Journal of Forecasting
J. Forecast. **18**, 95–109 (1999)

The Impact of Measurement Errors on ARMA Prediction

SERGIO G. KOREISHA* AND YUE FANG
University of Oregon, USA

ABSTRACT

Measurement errors can have dramatic impact on the outcome of empirical analysis. In this article we quantify the effects that they can have on predictions generated from ARMA processes. Lower and upper bounds are derived for differences in minimum mean squared prediction errors (MMSE) for forecasts generated from data with and without errors. The impact that measurement errors have on MMSE and other relative measures of forecast accuracy are presented for a variety of model structures and parameterizations. Based on these results the need to set up the models in state space form to extract the signal component appears to depend upon whether processes are nearly non-invertible or non-stationary or whether the noise-to-signal ratio is very high. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS autocovariance; forecasts; minimum mean square error

INTRODUCTION

Virtually all data contain observation errors. In addition, the measures often available to study economic and scientific phenomena are imperfect measurements of what is actually needed. The presence of measurement errors introduces added complexity in model identification since they will have an impact on the covariance structure of the observed data. Moreover, as demonstrated by Bell and Hilmer (1990) among others, parameter estimation is also very sensitive to the inclusion of error component factors in models.

Although much research has been conducted to deal with many of the problems associated with errors in variables, particularly in the area of parameter estimation or in the case of repeated time series measurements (Wong and Miller, 1990), as pointed in Wilcox (1992), most researchers do not spend much effort investigating the quality of the data, and often ignore measurement errors which may be important enough to influence the conclusions of empirical work as typically conducted.

* Correspondence to: Sergio G. Koreisha, Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA.

CCC 0277–6693/99/020095–15$17.50

Received October 1997
Accepted August 1998

The presence of measurement errors naturally will also have an effect on predictability. Consider, for example, the following discrete-time system,

$$y_t = x_t + \varepsilon_t \tag{1}$$

where y_t is the observed series, x_t is a stationary and invertible ARMA signal process and the series ε_t, the error component, also referred to as the noise, is a sequence of uncorrelated random variables with mean zero and variance σ_ε^2, assumed to be independent to x_t. Then, for a given specific signal process x_t with an ARMA (p,q) structure, the observed series, y_t, will be identified as an ARMA(p,Q) process with $Q \leqslant \max(p,q)$, which may be different from the ARMA(p,q) structure of $\{x_t\}$ (Granger and Morris, 1976).

In the case that $q < p$, the possible misspecification may introduce prediction bias. If $q \geqslant p$, the observed series may be identified correctly as an ARMA(p,q) process, but the extra volatility from the measurement errors will distort the parameter estimation, and hence may cause problems with prediction particularly for small samples.

If a model without measurement errors is used in a situation where measurement errors are present, the l-step ahead prediction of x_t based on the observed sequence $\{y_t\}$ is $\tilde{x}_t(l) = E(y_{t+l} \mid y_t, y_{t-1}, \ldots)$. The impact of measurement errors on prediction can then be defined as the difference between the minimum mean squared predictive errors (MMSE) of $\tilde{x}_t(l)$ and $\hat{x}_t(l)$, where $\hat{x}_t(l) = E(x_{t+l} \mid x_t, x_{t-1}, \ldots)$ is the prediction based on the signal $\{x_t\}$. Denote this difference as $\Delta MMSE$, that is,

$$\Delta MMSE = MMSE(\tilde{x}_t(l)) - MMSE(\hat{x}_t(l)) \tag{2}$$

Consequently,

$$\nabla MMSE \equiv \frac{\Delta MMSE}{MMSE(\hat{x}_t(l))} \tag{3}$$

provides a measurement of the relative impact of measurement errors on MMSE predictions.

Our study will focus on the effects that $\{\varepsilon_t\}$ have on $\Delta MMSE$ and $\nabla MMSE$. The results obtained here should provide useful guidelines for evaluating predictions based on ARMA models when measurement errors are present but not treated explicitly.

The paper is organized as follows. In the next section we investigate the problem of making linear predictions when measurement errors are present, and discuss some of the properties of $\Delta MMSE$ and $\nabla MMSE$. We also derive lower and upper bounds for both $\Delta MMSE$ and $\nabla MMSE$ for stationary linear models, and show that the results are only dependent on the autocovariance structures, and *not* on the model specifications of $\{x_t\}$ and $\{y_t\}$. In the third section, we derive exact expressions for $\Delta MMSE$ and $\nabla MMSE$ for some low-order ARMA processes, and show that the impact of measurement errors on MMSE predictions is not significant unless the signal process, $\{x_t\}$, is nearly non-invertible or non-stationary. In the fourth section we extend our findings to seasonal ARMA processes. In the final section we offer some conclusions and directions for further research.

MMSE PREDICTION WITH MEASUREMENT ERRORS

In this section we investigate the problem of predicting the value of x_{t+l} in terms of $\{y_i\}_{i=1}^{t}$, and discuss the properties of $\Delta MMSE$ and $\nabla MMSE$. We will start by assuming that $\{x_t\}$ is a stationary linear process, and then proceed to develop lower and upper bounds for these measures of predictability when the signal process follows a stationary and invertible ARMA process, $\Phi(B)x_t = \Theta(B)a_t$, where $\Phi(B)$ and $\Theta(B)$ are finite polynomials in the backshift operator B such that $B^j w_t = w_{t-j}$ and $\{a_t\}$ is a white noise process with variance σ_a^2. Such a signal process can also be expressed as autoregressive or moving average processes respectively, i.e. $\Pi(B)x_t = a_t$, or $x_t = \Psi(B)a_t$, where $\Pi = \Theta^{-1}\Phi$ and $\Psi = \Psi^{-1}\Theta$.

MMSE predictors

If measurement errors $\{\varepsilon_t\}$ are not present, the observed time series is simply $y_t = x_t$. The MMSE l-step ahead of prediction of x_t, based on a weighted average of previous observations and the forecasts made at previous lead times from the same origin, is defined in terms of the conditional expectation,

$$\hat{x}_t(l) = E(x_{t+l} \mid y_t, y_{t-1}, \ldots, y_1) = E(x_{t+l} \mid x_t, x_{t-1}, \ldots, x_1) \tag{4}$$

If we restrict ourselves to the class of linear predictors, then

$$\hat{x}_t(l) = E(x_{t+l} \mid y_t, y_{t-1}, \ldots, y_1) = E(x_{t+l} \mid x_t, x_{t-1}, \ldots, x_1) = \sum_{i=1}^{t} \pi_{i,x}^{l} x_i \tag{5}$$

where $\Pi_x^l \equiv (\pi_{1,x}^l, \pi_{2,x}^l, \ldots, \pi_{t,x}^l)'$ is the weight vector estimated from the data which will yield the MMSE for $\hat{x}_t(l)$ (We have added the subscript x to the Π's to emphasize that we are dealing with the series $\{x_t\}$. The theoretical expectation (5) requires knowledge of the x_i's going all the way back to the infinite past. However, because we have assumed that the model is invertible, the Π weights in (5) form a convergent series. Consequently, for computational purposes, given a certain level of accuracy, the dependence on the distant past should be negligible.)

