

# Testing for measurement errors with discrete-time data sampled from a CARMA model

HENGHSIU TSAI<sup>†\*</sup>, KUNG-SIK CHAN<sup>‡</sup> AND PATRICK FAYARD<sup>†</sup>

We consider the problem of testing for measurement errors with discrete-time data sampled from a continuous-time autoregressive moving-average process. We develop an efficient algorithm for computing the likelihood ratio test (LRT) statistic, and derive the non-standard asymptotic null distribution of the LRT. The efficacy of the proposed test is illustrated by simulations and a real application from an environmental study.

KEYWORDS AND PHRASES: Forecasting, Gaussian distribution, Kalman filter, Likelihood ratio test, Non-standard asymptotics.

## 1. INTRODUCTION

Time-series data are often subject to two problems, namely, irregular spacing and corruption by measurement errors. The first problem may be circumvented by treating the data as sampled from some continuous-time process, often modeled by a continuous-time autoregressive moving-average (CARMA) model, see Section 2, and Tsai and Chan (2005a) and the references therein. However, to our knowledge, the problem of testing for (and incorporating) measurement errors in the continuous-time modeling with discrete-time data remains an open problem. On the other hand, ignoring measurement errors may result in incorrect inference, further lessening the capacity to detect significant effects. More importantly, forecasting may be less accurate when measurement errors are not incorporated in the model, as shown by the simulation results reported in Section 5.3.

Here, we consider the case of time-series data, possibly irregularly spaced, sampled from a CARMA process in the presence of additional measurement errors. The model is formulated in Section 2. Maximum likelihood estimation of the CARMA model with measurement errors can be efficiently carried out via Kalman filters, as detailed in Section 3. A fundamental problem concerns testing for measurement errors, which is equivalent to investigating whether the variance of the measurement error is zero. As the variance must

be non-negative, its true parameter value under the null hypothesis of no measurement error lies on the boundary of the parameter space. Consequently, the likelihood ratio test (LRT) has a non-standard null distribution. We derive in Section 4 the asymptotic null distribution of the LRT which is either 0 with 0.5 probability or is distributed as  $\chi^2$  with one degree of freedom. The empirical performance of the LRT is studied by simulations in Section 5, where we also study the impact of ignoring measurement errors on the forecasting performance. We illustrate the new approach with a real application in Section 6. Section 7 concludes. All proofs are collected in the Appendix.

## 2. CARMA MODEL WITH MEASUREMENT ERRORS

Let  $Y = \{Y_{t_j}\}_{j=0,1,2,\dots,N}$  be data sampled from a CARMA( $p, q$ ) process, where  $0 \leq q < p$ , and  $t_0 = 0$ . Moreover, suppose  $Y$  may be subject to serially independent additive measurement errors. Specifically, let

$$\begin{aligned} (1) \quad & \tilde{Y}_t^{(p)} - \alpha_p \tilde{Y}_t^{(p-1)} - \dots - \alpha_1 \tilde{Y}_t^{(0)} - \alpha_0 \\ & = \sigma \{W_t^{(1)} + \beta_1 W_t^{(2)} + \dots + \beta_q W_t^{(q+1)}\}, \\ (2) \quad & Y_{t_i} = \tilde{Y}_{t_i} + \epsilon_{t_i}, \quad i = 0, \dots, N, \end{aligned}$$

where the superscript  $(j)$  denotes  $j$ -fold differentiation with respect to  $t$ ;  $\{W_t\}$  is the standard Brownian motion;  $\alpha_0, \dots, \alpha_p, \beta_1, \dots, \beta_q$  and  $\sigma$  are unknown parameters. We assume that  $\sigma > 0$ ,  $\alpha_1 \neq 0$  and  $\beta_q \neq 0$ . The solution to equation (1) is assumed to be asymptotically stationary;  $\{\epsilon_{t_i}\}_{i=0}^N$  is a sequence of i.i.d. normal random variables with mean zero and variance  $\nu\sigma^2$ , and is independent of  $\{W_t, t \geq 0\}$ . Our objective here is to test  $H_0 : \nu = 0$  against  $H_1 : \nu > 0$ , via the likelihood ratio test.

Similar to Brockwell (1993), Eqns. (1) and (2) can equivalently be cast in terms of the *observation* and *state* equations,

$$\begin{aligned} (3) \quad & Y_{t_i} = \beta' X_{t_i} + \epsilon_{t_i}, \quad i = 0, \dots, N, \\ (4) \quad & dX_t = (AX_t + \alpha_0 \delta_p) dt + \sigma \delta_p dW_t, \end{aligned}$$

\*Corresponding author.

<sup>†</sup>We wish to thank Academia Sinica, the National Science Council (NSC 98-2118-M-001-023-MY2 and SIT scholarship), R.O.C.,

<sup>‡</sup>and the US National Science Foundation (DMS-1021292) for partial support.

