Society, Series B*, **63**, 515–531.

Marcus, M. and Ming, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston.

Pierce, D. (1971). Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika*, **58**, 299–312.

Pukkila, T., Koreisha, S. and Kallinen, A. (1990). The identification of ARMA models. *Biometrika*, **77**, 537–549.

Schmidt, P. (1976). *Econometrics*. Marcel Dekker, New York.

Spitzer, J. (1979). Small-sample properties of nonlinear least squares and maximum likelihood estimators in the context of autocorrelated errors. *Journal of American Statistical Association*, **74**, 41–47.

Yamamoto, T. (1979). On the prediction efficiency of the generalized least squares model with an estimated variance covariance matrix. *International Economic Review*, **20**, 693–705.

Zinde-Walsh, V. and Galbraith, J. (1991). Estimation of a linear regression model with stationary ARMA(p, q) errors. *Journal of Econometrics*, **47**, 333–357.

APPENDIX

Proofs of Theorems and Corollaries

Proof of Corollary 2 Expanding the first row of (11) and noting that η' is just the first row of Λ^h, we see that

$$y^*_{(III)}(T+h) = (x_{1,T+h}, x_{2,T+h}, \ldots, x_{k,T+h})\beta + \eta'\mathcal{E}_T,$$

and

$$\text{PMSE}(y^*_{(III)}(T+h)) = E(\eta'\mathcal{E}_T - \varepsilon_{T+h})(\eta'\mathcal{E}_T - \varepsilon_{T+h})' = (\Upsilon\eta)'(\Upsilon\eta) - 2\eta'\tilde{V} + \sigma^2.$$

Hence,

$$\text{PMSE}(y^*_{(III)}(T+h)) - \text{PMSE}(y^*_{(II)}(T+h)) = (\Upsilon\eta)'(\Upsilon\eta) - 2\eta'\tilde{V}. \tag{A1}$$

Let $W = \Upsilon\eta$, $B = \Upsilon'^{-1}\tilde{V}$. Then,

$$\mathrm{PMSE}(y^*_{(III)}(T+h)) - \mathrm{PMSE}(y^*_{(II)}(T+h)) = W'W - 2W'B,$$

which can also be written as

$$\mathrm{PMSE}(y^*_{(III)}(T+h)) - \mathrm{PMSE}(y^*_{(II)}(T+h)) = W'W - 2W'GW,$$

where G is a diagonal matrix with its (i, i) elements given by (b_i/w_i), provided that $w_i \neq 0$. For $\mathrm{PMSE}(y^*_{(III)}(T+h)) - \mathrm{PMSE}(y^*_{(II)}(T+h))$ to be less than or equal to zero, it is necessary and sufficient to have $(2W'GW)/(W'W) \geq 1$, or equivalently, according to the Rayleigh–Ritz theorem (Marcus and Ming, 1964), for the smallest eigenvalue of $2G$ to be greater than or equal to 1. Since the smallest eigenvalue of $2G$ is $\min_i(2b_i/w_i)$, the Corollary follows.

Proof of Corollary 3 From (A1) and noting that $\tilde{V}' = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+\tilde{p}-1))$, $\eta' = (\pi_1^{(h)}, \pi_2^{(h)}, \ldots, \pi_{\tilde{p}}^{(h)})$, and $\Upsilon'\Upsilon = \{\gamma(i-j)\}_{i,j=1}^{\tilde{p}}$, it can be verified that

$$(\Upsilon\eta)'(\Upsilon\eta) = \sum_{i=1}^{\tilde{p}}\sum_{j=1}^{\tilde{p}} \gamma(i-j)\pi_i^{(h)}\pi_j^{(h)} \quad \text{and} \quad \eta'\tilde{V} = \sum_{i=1}^{\tilde{p}} \gamma(h+i-1)\pi_i^{(h)}.$$