If the model contains measurement errors, the l-step ahead of linear prediction of x_t, $\tilde{x}_t(l)$, will be based on $\{y_t\}$, which will be distorted by the noise $\{\varepsilon_t\}$, and will be defined as

$$\tilde{x}_t(l) = E(y_{t+l} \mid y_t, y_{t-1}, \ldots, y_1) = \sum_{i=1}^{t} \pi_{i,y}^{l} y_i \tag{6}$$

Although the forecasts obtained from equation (6) are unbiased, as we shall show, $\Delta MMSE \neq 0$.

Bounds for $\Delta MMSE$ and $\nabla MMSE$

Let $X_t = (x_t, x_{t-1}, \ldots, x_1)'$, $Y_t = (y_t, y_{t-1}, \ldots, y_1)'$, $\mathcal{R}_{t,x}^l = (\gamma_x(l), \gamma_x(1+l), \ldots, \gamma_x(t-1-l))'$, where $\gamma_x(\cdot)$ is the autocovariance function of x_t, and $\Gamma_{t,x} = [\gamma_x(i-j)]_{i,j=1,2,\ldots,t}$. Then, the Yule–Walker-type estimators for Π_x^l are given by $\Pi_x^l = \Gamma_{t,x}^{-1}\mathcal{R}_{t,x}^l$. Thus,

$$\hat{x}_t(l) = (\Pi_x^l)' X_t = (\Gamma_{t,x}^{-1}\mathcal{R}_{t,x}^l)' X_t \tag{7}$$

and consequently, the MMSE of equation (7) becomes

$$E(\hat{x}_t(l) - x_{t+l})^2 = \gamma_x(0) - \mathcal{R}^{l}_{t,x}{}'\Gamma^{-1}_{t,x}\mathcal{R}^{l}_{t,x} \tag{8}$$

(Corollary 5.1.1 in Brockwell and Davis, 1991).

Now let $\mathcal{R}^l_{t,y}$, $\Gamma_{t,y}$ and Π^l_y be similarly defined for y_t. Hence, $\Pi^l_y = \Gamma^{-1}_{t,y}\mathcal{R}^l_{t,y} = \Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x}$ since the autocovariance functions for $\{x_t\}$ and $\{y_t\}$ are the same. Moreover,

$$\tilde{x}_t(l) = (\Pi^l_y)'Y_t = (\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x})'Y_t \tag{9}$$

and the MMSE of equation (9) can be calculated as follows:

$$\begin{aligned} E(\tilde{x}_t(l) - x_{t+l})^2 &= E(\tilde{x}_t(l) - y_{t+l} + \varepsilon_{t+l})^2 \\ &= E(\tilde{x}_t(l) - y_{t+l})^2 + 2E(\tilde{x}_t(l) - y_{t+l})\varepsilon_{t+l} + E\varepsilon^2_{t+l} \\ &= \gamma_y(0) - \mathcal{R}^l_{t,y}{}'\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,y} - 2\sigma^2_\varepsilon + \sigma^2_\varepsilon \\ &= \gamma_x(0) + \sigma^2_\varepsilon - \mathcal{R}^l_{t,y}{}'\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,y} - \sigma^2_\varepsilon \\ &= \gamma_x(0) - \mathcal{R}^l_{t,x}{}'\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x} \end{aligned} \tag{10}$$

Hence,

$$\Delta MMSE = \mathcal{R}^l_{t,x}{}'\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} - \mathcal{R}^l_{t,x}{}'\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x} \tag{11}$$

and

$$\nabla MMSE = \frac{\mathcal{R}^l_{t,x}{}'\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} - \mathcal{R}^l_{t,x}{}'\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x}}{MMSE(\hat{x}(l))} \tag{12}$$

The conditions that $\gamma_x(0) > 0$ and $\gamma_x(h) \to 0$ as $h \to \infty$ are sufficient to ensure that both $\Gamma_{t,x}$ and $\Gamma_{t,y}$ are non-singular for every t (Proposition 5.1.1 in Brockwell and Davis, 1991). It should be noted that simplification of the matrix $[\Gamma^{-1}_{t,x} - \Gamma^{-1}_{t,y}]$ is, in general, very difficult, even though $[\Gamma_{t,y} - \Gamma_{t,x}]$ is of the form $\sigma^2_\varepsilon I$, where I is a $t \times t$ identity matrix. The following theorem, which is the main result in this section, provides lower and upper bounds for $\Delta MMSE$ and $\nabla MMSE$.

Theorem 1 If $\Gamma_{t,x}$ is positive definite, then

$$\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + max_i\lambda_i}B_{1,t} \leqslant \Delta MMSE \leqslant \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + min_i\lambda_i}B_{1,t} \tag{13}$$

and

$$\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + max_i\lambda_i}B_{2,t} \leqslant \nabla MMSE \leqslant \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + min_i\lambda_i}B_{2,t} \tag{14}$$

where $\lambda_i s$ are eigenvalues of $\Gamma_{t,x}$, $B_{1,t} = \gamma_x(0) - MMSE(\hat{x}(l))$ and $B_{2,t} = [\gamma_x(0) - MMSE(\hat{x}(l))]/MMSE(\hat{x}(l))$.

The proof of the theorem is given in the Appendix.

As can be seen, the bounds depend on σ_ε^2, and on the quality of the predictor $\hat{x}(l)$, which is measured by $B_{1,t}$ or $B_{2,t}$. In addition, the distribution of eigenvalues of $\Gamma_{t,x}$ also plays a critical role in the evaluation of $\Delta MMSE$ and $\nabla MMSE$. Furthermore, in the following sections, we will see that the eigenvalues of $\Gamma_{t,x}$ are closely related to process stationarity and invertibility.

In addition, the expected result that $\Delta MMSE \geqslant 0$ and $\nabla MMSE \geqslant 0$, which we state below as a corollary, can be easily verified, because γ_i's, $B_{1,t}$ and $B_{2,t}$ in equations (13) and (14) are non-negative.