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p \end{bmatrix},$$

$$X_t = \begin{bmatrix} X_t^{(0)} \\ X_t^{(1)} \\ \vdots \\ X_t^{(p-2)} \\ X_t^{(p-1)} \end{bmatrix}, \quad \delta_p = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 \\ \beta_1 \\ \vdots \\ \beta_{p-2} \\ \beta_{p-1} \end{bmatrix},$$

where we define  $\beta_j := 0$  for  $j > q$ , and the superscript prime denotes transpose. For  $(p, q) = (1, 0)$  and  $(p, q) = (2, 1)$  without measurement error, the equivalence of (1)/(2) and (3)/(4) was discussed in Tsai and Chan (2005b, pp. 585–586). Equation (4) is an Ito differential equation for the state vector  $X_t$ . We assume that  $X_0$  is independent of both  $\{W_t, t \geq 0\}$  and  $\{\epsilon_{t_i}\}_{i=0}^N$ , and that  $X_0$  is determined by initial conditions that could be random or deterministic. Let  $\alpha(z) = z^p - \alpha_p z^{p-1} - \cdots - \alpha_1$ , and  $\beta(z) = 1 + \beta_1 z + \beta_2 z^2 + \cdots + \beta_q z^q$ . For model identification, it is assumed that these two polynomials share no common roots. Furthermore, we assume all roots of  $\alpha(z) = 0$  and those of  $\beta(z) = 0$  have negative real parts. The condition on the roots of  $\alpha(z) = 0$  is necessary for the asymptotic stationarity of the process whereas that on  $\beta(z) = 0$  is akin to the invertibility condition for discrete-time processes. The preceding conditions ensures that the underlying continuous-time process is identifiable. However, these conditions may not be sufficient for identifying the model with discrete-time data obtained from regularly sampling the latent continuous-time process, plus measurement errors. Indeed, even without measurement error, a CARMA model may not be identifiable with regularly spaced discrete-time observations (i.e.  $t_i - t_{i-1} \equiv h > 0, \forall i$ ), see Pandit and Wu (1975, 1983), Hansen and Sargent (1983) and McCrorie (2003). However, Pandit and Wu (1975, 1983) showed that any stationary CAR(2) with non-zero  $\alpha_2$  is identifiable with regular discrete-time observations, for any sampling interval  $h$ . Moreover, they showed that a stationary CAR( $p$ ) model is identifiable with regular discrete-time data, if  $h$  is sufficiently small. While the results of Pandit and Wu (1975) assume no measurement errors, these results are valid for the case of sampling a stationary CAR model over regular epochs, plus IID measurement errors, because, based on the discrete-time observations, the variance of the measurement errors and the discrete-time auto-regressive moving average model, implied by the continuous-time model under the case of no measurement errors, are identifiable. The general problem of the identifiability of a CARMA model with discrete-time observations is, however, still open. Henceforth, we shall assume that the underlying continuous-time model is identifiable with regular discrete-time observations,

plus measurement errors. These regularity conditions on the stationarity and invertibility of the CARMA model, as well as model identifiability with discrete-time observations, will be called **Condition (C)** below.

**Lemma 2.1.** *If all roots of  $\alpha(z) = 0$  and those of  $\beta(z) = 0$  have negative real parts, then  $\alpha_j < 0$ , for  $j = 1, \dots, p$ , and  $\beta_j > 0$ , for  $j = 1, \dots, q$ .*

The states and the observations,  $X_{t_i}$  and  $Y_{t_i}$ , at the sampling times  $t_0, t_1, \dots, t_N$ , satisfy the discrete-time state and observation equations:

$$(5) \quad X_{t_{i+1}} = \mu + e^{A_{i+1}}(X_{t_i} - \mu) + Z_{t_i}, \quad i = 0, 1, \dots, N,$$

$$(6) \quad Y_{t_i} = \beta' X_{t_i} + \epsilon_{t_i}, \quad i = 0, 1, \dots, N,$$

where  $A_{i+1} = A(t_{i+1} - t_i)$ ,  $\mu = (-\alpha_0/\alpha_1, 0, \dots, 0)'$ ,  $Z_{t_i}$  is independent of  $X_{t_i}$  and  $\{Z_{t_i}, i = 0, 1, \dots, N\}$  is an independent sequence of Gaussian random vectors with zero mean and covariance matrices

$$(7) \quad \begin{aligned} \Sigma_i &= E(Z_{t_i} Z_{t_i}') = \sigma^2 \int_{t_i}^{t_{i+1}} e^{A(t_{i+1}-u)} \delta_p \delta_p' e^{A'(t_{i+1}-u)} du \\ &= V - e^{A(t_{i+1}-t_i)} V e^{A'(t_{i+1}-t_i)}, \end{aligned}$$

where  $V = \sigma^2 \int_0^\infty e^{Au} \delta_p \delta_p' e^{A'u} du$  is the asymptotic stationary variance of  $\{X_t\}$ . Refer to Tsai and Chan (2000) for an efficient method to compute  $V$ .

Define  $\hat{X}_{t|s}$  as the conditional expectation of  $X_t$  based on the observations up to time  $s$  and  $P_{t|s}$  the corresponding covariance matrix, i.e.,

$$\begin{aligned} \hat{X}_{t|s} &= E(X_t | y_j, j \leq s) = (\hat{X}_{t|s}^{(0)}, \hat{X}_{t|s}^{(1)}, \dots, \hat{X}_{t|s}^{(p-1)})', \\ P_{t|s} &= \text{Var}(X_t | y_j, j \leq s). \end{aligned}$$

### 3. MAXIMUM LIKELIHOOD ESTIMATION

Let the parameter space, for which **Condition (C)** holds, be denoted by  $\Omega$  and an arbitrary element of  $\Omega$  be denoted by  $\theta = (\nu, \psi)$ , where  $\nu \geq 0$  is the parameter of interest and  $\psi = (\alpha_0, \dots, \alpha_p, \beta_1, \dots, \beta_q)$  are nuisance parameters. Using equations (5) and (6), the log-likelihood function of the observed data,  $l_Y(\nu, \psi, \sigma^2)$ , can be computed via Kalman filters (see, e.g. Jones, 1980, 1985 and Tong, 1990), which we briefly outline below. First, start with a diffuse initial condition as we do not assume stationarity, i.e., let

$$\begin{aligned} \hat{X}_{t_{-1}|t_{-1}} &= [\bar{y}, 0, \dots, 0]', \\ P_{t_{-1}|t_{-1}} &= \delta s_y^2 I, \end{aligned}$$

where  $t_{-1}$  can be taken as a time point such that  $(t_0 - t_{-1})$  equals the average of  $\{t_j - t_{j-1}, j = 1, \dots, N\}$ ,  $\delta$  is some positive number,  $I$  is the identity matrix,  $\bar{y}$  and  $s_y^2$  are the sample mean and sample variance of the observed data, respectively. A reasonable choice of  $\delta$  is, e.g., 5.

