Department of Economics

Working Paper No. 0217

http://www.fas.nus.edu.sg/ecs/pub/wp/wp0217.pdf

Maximum Likelihood Estimation of ARMA Model with Error Processes for Replicated Observations

Wing-Keung Wong
National University of Singapore

Robert B. Miller
University of Wisconsin-Madison

Keshab Shrestha
Concordia University

Abstract: In this paper we analyse the repeated time series model where the fundamental component follows a ARMA process. In the model, the error variance as well as the number of repetition are allowed to change over time. It is shown that the model is identified. The maximum likelihood estimator is derived using the Kalman filter technique. The model considered in this paper can be considered as extension of the models considered by Anderson (1978), Azzalini (1981) and Wong and Miller (1990).

Keywords: ARMA model, Kalman filter, maximum likelihood estimation

JEL Classification: C10, C13

 Correspondence: Department of Economics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260. Tel: (65)-874-6014 Fax: (65)-775-2646 E-mail: ecswwk@nus.edu.sg .

1. Introduction

In Economics and Finance, many time series are not directly observable without errors. However, they are observed with errors. In some cases, many proxy observations can be obtained for the value of the series at one particular point in time. For example, an equilibrium wage rate for an industry may not be observable directly. However, we can obtian many different wage rates earned by workers in that industry. We can consider these wage rates as measurements of the unobservable equilibrium wage and incorporate the fact that these measurements contain some errors. In Finance, the stock price is an important variable. However, there are different measures of stock prices, e.g. open price and closing price. Such time series can be modelled as follows

$$z_{t,j} = y_t + e_{t,j}, \qquad j = 1, 2, \cdots, k_t, \quad \text{and} \quad t = 1, 2, \cdots, T \tag{1}$$

where y_t is the value of the fundamental time series at time t. For each y_t, we have k_t observations $z_{t,j}$ that are associated with k_t error components $e_{t,j}$. The error components are assumed to be independent white noise processes that are independent of the fundamental proces $y(t)$. To make the model general and interesting, it is assumed that the fundamental process $\{y_t\}$ is a stationary autoregressive moving average (ARMA) process. Such models are called repeated time series (RTS) models.

For non-repeated time series measurements, Kendall (1944) and Quenouille (1947) study the case of AR(2) with error while Walker (1960) extends the case to autoregressive processes with error. For these measurements, it is well-known that the maximum likelihood estimates of the parameters in the ARMA with error process cannot be obtained without imposing many restrictions on the parameters.

However, The theory of RTS is not well developed in time series analysis. Anderson (1978) and Azzalini (1981) study the theory of repeated time series measurement. They introduce a method to obtain the maximum likelihood estimates of the parameters for stationary low order autoregressive process. Wong and Miller (1990) analysed the model represented by equation (1) where the variance of the error components and the number of repititions were assumed to be the same for all time t.

This paper extends the previous works by allowing the variance of the error components to change over time. Furthermore, the number of the repetitions is also allowed to change over time. The maximum likelihood estimation technique is derived using the Kalman filtering. It is shown that all the parameters of the model is identified and the system dynamics used is stable.

2. Assumptions and Properties

Suppose the number of repetitions, k_t, at time t is greater than zero and suppose the observation $\{z_{t,j}\}$, $j = 1, \ldots, k_t$, are taken randomly for each time $t, t = 1, \ldots, T$. The error component, $e_{t,j}$, is independently distributed as $N(0, \sigma_{et}^2)$, while the signal component, y_t, follows an ARMA(p, q) model such that

$$\Phi(B)y_t = \Theta(B)\varepsilon_t \tag{2}$$

where $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, $\{\varepsilon_t\}$ is $N(0, \sigma_\varepsilon^2)$ and B is the backward shift operator such that $B^i x_t = x_{t-i}$. It is assumed that $\Phi(B)$ and $\Theta(B)$ have no common zeros, zeros of $\Phi(B)$ and $\Theta(B)$ are outside the unit circle.