Corollary 1 If $\Gamma_{t,x}$ is positive definite, then

$$\Delta MMSE \geqslant 0 \text{ and } \nabla MMSE \geqslant 0$$

As measurement errors decrease, y_t converges to x_t, and the effects of measurement errors on predictions vanish. This result is a direct consequence of Theorem 1, which can be stated as follows:

Corollary 2 As $\sigma_\varepsilon \to 0$,

$$\Delta MMSE \to 0 \text{ and } \nabla MMSE \to 0$$

The results derived thus far are very general and depend only on the autocovariance structures. In the subsection below, we will assume that $\{x_t\}$ follows an ARMA process.

Lower and upper bounds for some low-order ARMA processes

The derivation of theoretical lower and upper bounds for $\Delta MMSE$ and $\nabla MMSE$ for general ARMA(p,q) models is very difficult and tedious. Here we provide derivations for just two processes—AR(1) and MA(1). Higher-order processes will be discussed in the following two sections. There we will obtain numerical values for both $\Delta MMSE$ and $\nabla MMSE$ based on some specific model parameterizations.

Example 1 Consider the case that $\{x_t\}$ follows an AR(1) process. Let $x_t = \phi x_{t-1} + a_t$ with $\sigma_a^2 = 1$ and define

$$f(z) = \frac{1 - \phi^2}{1 - 2\phi \cos(\mathrm{z}) + \phi^2}$$

Then the smallest and largest eigenvalues of $\Gamma_{t,x}$ are

$$\lambda_{min} \approx \frac{1}{1 - \phi^2} min_{z\in(0,\pi)} f(z) = \frac{1}{(1 + |\phi|)^2}$$

and

$$\lambda_{max} \approx \frac{1}{1 - \phi^2} max_{z\in(0,\pi)} f(x) = \frac{1}{(1 - |\phi|)^2}$$

(Grenander and Szegö, 1958, see also Proposition 4.5.3 in Brockwell and Davis, 1991 for results on λ_{min} and λ_{max} for more general ARMA processes). Hence, by Theorem 1,

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \dfrac{1}{(1-|\phi|)^2}} B_{1,t} \leqslant \Delta MMSE \leqslant \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \dfrac{1}{(1+|\phi|)^2}} B_{1,t} \tag{15}$$

and

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \dfrac{1}{(1-|\phi|)^2}} B_{2,t} \leqslant \Delta MMSE \leqslant \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \dfrac{1}{(1+|\phi|)^2}} B_{2,t} \tag{16}$$

In general, $B_{1,t} \geqslant B_{2,t}$ since $MMSE(\hat{x}_t(l)) \geqslant \sigma_a^2 = 1$. If $l = 1$, we have $B_{1,t} = B_{2,t} = \gamma_x(0) - MMSE(\hat{x}_t(1)) = \gamma_x(0) - \sigma_a^2 = 1/(1-\phi^2) - 1 = \phi^2/(1-\phi^2)$.

When $|\phi|$ is small, equations (15) and (16) provide tight limits for $\Delta MMSE$ and $\nabla MMSE$. As $|\phi|$ goes to zero, the lower and upper bounds in equation (15) approach

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 1} B_{1,t}$$

the lower and upper bounds in equation (16) approach

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 1} B_{2,t}$$

When ϕ increases, the range between the lower and upper bounds increases. As $|\phi|$ approaches one, the lower bounds in equations (15) and (16) go to zero, while the upper bounds approach

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 0.25} B_{1,t}$$

and

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 0.25} B_{2,t}$$

respectively.

Example 2 In the second example, we assume that $\{x_t\}$ follows an MA(1) process. Let $x_t = a_t - \theta a_{t-1}$ and $\sigma_a^2 = 1$. Then the eigenvalues of $\Gamma_{t,x}$ (see, Gregory and Karney, 1969) are

$$\lambda_i = \gamma_x(0) + 2\gamma_x(1)\cos\left(\frac{k\pi}{n+1}\right) \qquad k = 1, 2, \ldots, n$$

where $\gamma_x(0) = (1 + \theta^2)$ and $\gamma_x(1) = -\theta$. Hence, $\lambda_{max} \approx (1 + |\theta|)^2$ and $\lambda_{min} \approx (1 - |\theta|)^2$. Applying Theorem 1, we have

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1 + |\theta|)^2} B_{1,t} \leqslant \Delta MMSE \leqslant \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1 - |\theta|)^2} B_{1,t} \tag{17}$$

and

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1 + |\theta|)^2} B_{2,t} \leqslant \nabla MMSE \leqslant \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1 - |\theta|)^2} B_{2,t} \tag{18}$$

When $l = 1$, $B_{1,t} = B_{2,t} = \gamma_x(0) - MMSE(\hat{x}_t(1)) = \gamma_x(0) - \sigma_a^2 = (1 + \theta^2) - 1 = \theta^2$. If $l > 1$, $B_{1,t} = B_{2,t} = 0$ since $MMSE(\hat{x}_t(l)) = MMSE(\tilde{x}_t(l))$.

When $l = 1$, if $|\theta|$ is small, the range associated with equations (17) and (18) is very narrow. As $|\theta|$ goes to zero, the lower and upper bounds in these equations approach

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 1}\theta^2$$

When θ increases, the difference between the lower and upper bounds increases. As $|\theta|$ goes to one, the lower and upper bounds in equations (17) and (18) approach

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + 4}\theta^2 \text{ and } \theta^2,$$

respectively.

Model misspecification and MMSE predictors

Granger and Morris (1976) have proven that the sum of two uncorrelated processes, ARMA(p_1, q_1) and $ARMA(p_2, q_2)$ is an ARMA(p^*, q^*) process, where $p^* \leqslant p_1 + p_2$ and $q^* \leqslant max(p_1 + q_2, p_2 + q_1)$. Thus, as pointed out earlier, given a specific ARMA (p, q) signal process $\{x_t\}$, the observed series $\{y_t\}$ will be identified as an ARMA(p, Q) process, where $Q \leqslant max(p, q)$.

If $q < p$, the number of past observations used in equations (5) and (6) to predict x_{t+l} may be different because of the structural difference between $\Gamma_{t,x}^{-1}$ and $\Gamma_{t,y}^{-1}$. If $q \geqslant p$, the observed series will probably be correctly identified. However, the projection coefficients $\Pi_{t,x}^l$ and $\Pi_{t,y}^l$ will be different due to the impact of σ_ε^2 in $\Gamma_{t,y}^l$.

Example 3 Suppose that $\{x_t\}$ follows an AR(1) process, $x_t - \phi x_{t-1} = a_t$. If there is no measurement error, only the last observation x_t is useful in predicting x_{t+l}. Therefore, the projection coefficients of X_t, $\Pi_x^l = (1, 0, 0, \ldots, 0)'$ and $\hat{x}_t(l) = \phi^l x_t$. If the measurement errors are not zero, y_t will follow an ARMA(1,1) process. Consequently, $\pi_{i,y}^l$ will not be equal to zero for all $i = t, t-1, \ldots, 1$, hence the whole past history of $\{y_t\}$ will be used in the calculation of the MMSE prediction of y_{t+l}. The projection coefficients of y_t, i.e. $\pi_{i,y}^l$, approach zero asymptotically instead of cutting off abruptly.