The log likelihood function can now be computed recursively as follows. For  $i = 0$  to  $i = N$ :

- (i) Calculate the one-step ahead predictor

$$\hat{X}_{t_i|t_{i-1}} = \mu + e^{A_i}(\hat{X}_{t_{i-1}|t_{i-1}} - \mu).$$

- (ii) Calculate the covariance matrix of the preceding predictor

$$P_{t_i|t_{i-1}} = e^{A_i} P_{t_{i-1}|t_{i-1}} e^{A_i'} + V - e^{A_i} V e^{A_i'}.$$

- (iii) The prediction of the next observation is

$$\hat{y}_{t_i|t_{i-1}} = \beta' \hat{X}_{t_i|t_{i-1}}.$$

- (iv) The innovation,  $\tilde{y}_{t_i}$ , is the predictive residual, i.e.,

$$\tilde{y}_{t_i} = y_{t_i} - \hat{y}_{t_i|t_{i-1}}.$$

- (v) The innovation covariance matrix is

$$\Lambda_{t_i} = \beta' P_{t_i|t_{i-1}} \beta + \nu \sigma^2.$$

- (vi) The Kalman gain matrix is

$$\Delta_{t_i} = P_{t_i|t_{i-1}} \beta \Lambda_{t_i}^{-1}.$$

- (vii) Update the estimate of the state vector to

$$\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + \Delta_{t_i} \tilde{y}_{t_i},$$

with the corresponding covariance matrix being

$$P_{t_i|t_i} = P_{t_i|t_{i-1}} - \Delta_{t_i} \beta' P_{t_i|t_{i-1}}.$$

These steps are cycled through to yield minus twice the log-likelihood which is equal to

$$(8) \quad -2l_Y(\nu, \psi, \sigma^2) = \sum_{i=0}^N \left( \frac{\tilde{y}_{t_i}^2}{\Lambda_{t_i}} + \log \Lambda_{t_i} \right) + \text{constant}.$$

By the same token, we can also obtain a closed form expression for the maximum likelihood estimator of the parameter  $\sigma^2$ . This can be done by calculus, and the recursion defined by (i) – (vii) takes a slightly different form (see, e.g., Jones, 1980). We only need to replace  $P_{t|s}$  by  $Q_{t|s} = P_{t|s}/\sigma^2$  and  $\Lambda_{t_i}$  by  $\Lambda_{t_i}^* = \Lambda_{t_i}/\sigma^2$  in (i) to (vii), and equation (8) becomes

$$(9) \quad -2l_Y(\nu, \psi, \sigma^2) = \sum_{i=0}^N \left\{ \frac{\tilde{y}_{t_i}^2}{\sigma^2 \Lambda_{t_i}^*} + \log(\sigma^2 \Lambda_{t_i}^*) \right\} + \text{constant}.$$

Differentiating (9) with respect to  $\sigma^2$  and equating to zero gives

$$(10) \quad \sigma^2 = \frac{1}{N+1} \sum_{i=0}^N \frac{\tilde{y}_{t_i}^2}{\Lambda_{t_i}^*},$$

and substituting into (9), the objective function becomes

$$(11) \quad -2l_Y(\nu, \psi) = (N+1) \log \left( \sum_{i=0}^N \frac{\tilde{y}_{t_i}^2}{\Lambda_{t_i}^*} \right) + \sum_{i=0}^N \log \Lambda_{t_i}^* + \text{constant}.$$

Under the alternative hypothesis  $H_1 : \nu > 0$ , the function (11) is minimized with respect to  $\nu$  and  $\psi$  to get the maximum likelihood estimates  $\hat{\nu}$  and  $\hat{\psi}$ . The parameter estimate  $\hat{\sigma}^2$  is then calculated from (10). Under the null hypothesis  $H_0 : \nu = 0$ , the parameter  $\nu$  is restricted to be zero, and the objective function (11) is minimized with respect to  $\psi$  to get the maximum likelihood estimate  $\hat{\psi}_0$ .

## 4. THE LIKELIHOOD RATIO TEST STATISTIC

We are interested in testing the following hypothesis:

$$\begin{cases} H_0 : \nu = 0, \\ H_1 : \nu > 0. \end{cases}$$

Recall that  $\hat{\psi}_0$  is the maximum likelihood estimate of  $\psi$  under  $H_0$ , and  $\hat{\nu}$  and  $\hat{\psi}$  are the maximum likelihood estimates of  $\nu$  and  $\psi$  under  $H_1$ . (Note for real applications,  $\hat{\nu}$  might be zero.) Thereafter, depending on the  $-2$  times the log-likelihood ratio statistic given by  $\ell = 2 \max\{l_Y(\hat{\nu}, \hat{\psi}) - l_Y(0, \hat{\psi}_0), 0\}$ , where  $-2l_Y$  is defined by equation (11), we reject  $H_0$  if  $\ell$  is too large. The following main result derives the asymptotic null distribution of the likelihood ratio test.