Proof of Theorem 2 Since

$$\hat{\beta} - \beta = (X'\Xi^{-1}X)^{-1}X'\Xi^{-1}\varepsilon = A\varepsilon$$

and

$$Y - X\hat{\beta} = (I - XA)\varepsilon,$$

it follows that

$$\begin{aligned}
\mathrm{PMSE}(\hat{Y}^*) &= E(\hat{Y}^* - Y^*)(\hat{Y}^* - Y^*)' \\
&= E(X^*\hat{\beta} + C'(I-XA)\varepsilon - (X^*\beta + \varepsilon^*))(X^*\hat{\beta} + C'(I-XA)\varepsilon - (X^*\beta + \varepsilon^*))' \\
&= E([X^*A + C'(I-XA)]\varepsilon - \varepsilon^*)([X^*A + C'(I-XA)]\varepsilon - \varepsilon^*)' \\
&= \Omega^* + D\Omega D' - DV - V'D'.
\end{aligned}$$

Proof of Theorem 3 Part (1) follows from Theorem 1 in Koreisha and Fang (2001). To prove Eq. (2) in Theorem 3, note that

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T-l}\hat{\varepsilon}_t\hat{\varepsilon}_{t+l} &= T^{-1}\sum_{t=1}^{T-l}\varepsilon_t\varepsilon_{t+l} - T^{-1}\sum_{t=1}^{T-l}\varepsilon_t w_{t+h}(\hat{\hat{\beta}} - \beta) - T^{-1}\sum_{t=1}^{T-l}\varepsilon_{t+h}w_t(\hat{\hat{\beta}} - \beta) \\
&\quad + T^{-1}\sum_{t=1}^{T-l}(\hat{\hat{\beta}} - \beta)w_t'w_{t+h}(\hat{\hat{\beta}} - \beta). \qquad \text{(A2)}
\end{aligned}$$

Arguing as Fuller (1996) did in proving Theorem 9.3.1, we can show that

$$\sum_{t=1}^{T-l} \varepsilon_t w_{t+h} = O(T^{1/2}), \quad T^{1/2}(\hat{\hat{\beta}} - \beta) = O(1) \quad \text{and} \quad \sum_{t=1}^{T-l} w_t' w_{t+h} = O(T).$$

Hence, the last three terms on the right-hand-side of (A2) are $O(T^{-1})$. Since $T^{-1} \times \sum_{t=1}^{T-l} \varepsilon_t \varepsilon_{t+l} = \gamma(l) + O(T^{-1})$, the result in Eq. (2) follows.

Proof of Theorem 4 From Eq. (17) and assuming that β and Γ have been estimated consistently, it follows that

$$\text{APMSE}(\hat{\hat{Y}}^*) \equiv \underset{T\to\infty}{plim}\ \text{PMSE}(\hat{\hat{Y}}^*) \to E(\mathcal{C}'\varepsilon - \varepsilon^*)(\mathcal{C}'\varepsilon - \varepsilon^*) = \Omega^* + \mathcal{C}'\Omega\mathcal{C} - V'\mathcal{C} - \mathcal{C}'V.$$

Proof of Theorem 5 From Theorem 4 we know that as $T \to \infty$,

$$\text{APMSE}(\hat{\hat{Y}}^*) = \Omega^* + \mathcal{C}'\Omega\mathcal{C} - V'\mathcal{C} - \mathcal{C}'V.$$

Noting that $\text{APMSE}(\hat{\hat{Y}}^*_{\text{OLS}}) = \Omega^*$, it is only necessary to show that the (1, 1)th element of $\nabla \equiv \mathcal{C}'\Omega C - V'C - C'V$ goes to zero as $h \to \infty$ for $C' = V'\Omega^{-1}$ and ϱ'. When $\mathcal{C}' = V'\Omega^{-1}$, $\nabla = -V'\Omega^{-1}V$, and hence, the (1, 1)th element of ∇ converges to zero since the first row of $V' = E(\varepsilon^*\varepsilon') \to 0$ as h increases. In the case that $\mathcal{C}' = \varrho'$, the first row of ϱ' depends only on $\{\pi_i^{(h)}\}$, which goes to zero as $h \to \infty$ (Hamilton, 1994), and so does ∇.