Example 4 Assume that $\{x_t\}$ follows MA(1) process, that is $x_t = a_t - \theta_1 a_{t-1}$ with $\sigma_a^2 = 1$. Then $y_t = x_t + \varepsilon_t$ will also follow an MA(1) process. Now assume that $y_t = b_t - \vartheta b_{t-1}$, where $\{b_t\}$ is a white noise process uncorrelated with $\{a_t\}$. The non-zero autocovariances of y_t are given by $\gamma_t(0) = (1 + \theta^2) + \sigma_\varepsilon^2 = (1 + \vartheta^2)\sigma_b^2$ and $\gamma_y(1) = -\theta = -\vartheta\sigma_b^2$. These two equations yield

$$(1 + \theta^2) + \sigma_\varepsilon^2 = (1 + \vartheta^2)\theta/\vartheta \tag{19}$$

Using equation (19) one can easily show that ϑ and θ have equal sign and that $|\vartheta| \leqslant |\theta|$. This implies that the projection coefficients Π_y^l and $\tilde{x}_t(l)$ are determined by an MA(1) process with $|\vartheta| \leqslant |\theta|$ and $\sigma_b^2 \geqslant 1$ $(=\sigma_a^2)$. Both $\Gamma_{t,x}$ and $\Gamma_{t,y}$ are tridiagonal matrices. A closed form of the inverse of tridiagonal matrices has been given by Gregory and Karney 1969). Thus, applying Gregory and Karney's result, one can find both $\Gamma_{t,x}^{-1}$ and $\Gamma_{t,y}^{-1}$.

PREDICTIONS BASED ON ARMA MODELS

In this section, we derive exact values for $\Delta MMSE$ and $\nabla MMSE$ associated with some low-order ARMA models. We have chosen model parameterizations of $\{x_t\}$ which not only conform to other previously published studies, such as Pukkila, Koreisha and Kallinen (1990) but which also cover a wide spectrum of parameter values ranging from well-defined stationary and invertible processes to nearly non-invertible or non-stationary processes. Without loss of generality, we will assume that the variance of $\{a_t\}$, σ_a^2, is one.

The assumption $\sigma_a^2 = 1$ implies that if $l = 1$, $MMSE(\hat{x}(l)) = \sigma_a^2 = 1$. Hence, we have $\Delta MMSE = \nabla MMSE$ for the process without an MA part and $\Delta MMSE \approx \nabla MMSE$ for the process with an MA part. When $l > 1$, $\Delta MMSE > \nabla MMSE$ since $MMSE(\hat{x}(l)) > 1$.

To provide a measure of the magnitude of the measurement error relative to the signal process, let us define the noise-to-signal ratio as $\rho = \sigma_\varepsilon/\sigma_x$. In this study, we will evaluate the consequences on several-steps-ahead forecasts when ρ will be set equal to 1%, 10% or 25%. Since $\nabla MMSE$ is calculated in terms of second moments, we will compare $\nabla MMSE$ with ρ^2 instead of ρ. The corresponding ρ^2 values are 0.0001, 0.01 and 0.0625, respectively. These three levels of ρ (or ρ^2) cover a wide range of situations usually encountered in practice (Bell and Wilcox, 1993).

It should be noted that if x_t is an AR(p) process, $\nabla MMSE$ does not depend on the sample size t as long as $t > p$. This is because the only last p observations are used for prediction. On the other hand, if x_t has an MA component, $\nabla MMSE$ does depend on t since all past observations are required to calculate the MMSE prediction. However, these statistics are not noticeably affected by the sample size, especially when roots of the polynomials in equation (4) are not near the unit circle. Hence, for brevity, we will report results only for $t = 100$. For illustrative purposes, however, we will also provide rests on $t = 400$ for MA(1) structures.

Tables I–V contain $\nabla MMSE$'s for some of the low-order ARMA(p,q) structures we studied for several prediction horizons, l, and levels of ρ^2. (For brevity we have omitted the results associated with $\Delta MMSE$ for most of the ARMA(p,q) structures in these tables. Table IV, however, contrasts $\Delta MMSE$ and $\nabla MMSE$ for various ARMA(1,1) parameterizations). In general we see that the impact that measurement errors have on MMSE predictions depends mainly on the invertibility and stationary properties of the $\{x_t\}$ processes, and naturally, on the magnitude of ρ^2. Unless the roots of the characteristic polynomials are near the unit circle, the impact is relatively negligible and concentrated primarily around the first few-steps-ahead forecasts. When the processes are nearly non-stationary, as exemplified by the AR(1) process with $|\phi| \geqslant 0.99$ and the AR(2) process with $\phi_1 = 1.8$, and $\phi_2 = -0.9$, the impact of measurement

Table I. $\nabla MMSE$ of AR(1) and AR(2) processes for several forecast horizons and different levels of measurement errors

		AR(1): ϕ_1				AR(2): (ϕ_1, ϕ_2)			
ρ^2	l	0.30 or −0.30	0.50 or −0.50	0.90 or −0.90	0.99 or −0.99	(1.42, −0.73)	(1.80, −0.90)	(0.50, 0.30)	(−0.30, 0.50)
0.0001	1	9.889e-06	3.333e-05	0.0004261	0.0049006	0.0016699	0.0204784	2.627e-05	7.082e-05
	2	8.165e-07	6.666e-06	0.0001907	0.0024257	0.0005925	0.0096284	5.832e-05	7.082e-05
	5	5.904e-10	9.774e-08	5.351e-05	0.0009411	1.457e-05	0.0022898	1.350e-05	1.444e-05
	10	3.486e-15	9.535e-11	1.383e-05	0.0004469	2.075e-06	8.338e-05	2.390e-06	3.376e-06
0.01	1	0.0009783	0.0032896	0.0405791	0.3596025	0.1455287	1.1615970	0.0075236	0.0069681
	2	8.077e-05	0.0006579	0.0181597	0.1779943	0.0499810	0.4802442	0.0057359	0.0069593
	5	5.841e-08	9.647e-06	0.0050957	0.0690587	0.0014318	0.0984472	0.0013287	0.0014225
	10	3.449e-13	9.412e-09	0.0013174	0.0327956	0.0001962	0.0059951	0.0002332	0.0003324
0.0625	1	0.0057862	0.0192588	0.2094753	1.3018980	0.6219203	3.8795890	0.0439953	0.0402842
	2	0.0004778	0.0038518	0.0937431	0.6444068	0.1951217	1.3721710	0.0330805	0.0399863
	5	3.455e-07	5.648e-05	0.0263046	0.2500185	0.0083004	0.2300115	0.0077032	0.0082634
	10	2.040e-12	5.510e-08	0.0068006	0.1187326	0.0010888	0.0252748	0.0013516	0.0019278