**Theorem 4.1.** *Assume Condition (C) holds and that the data are regularly spaced. Let  $W$  be the matrix such that its  $(i, j)$ -th element equals  $\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log g}{\partial \theta_i} \frac{\partial \log g}{\partial \theta_j} d\omega$ , with  $\theta_i$  being the  $i$ -th component of  $\theta$  and the integrand evaluated at the true value  $\theta_0$ . Assume that  $W$  is positive definite, under  $H_0$ . Then, the asymptotic null distribution of  $\ell$  is that of a chance variable which is zero half the time and which behaves like  $\chi^2$  with one degree of freedom the other half of the time.*

For CAR( $p$ ) models, if the roots of  $\alpha(z) = 0$  are all distinct, positive definiteness of  $W$  can be seen as follows. First, regular sampling from a CARMA( $p, q$ ) process with  $q < p$  generally yields an ARMA( $p, p-1$ ) process; see Pandit and Wu (1983, Chapter 6). Second, if the roots of the characteristic equation for the CAR( $p$ ) model are all distinct and assuming the CAR( $p$ ) model is identifiable based on the regularly sampled discrete-time data, then the CAR parameter vector is a differentiable, invertible function of the parameter vector of the ARMA model for the discrete-time data sampled regularly from the CAR model. Invertibility follows from the assumption that the CAR( $p$ ) model is identifiable with regularly sampled data. The preceding claim on differentiability follows from two facts: (i) the roots, denoted by  $\lambda_i$ , of the AR characteristic equation of the ARMA model

relate to the roots, denoted by  $\mu_i$ , of the AR characteristic root of the CAR model by the simple transformation  $\lambda_i = \exp(\mu_i h)$ , where  $h$  is the sampling interval, (ii) the AR (CAR) coefficients of the ARMA (CAR) model are elementary symmetric functions of the roots of the AR (CAR) characteristic equation. Moreover, specific formulas for the MA coefficients of the ARMA model given by Pandit and Wu (1975 and 1983) shows that the transformation from the CAR coefficients to the MA coefficients of the ARMA model is a differentiable function. Hence, the Jacobian matrix, denoted by  $J$ , for the transformation linking the CAR( $p$ ) parameters to the parameters of the ARMA model is of full-rank. Denote by  $W_d$  the matrix that is similar to  $W$  except that the CAR spectrum  $g$  is replaced by that of the corresponding ARMA process and  $\theta$  replaced by the vector of ARMA parameters. It then follows that  $W = JW_d J^T$ , which is positive definite because of the positive-definiteness of  $W_d$ ; see Hannan (1970, Theorem 8 on p. 392). For the CARMA case, the relationship between the CARMA parameters with the parameters of the corresponding ARMA model for the regularly sampled data is quite complex, and it is an interesting future problem to extend the preceding result concerning the positive-definiteness of  $W$  to the CARMA case.

Even though Theorem 4.1 holds for equally spaced data, we conjecture that it also holds for irregularly spaced data under some regularity conditions, e.g. when the spacing is uniformly bounded away from 0 and infinity. Also, the asymptotic distribution of the maximum likelihood estimator under the alternative is known to be asymptotically normal, under mild regularity conditions; see Hannan (1973).

## 5. SIMULATION

In this section we use simulation examples to study the performance of the Likelihood Ratio test statistic in finite samples.

### 5.1 Empirical sizes of the likelihood ratio test statistic

In this subsection we use Monte Carlo methods to compute the empirical rejection frequencies of the Likelihood Ratio test statistic with nominal values equal to 0.10, 0.05 and 0.01, respectively. Three stationary CARMA( $p, q$ ) processes are considered:

$$\begin{aligned} \text{Model I,} \quad & \tilde{Y}_t^{(1)} + 0.25\tilde{Y}_t^{(0)} = W_t^{(1)}, \\ \text{Model II,} \quad & \tilde{Y}_t^{(2)} + 0.20\tilde{Y}_t^{(1)} + 0.3\tilde{Y}_t^{(0)} = W_t^{(1)}, \\ \text{Model III,} \quad & \tilde{Y}_t^{(2)} + 0.20\tilde{Y}_t^{(1)} + 0.3\tilde{Y}_t^{(0)} = W_t^{(1)} + 0.5W_t^{(2)}. \end{aligned}$$

The observations  $\{Y_{t_i}\}_{i=0, \dots, N}$  are generated via equations (5) and (6) without measurement errors. We consider both regularly spaced and irregularly spaced data. For regularly spaced data, we set  $t_i = i$ , for all integer  $i$ , whereas for irregularly spaced time series data, we follow the following

Table 1. Empirical frequencies of rejecting the null hypothesis of a CARMA( $p, q$ ) model without measurement errors when 0.01, 0.05, and 0.10 asymptotic critical values are used. The results are based on 1,000 replications

N	regularly spaced			irregularly spaced		
	Nominal	Level		Nominal	Level	
I	0.100	0.050	0.010	0.100	0.050	0.010
100	0.078	0.041	0.005	0.061	0.029	0.001
200	0.076	0.036	0.006	0.077	0.039	0.006
400	0.080	0.041	0.005	0.084	0.038	0.005
II	0.100	0.050	0.010	0.100	0.050	0.010
100	0.131	0.074	0.012	0.115	0.053	0.016
200	0.121	0.057	0.007	0.110	0.056	0.009
400	0.111	0.053	0.010	0.128	0.069	0.017
III	0.100	0.050	0.010	0.100	0.050	0.010
100	0.113	0.080	0.043	0.173	0.134	0.090
200	0.114	0.085	0.072	0.148	0.114	0.092
400	0.125	0.085	0.058	0.168	0.119	0.084

simulation scheme. First, simulate  $s_i, i = 1, \dots, N$ , independently from the exponential distribution with mean equals 0.5. Then, set  $t_0 = 0$ , and  $t_i = t_{i-1} + s_i + 0.5, i = 1, \dots, N$ . Second, irregularly spaced time series data are simulated at these particular time values. The sample sizes considered are  $N = 100, 200$ , and 400.

From Table 1, we can observe that, in general, the empirical sizes are close to their corresponding nominal sizes, except for Model III, especially for the case of irregularly spaced data.

### 5.2 Empirical power of the test

In this subsection, we consider the empirical power of the test when the true models are subject to measurement errors. The true model under investigation is Model II in subsection 5.1 with additional measurement errors. Specifically, the values of  $\nu$  are chosen to be  $\nu = 0.00, 0.01, 0.05, 0.09$  and 9.00. We define the signal-to-noise ratio (SNR) to be the ratio of the asymptotic stationary variance of  $\{\tilde{Y}_{t_j}\}$  to the variance of  $\{\epsilon_{t_j}\}$  if the measurement noise's variance is non-zero, or infinity otherwise. Note that for this model with non-zero  $\nu$ , the SNR is  $(\beta'V\beta)(\nu\sigma^2)^{-1} = (2\alpha_1\alpha_2\nu)^{-1}$ , by Theorem 2.1 (b) of Tsai and Chan (2000). Therefore, the corresponding SNR's are  $\infty, 833.3, 166.7, 92.6$ , and 0.93, respectively. Only regularly spaced data with  $t_i = i$  are considered.

From Table 2, it is clear that, for a fixed SNR, the empirical power increases as the sample size increases. For fixed sample size  $N$ , the empirical power is largest when  $\nu = 0.09$ , corresponding to SNR = 92.6.

### 5.3 The impact of the measurement errors in forecasting

In this subsection, we study empirically the impact of ignoring measurement errors on forecasting. Consider the



Table 2. The empirical frequencies of rejecting the null hypothesis for different specifications of the value of  $\nu$ . The data are generated from Model II in subsection 5.1 with measurement errors corresponding to  $\nu = 0.00, 0.01, 0.05, 0.09, \text{ and } 9.00$ . The results are based on 1,000 replications and critical values of the nominal size 0.05

$\nu$	0.00	0.01	0.05	0.09	9.00	
(SNR)	( $\infty$ )	(833.3)	(166.7)	(92.6)	(0.93)	
N	100	0.074	0.276	0.857	0.963	0.847
	200	0.057	0.359	0.981	0.998	0.990
	400	0.053	0.586	1.000	1.000	1.000

true model to be again Model II of Section 5.1 with measurement errors corresponding to the  $\nu$  values considered in Subsection 5.2. Again, only regularly spaced data with  $t_i = i$  and sample size  $N = 100, 200, \text{ and } 400$  are considered. We are interested to see how the root mean squares of forecast errors differ if we do not consider the measurement errors in modelling. Specifically, we simulate  $\{Y_i\}_{i=0, \dots, N+m}$  from the true model. Then we estimate the continuous-time AR(2) models with and without measurement errors based on  $\{Y_i\}_{i=0, \dots, N}$ , assuming the order  $p = 2$  is known whereas the parameters  $\nu$ , and the  $\alpha$ 's are unknown, and have to be estimated from the data. Next, we compute the innovations,  $\tilde{y}_i$ , for  $i = N + 1, \dots, N + m$ , which is defined in step (iv) of Section 3. Finally, we compute the root mean squared forecast error:

$$(12) \quad \text{RMSFE} = \left( \frac{1}{m} \sum_{N+1}^{N+m} \tilde{y}_i^2 \right)^{1/2}.$$

In this subsection, the value of  $m$  is always 10. For each model and each sample size, we replicate the experiment 1,000 times, with the averages of the RMSFE's reported in Table 3. From the table, we see that for  $\nu = 0.00$  and 0.01, there are essentially no differences between using the continuous-time ARMA model with and without measurement errors, whereas for  $\nu = 0.05, 0.09, \text{ and } 9.00$ , fitting the model with measurement errors results in smaller RMSFE values.

## 6. APPLICATION

**Example:** Field values of pH of wet deposition at the McNay Research Station in the Lucas County of Iowa, U.S., have been collected on a more or less weekly basis since 1984. The McNay Research Station is one of the two monitoring sites of the Iowa Precipitation Monitoring Program for the National Trends Network. The monitoring program aims at providing an overview of chemical composition of atmospheric deposition in the U.S., see <http://ia.water.usgs.gov/projects/ia005.html>. Besides field values of pH, specific conductance and chemical analysis of the precipitation were recorded. Here, we focus

Table 3. Comparison of Forecasting performances of the continuous-time ARMA models with measurement errors. The data are generated from Model II in subsection 5.1. The results are based on averages of 1,000 replications, the value of  $m$  is always 10, and the sample sizes used are 100, 200 and, 400

$\nu$	0.00	0.01	0.05	0.09	9.00
(SNR)	( $\infty$ )	(833.3)	(166.7)	(92.6)	(0.93)
N	$H_0$	$H_0$	$H_0$	$H_0$	$H_0$
100	0.680	0.721	0.857	0.963	3.735
200	0.683	0.723	0.862	0.972	3.726
400	0.672	0.713	0.850	0.958	3.681
N	$H_1$	$H_1$	$H_1$	$H_1$	$H_1$
100	0.681	0.721	0.833	0.913	3.560
200	0.684	0.723	0.837	0.919	3.613
400	0.672	0.711	0.821	0.900	3.531

on the pH measurements which measured the acidity of the wet deposition; there are 562 observations, collected from October 1, 1985 to September 18, 2001. For simplicity, we round the sampling times to days. The data are irregularly spaced with the sampling intervals between consecutive observations ranging from 1 to 98 days; the average and median sampling intervals are 10 and 7 days, respectively. The unit of time is taken as one day. The same data was analyzed in Tsai and Chan (2005a). Here, we are interested in analyzing the log-transformed data, the time series plot of which is displayed in Fig. 1.