Notes:
(1) The observations follow $y_t = x_t + \varepsilon_t$.
(2) In the AR(1) case, $x_t - \phi_1 x_{t-1} = a_t$ with $\sigma_a^2 = 1$ and

$$\sigma_\varepsilon^2 = \rho^2 \frac{1}{1-\phi_1^2}.$$

(3) In the AR(2) case, $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = a_t$ with $\sigma_a^2 = 1$ and

$$\sigma_\varepsilon^2 = \rho^2 \frac{1-\phi_2}{(1+\phi_2)((1-\phi_2)^2-\phi_1^2)}.$$

(4) The result does not depend on the sample size t; results shown here are for $t = 100$ observations.

errors as measured by $\nabla MMSE$ for the first-step-ahead forecast can be quite dramatic: 36% when $\rho = 10\%$, and 130% when $\rho = 25\%$ for the AR(1) process; and 116% when $\rho = 10\%$, and 388% when $\rho = 25\%$ for the AR(2) process. Note, however, how small the impact becomes, even for these processes, when the forecast horizon increases to ten periods: 3.28% and 11.9% for the AR(1) process, and 0.6% and 2.53% for the AR(2) process, when $\rho = 10\%$ and 25%, respectively. For pure MA processes the impact of measurement errors are only felt for forecast horizons $l \leqslant q$ since $\gamma_x(l)$ is zero for $l > q$. As can be seen, the impact of measurement errors in the model prediction follows a pattern very similar to pure AR processes, more noticeably when the processes are nearly non-invertible, but they are generally of lower magnitude.

For mixed processes the impact also appears to be dependent on model parameterization. Those structures which when converted into equivalent AR(∞) representations have π-weights which go to zero relatively slowly as the lag length increases, feel more of the impact of measurement errors than those for which the π-weights go to zero rapidly. Note, for example, when $l = 1$, how large is the difference in $\nabla MMSE$ for $\rho = 10\%$ and $\rho = 25\%$, respectively for ARMA(1,1) processes with $\phi = 0.8$, $\theta = 0.7$ ($\pi_1 = 0.1$, $\pi_2 = 0.1 \times 0.7$, $\pi_3 = 0.1 \times 0.7^2, \ldots$) and with $\phi = 0.8$ and $\theta = -0.7$ ($\pi_1 = 1.5$, $\pi_2 = 1.5 \times 0.7$, $\pi_3 = 1.5 \times 0.7^2, \ldots$). The impact of the measurement errors also decreases dramatically as the forecast horizon increases. For the

Table II. $\nabla MMSE$ of MA(1) process for several forecast horizons and different levels of measurement errors ($t = 100$ and $t = 400$)

		θ_1							
		0.30 or −0.30		0.50 or −0.05		0.90 or −0.90		0.99 or −0.99	
ρ^2	l	$t = 100$	$t = 400$	$t = 100$	$t = 400$	$t = 100$	$t = 400$	$t = 100$	$t = 400$
0.0001	1	1.078e-05	1.078e-05	4.166e-05	4.166e-05	0.0007678	0.0007678	0.0051690	0.0071218
	2, 5, 10	0	0	0	0	0	0	0	0
0.01	1	0.0010640	0.0010640	0.0040766	0.0040766	0.0555977	0.0559775	0.1175039	0.1208585
	2, 5, 10	0	0	0	0	0	0	0	0
0.0625	1	0.0062287	0.0062287	0.0229490	0.0229490	0.1869673	0.1869673	0.2799237	0.2837658
	2, 5, 10	0	0	0	0	0	0	0	0

Notes:
(1) The signal is an MA(1) process $x_t = a_t - \theta_1 a_{t-1}$ with $\sigma_a^2 = 1$ and the observations follow $y_t = x_t + \varepsilon_t$.
(2) $\sigma_\varepsilon^2 = \rho^2(1 + \theta_1^2)$.
(3) The sample size effects are negligible, especially when the signal process is not nearly non-invertible. When the signal process is nearly non-invertible ($|\theta_1| \geqslant 0.9$), some of $\nabla MMSE$s at $t = 400$ are greater than those at $t = 100$ due to the different decay speeds of $MMSE(\tilde{x}_t(l))$ and $MMSE(\hat{x}_t(l))$. For example, when $|\theta_1| = 0.99$, $MMSE(\tilde{x}_t(1)) = 1.008193$ and $MMSE(\hat{x}_t(1)) = 1.003008$ and for $t = 100$, and $MMSE(\tilde{x}_t(1)) = 1.007128$ and $MMSE(\hat{x}_t(1)) = 1.000006$ for $t = 400$.

ARMA(1,2) structure with $\phi_1 = -0.8$, $\theta_1 = 1.4$ and $\theta_2 = -0.6$ for $\rho = 10\%$, $\nabla MMSE$ changes from 115% when $l = 1$ to 0.22% when $l = 10$; for the ARMA(2,1) with $\phi_1 = -0.5$, $\phi_2 = -0.9$ and $\theta = 0.6$, the corresponding values when $\rho = 25\%$ are 87% and 4.1% for $l = 1$ and 10 respectively.

Finally, as mentioned earlier, sample size does not have a noticeable effect on the prediction measures as can be seen by contrasting the values of these measures for $t = 100$ and $t = 400$ in Table II.

SEASONAL ARMA PROCESSES

Seasonal models can be viewed as special forms of the ARMA models. Results for low-order non-seasonal ARMA processes in the previous sections, as we shall demonstrate, can be used to show the impact that measurement errors have on the predictability of seasonal models. To illustrate how one can apply the results for low-order non-seasonal ARMA models to seasonal ARMA processes, we will consider two often used multiplicative structures.