Specifically, we consider testing a continuous-time AR( $p$ ) model without measurement errors versus a continuous-time AR( $p$ ) model with measurement errors with  $p = 1, 2, \text{ and } 3$ . Here, the diffuse initial condition for the Kalman filter is  $P_{t-1|t-1} = 5s_y^2 I$ . For each model, the Akaike information criterion (AIC) is reported, which is defined as  $\text{AIC} = -2(l_Y(\hat{\theta}) - r)$ , where  $-2l_Y$  is defined in Equation (11) without the constant term,  $r$  is the number of parameters in the model, and  $\hat{\theta}$  is the maximum likelihood estimate of  $\theta$ . See Table 4 for the values of  $\hat{\nu}$ , the  $p$ -values of the LRT and the AIC's.

From Table 4, we see that, the CAR(1) plus measurement error model has the smallest AIC value, for  $p = 1, \text{ and } 2$ , the  $p$ -values both are equal to 0.00, resulting in rejection of the null hypothesis of no measurement error, at 5% significance level. Furthermore, because  $\hat{\nu} = 0.00$  for  $p = 3$ , the CAR(3) models are the same under  $H_0$  and  $H_1$ .

Because the AIC's are close for  $p = 1, 2, \text{ and } 3$  under  $H_0$ , and close for  $p = 1 \text{ and } 2$  under  $H_1$ , we compare these five models by examining their corresponding model fits by comparing the sample autocorrelation function (ACF) of the standardized residuals of these models in Figure 2. It is clear that the CAR models without measurement errors show similar ACF's for  $p = 1, 2, \text{ and } 3$ . Similarly, the CAR models with measurement errors show similar ACF's for  $p = 1, \text{ and}$

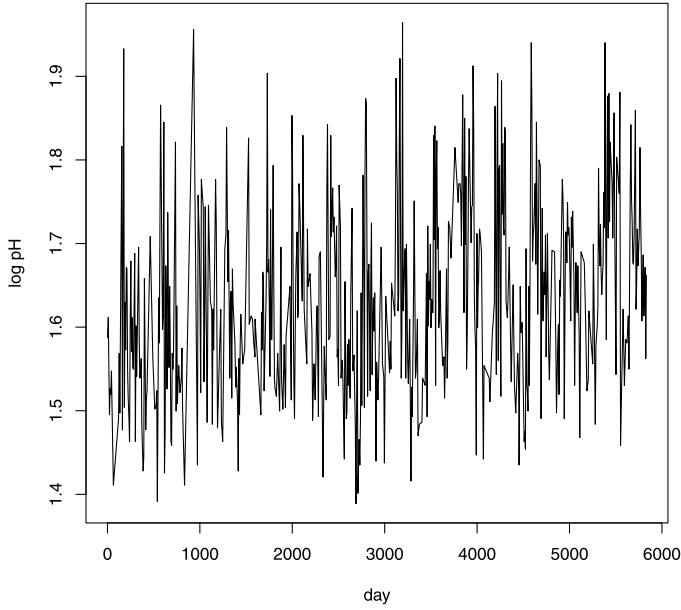


Figure 1. Time series plot of the log-transformed field values of pH of wet deposition at the McNay Research Station.

Table 4. The  $p$ -values of likelihood ratio test, the values of  $\hat{\nu}$ , and the AIC's for various CAR( $p$ ) models fitted to the McNay Data

$p$	$\hat{\nu}$	$p$ -value	reject $H_0$	SNR under $H_1$	AIC under $H_0$	AIC under $H_1$
1	38.5	0.000	yes	0.70	1042.2	1016.9
2	1.29	0.000	yes	0.66	1045.4	1018.9
3	0.00	0.500	no	$\infty$	1043.9	1045.9

2. Moreover, the ACF's of the standardized residuals show that the models with measurement errors fit the data better.

Finally, we reserved the last 10 data cases for validating the model predictive performance as follows. To the reduced data without the last 10 data we refitted the three CAR models with/without measurement errors, and then computed their RMSEF's defined in Equation (12). The results are summarized in Table 5. From the table, we see that for  $p = 1$  and 2, the model with measurement error has smaller RMSFE. Overall, the continuous-time AR(1) with measurement error model has the smallest RMSFE. The parameter estimates for this model are  $(\hat{\nu}, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\sigma}^2) = (38.5, 3.01 \times 10^{-2}, -1.85 \times 10^{-2}, 1.97 \times 10^{-4})$ . We then assess the uncertainty of the estimators by computing the observed Fisher information based on equation (11), based on which the asymptotic standard errors turn out to be  $(15.1, 8.87 \times 10^{-3}, 5.45 \times 10^{-3})$  for the estimates of  $(\nu, \alpha_0, \alpha_1)$ , so the estimates are all significant at a 5% level.

## 7. CONCLUSION

We have shown that the LRT provides a powerful approach for checking the presence of measurement error with

Table 5. Comparison of the RMSFE Values for the log-transformed Field values Data, for  $p = 1, 2$ , and 3

$p$	1	2	3
without measurement errors	0.0712	0.0715	0.0708
with measurement errors	0.0657	0.0685	

data sampled from a CARMA process. However, nonlinear time series data abound, and it is an interesting future research problem to study nonlinear continuous-time processes with measurement errors, e.g. threshold type nonlinear processes (Tong, 1990) with measurement errors.