Example 5 Suppose that x_t is an ARMA(0,1) × SARMA$(1,0)_{12}$ process, that is $x_t = \Phi x_{t-12} + a_t - \theta a_{t-1}$. Then $\gamma_x(0) = (1 + \theta^2)/(1 - \Phi^2)$, $\gamma_x(12k) = \Phi^k \gamma_x(0)$, and $\gamma_x(12k - 1) = \gamma_x(12k + 1) = -\theta\Phi^k\gamma_x(0)/(1 + \theta^2)$ for $k = 1, 2, \ldots$, $\gamma_x(\cdot) = 0$ at all other lags. Define the ARMA(1,1) process $\{x_t^*\}$ as $x_t^* = \Phi x_{t-1}^* + a_t - \theta a_{t-1}$. If we compare the covariances of $\{x_t\}$ with those of $\{x_t^*\}$, we see that

$$MMSE(\hat{x}_t(l)) = MMSE(\hat{x}_t^*(l^*)) \quad \text{for } 12(l^* - 1) < l \leqslant 12l^*, \qquad l^* = 1, 2, 3, \ldots \tag{20}$$

Table III. $\nabla MMSE$ of MA(2) and MA(3) processes for several forecast horizons and different levels of measurement errors

		MA(2): (θ_1, θ_2)				MA(3): $(\theta_1, \theta_2, \theta_3)$			
ρ^2	l	(1.42, −0.73)	(1.8, −0.9)	(0.5, 0.3)	(−0.3, 0.5)	(1.0, 0.8, −0.9)	(0.5, 0.3, −0.3)	(0.8, 0.8, −0.64)	(−0.4, −0.55, −0.22)
0.0001	1	0.0019661	0.0228484	0.0001665	0.001450	0.0211620	8.482e-05	0.0058830	6.506e-05
	2	0.0004091	0.0043595	2.163e-05	6.398e-05	0.0007031	1.511e-05	0.0002882	5.647e-05
	3	0	0	0	0	0.0064248	1.530e-05	0.0010984	7.152e-06
	5	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0
0.01	1	0.1495970	0.5493175	0.0150871	0.0131661	0.3621351	0.0082420	0.1970614	0.0063627
	2	0.0275923	0.0716249	0.0019943	0.0059355	0.0290324	0.0014899	0.0218090	0.0054987
	3	0	0	0	0	0.0866441	0.0014806	0.0328719	0.0006956
	5	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0
0.0625	1	0.5269499	1.2148580	0.0670375	0.0589886	0.7721829	0.0450623	0.4749465	0.0358288
	2	0.0756443	0.1155425	0.0094341	0.0286489	0.1181814	0.0087035	0.0947855	0.0303103
	3	0	0	0	0	0.1515817	0.0079593	0.0703101	0.0038140
	5	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0

Notes:
(1) The observations follow $y_t = x_t + \varepsilon_t$.
(2) For the MA(2) case, $x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ with $\sigma_a^2 = 1$ and $\sigma_\varepsilon^2 = \rho^2(1 + \theta_1^2 + \theta_2^2)$.
(3) For the MA(3) case, $x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$ with $\sigma_\varepsilon^2 = 1$ and $\sigma_\varepsilon^2 = \rho^2(1 + \theta_1^2 + \theta_2^2 + \theta_3^2)$.
(4) The result does not depend on the sample size $t = 100$ observations.

where $\hat{x}_t(l)$ is the MMSE predictor of $x_t(t + l)$ based on $\{x_t\}$ and $\hat{x}_t^*(l^*)$ is the MMSE predictor of $x_t^*(t + l^*)$ based on $\{x_t^*\}$. A similar relationship exists for $MMSE(\tilde{x}_t(l))$ and $MMSE(\tilde{x}_t^*(l^*))$, where $\tilde{x}_t(l)$ is the MMSE predictor of $y_t(t + l)$ based on $\{y_t\}$ and $\tilde{x}_t^*(l)$ is the MMSR predictor of $y_t^*(t + l^*)$ based on $y_t^* \equiv x_t^* + \varepsilon_t$, namely,

$$MMSE(\tilde{x}_t(l)) = MMSE(\tilde{x}_t^*(l^*)) \quad \text{for } 12(l^* - 1) < l \leqslant 12l^*, \qquad l^* = 1, 2, 3, \ldots \tag{21}$$

Therefore, in order to analyze the impact of measurement errors on MMSE predictions for ARMA(0,1) $\times$ SARMA$(1,0)_{12}$, one only needs to evaluate the corresponding statistics for the ARMA(1,1) process, $\{x_t^*\}$.

Example 6 Airline Model (The actual airline passenger model has an integrated component (see Box and Jenkins, 1976, p. 531.)) Suppose that x_t follows an ARMA(0,1) $\times$ SARMA$(0,1)_{12}$, that is, $x_t = a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-12-1}$. If $\sigma_a^2 = 1$, the autocovariance function for x_t is $\gamma_x(0) = (1 + \theta^2)(1 + \Theta^2)$, $\gamma_x(1) = \theta(1 + \Theta^2)$, $\gamma_x(k) = 0$ for $k = 2, 3, \ldots, 12 - 2$, $\gamma_x(12 - 1) = \theta\Theta$, $\gamma_x(12) = \Theta(1 + \theta^2)$, $\gamma_x(12 + 1) = \theta\Theta$, and $\gamma_x(l) = 0$ for $l \geqslant 12 + 2$.

Table IV. $\Delta MMSE$ and $\nabla MMSE$ of ARMA(1,1) process for several forecast horizons and different levels of measurement errors

		(ϕ_1, θ_2)							
		(0.8, −0.7) or (−0.8, 0.7)		(0.8, 0.7) or (−0.8, −0.7)		(0.9, −0.9) or (−0.9, 0.9)		(0.3, 0.5) or (−0.3, −0.5)	
ρ^2	l	$\Delta MMSE$	$\nabla MMSE$	$\Delta MMSE$	$\nabla MMSE$	$\Delta MMSE$	$\nabla MMSE$	$\Delta MMSE$	$\nabla MMSE$
0.0001	1	0.0031771	0.0031771	2.015e-06	2.015e-06	2.015e-06	0.0269167	5.567e-06	5.567e-06
	2	0.0020334	0.0006256	1.290e-06	1.277e-06	0.0218025	0.0051421	5.010e-07	4.818e-07
	5	0.0005330	8.595e-05	3.381e-07	3.304e-07	0.0115868	0.0010817	3.653e-10	3.499e-10
	10	5.723e-05	8.109e-06	3.630e-08	3.534e-08	0.0040401	0.0002608	2.157e-15	2.066e-15
0.01	1	0.2166930	0.2166930	0.0002000	0.0002000	0.6608152	0.6608119	0.0005494	0.0005494
	2	0.1386835	0.0426719	0.0001280	0.0001267	0.5352603	0.1262405	4.945e-05	4.755e-05
	5	0.0363551	0.0058624	3.355e-05	3.280e-05	0.2844593	0.0265551	3.605e-08	3.453e-08
	10	0.0039036	0.0005469	3.603e-06	3.507e-06	0.0991848	0.0064019	2.129e-13	2.039e-13
0.0625	1	0.7277709	0.7277709	0.0012022	0.0012022	1.7204460	1.7204370	0.0032118	0.0032118
	2	0.4657734	0.1433149	0.0007694	0.0007618	1.3935610	0.3286697	0.0002891	0.0002779
	5	0.1220997	0.0196890	0.0002017	0.0001971	0.7405955	0.0691368	2.107e-07	2.019e-07
	10	0.0131104	0.0018369	2.166e-05	2.108e-05	0.2582297	0.0166674	1.244e-12	1.192e-12

Notes:
(1) The signal is an ARMA(1,1) process $x_t - \phi_1 x_{t-1} = a_t - \theta_1 a_{t-1}$ with $\sigma_a^2 = 1$ and the observations follow $y_t = x_t + \varepsilon_2$.
(2)
$$\sigma_\varepsilon^2 = \rho^2 \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}.$$
(3) The result does not depend on the sample size t; results shown here are for $t = 100$ observations.

Let $x_t^* \equiv a_t - \theta a_{t-1} - \Theta a_{t-2} + \theta\Theta a_{t-3}$. Furthermore, let $\hat{x}_t(l)$ be the MMSE predictor of $x_t(t + l)$ based on $\{x_t\}$, and $\hat{x}_t^*(l^*)$ be the MMSE predictor of $x_t^*(t + l^*)$ based on $\{x_t^*\}$. Following a similar line of logic as in the previous example, we have

$$MMSE(\hat{x}_t(l)) = MMSE(\hat{x}_t^*(l^*)) \tag{22}$$

for $l^* = 1$ if $l = 1$; $l^* = 2$ if $l = 2, 3, \ldots, 12$; $l^* = 3$ if $l = 12 + 1$; and $l^* > 3$ if $l > 12 + 1$. Equation (22) holds for $MMSE(\tilde{x}_t(l))$ and $MMSE(\tilde{x}_t^*(l^*))$, where $\tilde{x}_t(l)$ is the MMSE predictor of $y_t(t + l)$ based on $\{y_t\}$ and $\tilde{x}_t^*(l)$ is the MMSE predictor of $y_t^*(t + l^*)$ based on $y_t^* \equiv x_t^* + \varepsilon_t$.

In examining the results, for instance, for the MA(3) model in Table III with $\theta_1 = -0.40$, $\theta_2 = -0.55$ and $\theta_3 = -0.22$ (equivalent to the $w_t = (1 + 0.4B)(1 + 0.55B^{12})a_t$ parameterization obtained by Box and Jenkins, 1976 for the airline data), we see that the impact of measurement errors regardless of the value of ρ is rather negligible. For equivalent seasonal processes with eigenvalues nearer the unit circle we observe the same type of behaviour noted earlier for non-seasonal processes, namely that their impact decreases as the forecast horizon increases. Note, for example, how dramatic are the changes in $\nabla MMSE$ for the parameterization $\theta_1 = 1.0$, $\theta_2 = 0.8$ and $\theta_3 = -0.9$ (equivalent to the MA(13) model with only three non-zero parameters θ_1, θ_{12},

Table V. $\nabla MMSE$ of ARMA(1,2) and ARMA(2,1) processes for several forecast horizons and different levels of measurement errors

		ARMA(1,2): $(\phi_1, \theta_1, \theta_2)$				ARMA(2,1): $(\phi_1, \phi_2, \theta_1)$			
ρ^2	l	(−0.8, 1.4, −0.6)	(0.6, −0.5, −0.9)	(0.3, −0.5, 0.3)	(−0.3, 0.3, −0.5)	(1.4, −0.6, −0.8)	(−0.5, −0.9, 0.6)	(0.3, −0.5, 0.3)	(−0.3, 0.3, −0.5)
0.001	1	0.0385551	0.0049626	0.0003949	0.0001271	0.0373196	0.0047054	3.701e-05	1.032e-05
	2	0.0125130	0.0036421	8.082e-07	8.818e-05	0.0113879	0.0004566	3.701e-05	6.634e-06
	5	0.0011102	6.436e-05	5.878e-10	4.681e-08	0.0007381	0.0002473	2.336e-06	2.406e-07
	10	9.853e-05	3.773e-07	3.471e-15	2.763e-13	8.269e-05	0.0001450	5.223e-08	1.047e-08
0.01	1	1.1520600	0.2282193	0.0334433	0.0123377	1.0559050	0.2883617	0.0036461	0.0010210
	2	0.2772410	0.1594559	7.032e-05	0.0084952	0.2756534	0.0322743	0.0036473	0.0006560
	5	0.0245971	0.0028177	5.115e-08	4.509e-06	0.0150383	0.0202891	0.0002308	2.377e-05
	10	0.0021831	1.654e-05	3.020e-13	2.662e-11	0.0018817	0.0106396	5.157e-06	1.034e-06
0.0625	1	2.8910510	0.6771251	0.13252727	0.0675198	2.8509200	0.8721216	0.0211792	0.0060423
	2	0.5636513	0.4227167	0.0003029	0.0448998	0.6317397	0.1399589	0.0212131	0.0038709
	5	0.0500077	0.0074697	2.203e-07	2.383e-05	0.0282102	0.0894723	0.0013586	0.0001398
	10	0.0044384	4.384e-05	1.300e-12	1.407e-10	0.0040122	0.0409981	3.025e-05	6.083e-06

Notes:
(1) The observations follow $y_t = x_t + \varepsilon_t$.
(2) For the ARMA(1,2) case, $x_t - \phi_1 x_{t-1} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ with $\sigma_a^2 = 1$ and

$$\sigma_\varepsilon^2 = \rho^2 \frac{1 + \theta_1^2 + \theta_2^2 - \theta_1\phi_1 - \theta_2\phi_1^2\theta_1\theta_2\phi_2}{1 - \phi_1^2}.$$

(3) For the ARMA(2,1) case $x_t - \phi_1 x_{t-1} = \phi_2 x_{t-2} = a_t - \theta_1 a_{t-1}$ with $\sigma_a^2 = 1$ and

$$\sigma_\varepsilon^2 = \rho^2 \frac{1 - \phi_2 + \theta_1^2 - \phi_2\theta_1^2}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)}.$$

(3) The result does not depend on the sample size t; results shown here are for $t = 100$ observations.

and θ_{13}) when the forecast horizon changes from one to two periods. As one would expect, the rate of decrease is smaller for large values of ρ.

CONCLUSIONS AND PATHS FOR FURTHER RESEARCH

In this paper we examined the impact that measurement errors have on prediction generated by ARMA models. Using MMSE and other measures of forecast accuracy we showed that the effect on forecasts generated from observed data *vis-à-vis* the true signal process is generally very small when the noise-to-signal ratio is small and the model parameters are well within the unit circle. The impact, however, can be quite large, particularly for the early forecast horizons, when the parameters are near the unit circle, and they increase in magnitude as the noise-to-signal ratio increases. In these cases it would behoove the model builder to formulate the model in state space form to extract the signal to generate forecasts.