## APPENDIX A. APPENDIX SECTION

*Proof of Lemma 2.1.* The cases of  $p = 1$  and  $p = 2$  for  $\alpha(z)$  can be checked by algebra. The proof is completed on noting that the characteristic polynomial  $z^p - \alpha_p z^{p-1} - \dots - \alpha_1$  can be factorized into products of real polynomials of a degree not greater than two, all of which have positive coefficients based on the arguments presented for orders one and two. The proof for  $\beta(z)$  is similar and is hence omitted.  $\square$

*Proof of Theorem 4.1.* Let  $i = \sqrt{-1}$ . Without loss of generality, we shall restrict the proof to the case of zero mean so that  $\alpha_0$  will be dropped from the parameter. For regularly spaced  $\{Y_t, t = 0, 1, 2, \dots\}$  driven by (5) and (6), it essentially follows from Eqns. (3.2) and (8.1) on p. 45 and p. 405 of Hannan (1970), respectively, that its spectral density function is given by  $\sigma^2 g(\omega; \theta)$ ,  $-\pi \leq \omega < \pi$ , where

$$g(\omega; \theta) = \frac{1}{2\pi} \left\{ \nu + \sum_{j=-\infty}^{\infty} \frac{|\beta(i\omega + 2\pi i j)|^2}{|\alpha(i\omega + 2\pi i j)|^2} \right\}.$$

(Recall  $\alpha(z) = z^p - \alpha_p z^{p-1} - \dots - \alpha_1$ , and  $\beta(z) = 1 + \beta_1 z + \beta_2 z^2 + \dots + \beta_q z^q$ .) For simplicity, we first profile out  $\sigma^2$  from the likelihood function. It is well-known (Whittle, 1951, and Theorems 1–3 of Hannan, 1973) that for Gaussian time series, the normalized profile likelihood is asymptotically equal to the following profile Whittle likelihood:

$$(13) \quad \frac{\ell_Y(\theta)}{N} = -\frac{1}{2} \log(U_N(Y; \theta)/N) + \text{const},$$

where const is an expression that does not depend on the parameter, and

$$U_N(Y; \theta) = \int_{-\pi}^{\pi} I_Y(\omega) g^{-1}(\omega; \theta) d\omega,$$

where  $I_Y(\omega)$  is the periodogram. Theorem 2 of Hannan (1973) implies that the profile Whittle likelihood (and the profile Gaussian likelihood) has the following second-order Taylor expansion:

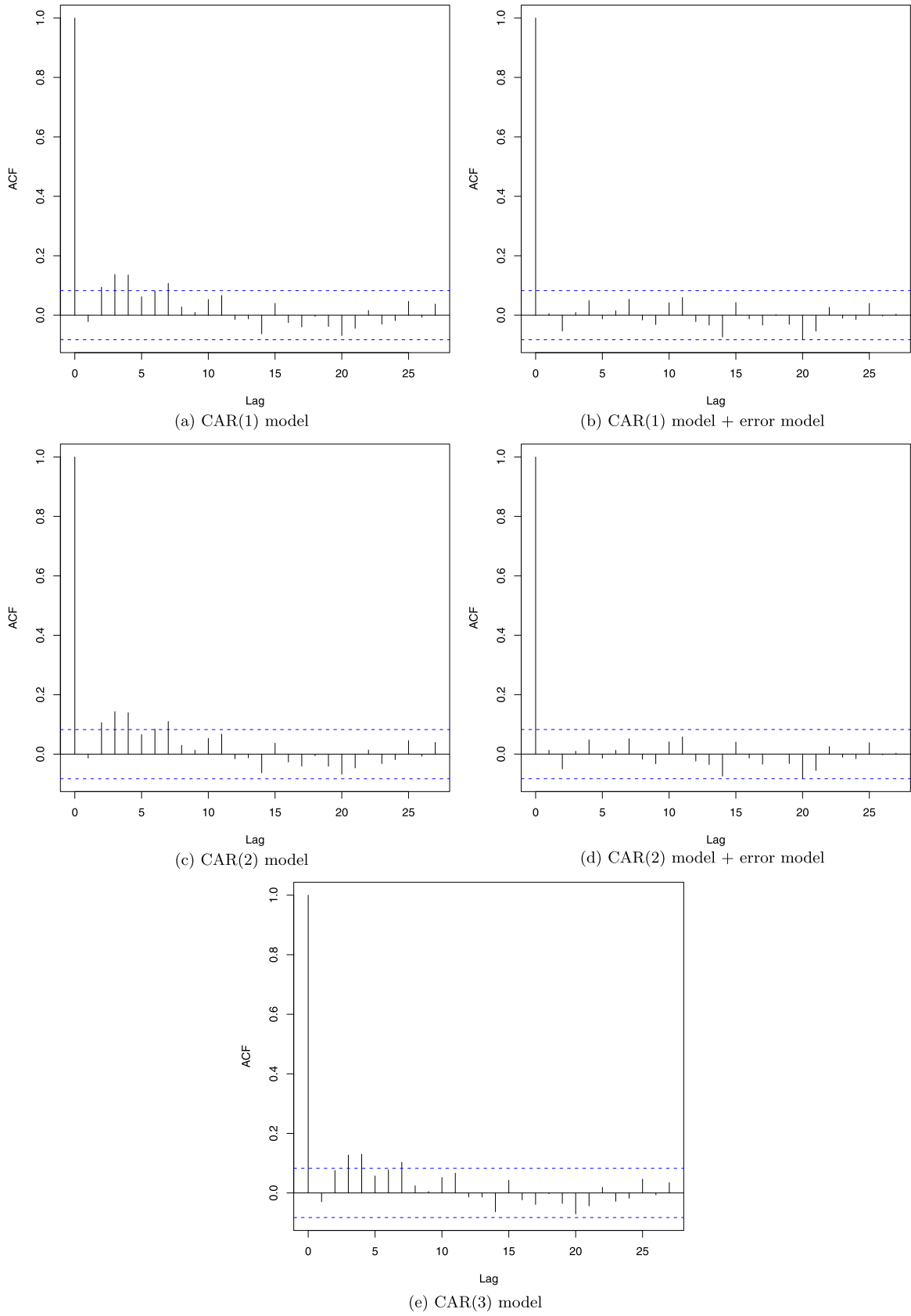


Figure 2. ACF of the standardized residuals for the various  $CAR(p)$  models fitted to the McNay data.

$$(14) \quad 2 \frac{\ell_Y(\theta)}{N} = 2 \frac{\ell_Y(\theta_0)}{N} - \frac{\dot{U}_N(Y; \theta_0)/N}{U_N(Y; \theta_0)/N} (\theta - \theta_0) \\ - (\theta - \theta_0)^T \frac{\ddot{U}_N(Y; \theta_0)/N}{2U_N^2(Y; \theta_0)/N^2} (\theta - \theta_0) \\ + o(|\theta - \theta_0|^2),$$

where  $\dot{U}_N$  and  $\ddot{U}_N$  are the first and second partial derivatives of  $U_N$  w.r.t.  $\theta$ . (Note the conditions A, B and (12) required by Theorems 1–3 of Hannan (1973) are satisfied for our model.) By the Law of Large Numbers,  $U_N(Y; \theta_0)/N \rightarrow 1$  almost surely as  $N \rightarrow \infty$ . Also, it is shown in the proof of Theorem 2 of Hannan (1973) that  $N^{-1/2}\dot{U}_N(Y; \theta_0)$  is asymptotically normally distributed with zero mean and variance-covariance matrix equal to  $2\sigma_0^2 W$  where its  $(i, j)$ -th element equals  $\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \log g}{\partial \theta_i} \frac{\partial \log g}{\partial \theta_j} d\omega$ , with  $\theta_i$  being the  $i$ -th component of  $\theta$  and the integrand is evaluated at the true value  $\theta_0$ . By ergodicity,  $\ddot{U}_N(Y; \theta_0)/N \rightarrow W$ , almost surely. The result stated in Theorem 4.1 follows from Eqn. (4) of Chernoff (1954), where the term  $|\theta|^3 |O_p(1)|$  can be replaced by  $o_p(|\theta|^2)$ , Lemma 1 and Theorem 1. (The consistency of the constrained estimator can be verified by using similar arguments as in the proof of Theorem 1 of Hannan (1973).)  $\square$

Received 15 November 2010

## REFERENCES

- BROCKWELL, P. J. (1993). Threshold ARMA processes in continuous time. In *Dimension Estimation and Models* (H. Tong, ed) 170–190. World Scientific Publishing, River Edge, NJ. [MR1307659](#)
- CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Statist.* **25** 573–578. [MR0065087](#)
- JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22** 389–395. [MR0585635](#)
- JONES, R. H. (1985). Time series analysis with unequally spaced data. *Handbook of Statistics* **5** (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds) 157–177. Elsevier. [MR0831747](#)
- HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. [MR0279952](#)
- HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Prob.* **10** 130–145. [MR0365960](#)

- HANSEN, L. P. and SARGENT, T. J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* **51** 377–87. [MR0688733](#)
- MCCRODIE, J. R. (2003). The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* **79** 9–16. [MR2021872](#)
- PANDIT, S. M. and WU, S. M. (1975). Unique estimates of the parameters of a continuous stationary stochastic process. *Biometrika* **62** 497–501. [MR0408152](#)
- PANDIT, S. M. and WU, S. M. (1983). *Time Series and System Analysis with Applications*. Wiley, New York.
- TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford. [MR1079320](#)
- TSAI, H. and CHAN, K. S. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statistica Sinica* **10** 989–998. [MR1789804](#)
- TSAI, H. and CHAN, K. S. (2005a). Maximum likelihood estimation of linear continuous time long memory processes with discrete time data. *J. Royal Stat. Soc., Ser. B* **67** 703–716. [MR2210688](#)
- TSAI, H. and CHAN, K. S. (2005b). Temporal aggregation of stationary and non-stationary continuous-time processes. *Scand. J. Stat.* **32** 583–597. [MR2232344](#)
- WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis*. Almqvist and Wiksell, Stockholm. [MR0040634](#)

Henghsiu Tsai  
Institute of Statistical Science  
Academia Sinica  
Taipei 11529  
Taiwan, R.O.C.  
Phone: 886-2-27835611 Ext. 215  
Fax: 886-2-27831523  
E-mail address: [htsai@stat.sinica.edu.tw](mailto:htsai@stat.sinica.edu.tw)

Kung-Sik Chan  
Dept. of Statistics and Actuarial Science  
University of Iowa  
Iowa City, Iowa 52242  
USA  
E-mail address: [kung-sik-chan@uiowa.edu](mailto:kung-sik-chan@uiowa.edu)

Patrick Fayard  
Dept. of Electrical and Computer Engineering  
McMaster University  
Hamilton, ON L8S 4K1  
Canada  
E-mail address: [fayardp@grads.ece.mcmaster.ca](mailto:fayardp@grads.ece.mcmaster.ca)