We are currently investigating the possibility of extending our results to non-stationary models. This is mathematically complicated because Theorem 1 and its corollaries are based on the assumption that measurement errors ε_t are uncorrelated in time. If x_t, and hence, y_t are stationary after applying the differencing operator $\delta(B)$, the new error term $\delta(B)\varepsilon_t$ will be stationary but *not* uncorrelated. This makes closed forms of $\Delta MMSE$ and $\nabla MMSE$ for $\delta(B)x_t$ and $\delta(B)y_t$ very difficult, if not impossible, to derive because $R^l_{t,\delta(B)x_t}$ is not the same as $R^l_{t,\delta(B)y_t}$, and the difference between $\Gamma_{t,\delta(B)x_t}$ and $\Gamma_{t,\delta(B)y_t}$ is no longer a diagonal matrix as in equations (11) and (12). Numerical calculations of the mean squared errors for the original, non-stationary series, although tedious, can be made by first obtaining the MMSE's of the stationary series $\delta(B)x_t$ and then applying the filter $\delta^{-1}(B)$ to those MMSE's.

It is also interesting to note that extension to non-stationary structures is closely related to models having a stationary signal component and measurement errors that are correlated in time. For example, if the model $y_t = x_t + \varepsilon_t$ is non-stationary, and if taking first differences leads to the stationary model $(1-B)y_t = (1-B)x_t + (1-B)\varepsilon_t$, then $(1-B)y_t$ is a combination of a stationary ARMA process with an AR(1) measurement error term. Having more information on different error structures could provide model builders with an even more flexible framework for which to evaluate forecasts.

APPENDIX

Proof of Theorem 1: Applying Theorem 2.4.7 in Mathai and Provost (1992), we have

$$\mu_{min}\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} \leqslant \mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x} \leqslant \mu_{max}\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} \tag{A1}$$

where μ_{min} and μ_{max} are smallest and largest eigenvalues of $\Gamma_{t,x}\Gamma^{-1}_{t,y}$, respectively. Since $\Gamma_{t,x}\Gamma^{-1}_{t,y} = (\Gamma_{t,y} - \sigma^2_\varepsilon I_{t\times t})\Gamma^{-1}_{t,y} = I_{t\times t} - \sigma^2_\varepsilon\Gamma^{-1}_{t,y}$. Hence we have,

$$\mu_{min} = min_i\left(1 - \sigma^2_\varepsilon\frac{1}{\sigma^2_\varepsilon + \lambda_i}\right) \text{ and } \mu_{max} = max_i\left(1 - \sigma^2_\varepsilon\frac{1}{\sigma^2_\varepsilon + \lambda_i}\right)$$

where $\{\lambda_i\}$ are eigenvalues of $\Gamma_{t,x}$. It is easy to verify that

$$\mu_{min} = 1 - \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + min\ \lambda_i} \text{ and } \mu_{max} = 1 - \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + max\ \lambda_i} \tag{A2}$$

Replacing μ_{min} and μ_{max} in equation (A1) by (A2) we obtain

$$\left(1 - \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + min_i\lambda_i}\right)\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} \leqslant \mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,y}\mathcal{R}^l_{t,x} \leqslant \left(1 - \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + max_i\lambda_i}\right)\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x}$$

or

$$\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + max_i\lambda_i}\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x} \leqslant \Delta MMSE \leqslant \frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + min_i\lambda_i}\mathcal{R}^{l\,\prime}_{t,x}\Gamma^{-1}_{t,x}\mathcal{R}^l_{t,x}$$

If we let

$$B_{1,t} \equiv \gamma_x(0) - MMSE(\hat{x}(l)) \equiv \mathcal{R}^{l\,\prime}_{t,x}\gamma^{-1}_{t,x}\mathcal{R}^l_{t,x}$$

and

$$B_{2,t} \equiv [\gamma_x(0) - MMSE(\hat{x}(l))]/MMSE(\hat{x}(l))$$

we obtain the two inequalities for $\Delta MMSE$ and $\nabla MMSE$ of Theorem 1.

ACKNOWLEDGEMENTS

We are grateful to William R. Bell for useful comments on the early version of the paper.

REFERENCES

Bell, W. and Wilcox, D., 'The effect of sampling error on the time series behavior of consumption data', *Journal of Econometrics*, **55** (1993), 235–65.

Bell, W. and Hilmer, S., 'The time series approach to estimation for repeated surveys', *Survey Methodology*, **16** (1990), 195–215.

Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1976.

Brockwell, P. and Davis, R., *Time Series: Theory and Methods*, New York: Springer-Verlag, 1991.

Granger, C. W. J. and Morris, M. J., 'Time series modelling and interpretation', *Journal of the Royal Statistical Society, Series A*, **139** (1976), 246–57.

Gregory, R. and Karney, D., *A Collection of Matrices for Testing Computational Algorithms*, New York: Wiley-Interscience, 1969.

Grenander, U. and Szegö, G., *Toeplitz Forms and their Applications*, New York: Chelsea Publishing Company, 1958.

Kron, G. and Kron, T., *Mathematical Handbook for Scientists and Engineers*, New York: McGraw-Hill, 1961.

Mathai, A. M. and Provost, S. B., *Quadratic Forms in Random Variables*, New York: Marcel Dekker, 1992.

Pukkila, T., Koreisha, S. and Kallinen, A., 'The identification of ARMA models', *Biometrika*, **77** (1990), 537–48.

Wilcox, D., 'The construction of U.S. consumption data: some facts and their implications for empirical work', *The American Economic Review*, **82** (1992), 922–41.

Wong, W.-K. and Miller, R. B., 'Repeated time series analysis of ARIMA-noise models', *Journal of Business & Economic Statistics*, **8** (1990), 243–50.

Authors' biographies:

Sergio A. Koreisha is the Charles H. Lundquist Research Scholar and Professor of Decision Sciences at the Lundquist College of Business, University of Oregon. He specializes in the areas of forecasting, mathematical model building, applied econometrics, production analysis, and manufacturing strategy. Some of his latest works in the field of forecasting include articles dealing with new fast estimation methods for large-scale multiple-time series models; new approaches for identifying the mathematical structures governing the behavior of economic time series; and how causality among economic and business variables might be ascertained and used to formulate more accurate models. He has taught courses in Europe, South America, and the Pacific. He also provides consulting services for companies interested in doing business in Brazil.

Yue Fang is assistant professor of Decision Sciences at the Lundquist College of Business, University of Oregon. He has a Ph.D. in Statistics from MIT. His publications include articles on financial econometrics, time series forecasting and statistical quality controls. His recent work focuses on the quantitative analysis of financial markets using mathematical and statistical models.

Authors' address:
Sergio A. Koreisha and **Yue Fang**, Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA.