Let $\mathbf{1}_t$ be a $k_t \times 1$ vector in which all elements are 1, I_t be the $k_t \times k_t$ identity matrix and $R_t = k_t^{-1}\mathbf{1}_t\mathbf{1}_t'$. We define the following variables

$$w_t = k_t^{-1}\mathbf{1}_t'\mathbf{z}_t\ , \quad a_t = k_t^{-1}\mathbf{1}_t'\mathbf{e}_t\ , \tag{3}$$

and

$$(I_t - R_t)\mathbf{z}_t = \mathbf{z}_t - w_t\mathbf{1}_t \qquad t = 1, \ldots, T\ ,$$

where $\mathbf{z}_t = (z_{t,1}, \cdots, z_{t,j}, \cdots, z_{t,k_t})'$ and $\mathbf{e}_t = (e_{t,1}, \cdots, e_{t,j}, \cdots, e_{t,k_t})'$, It is easy to show that $w_t = y_t + a_t$ where $\{a_t\}$ is $N(0, \sigma_{at}^2)$ with $\sigma_{at}^2 = \sigma_{et}^2/k_t$. In this model, $\{y_t\}$ is stationary. However, $\{w_t\}$ and $\{z_{t,j}\}$ may not be stationary because σ_{at}^2 and σ_{et}^2 may vary over time.

Under the normality and independence assumptions on $\{e_{t,j}\}$ and $\{y_t\}$, $\{w_t\}$ and $\{(I_t - R_t)z_t\}$ are independent. Therefore, the likelihood function

$$L\{\Phi, \Theta, \sigma_\varepsilon^2, \sigma_{at}^2 \mid \mathbf{z}_t, \quad t = 1, \ldots, T\} \tag{4}$$

can be written as

$$L_1\{\Phi, \Theta, \sigma_\varepsilon^2, \sigma_{at}^2 \mid \mathbf{w}\} L_2\{\sigma_{at}^2 \mid (I_t - R_t)\mathbf{z}_t \quad t = 1, \ldots, T\}\ .$$

Before finding estimates of Φ, Θ, σ_ε^2, σ_{at}^2 to maximize L, we first find $\{\hat{\sigma}_{at}^2\}$ which maximizes L_2. We define the index set $\Lambda = \{1, 2, \cdots, T\}$ which can be partitioned into Λ_i such that σ_{et}^2 is constant for each t in Λ_i and is different from those in other Λ_j. Then the maximum likelihood estimates of σ_{et}^2 and σ_{at}^2 are respectively

$$\hat{\sigma}_{et}^2 = \frac{\sum_{t\in\Lambda_i}\sum_{j=1}^{k_t}(z_{t,j} - w_t)^2}{\sum_{t\in\Lambda_i} k_t} \qquad \text{for each} \quad t \in \Lambda_i \tag{5}$$

and

$$\hat{\sigma}_{at}^2 = \frac{\hat{\sigma}_{et}^2}{k_t}\ ,$$

provided that $\sum_{t\in\Lambda_i}(k_t - 1)$ is greater than zero. In this paper we assume this condition holds and so σ_{et}^2, and consequently σ_{at}^2, can be estimated for any t. We note that in this paper, we only consider the situation in which σ_{et}^2 follows a step function. One may extend the theory by releasing this condition to include a more complicated situation. However, in practice, for example in our illustration, step function should be a good approximation. Similarly, one may release the ARIMA assumption on y_t and assume y_t follows a more complicated model like GARCH model.

The next step is to find the estimates which maximize L_1. Since $\hat{\sigma}_{at}^2$ can be obtained in (5), we treat it as a constant in L_1. That is, we find estimates of Φ, Θ and σ_ε^2 which maximize

$$L_3\{\Phi, \Theta, \sigma_\varepsilon^2 | \mathbf{w}, \sigma_{at}^2 \quad t = 1, \dots, T\} \tag{6}$$

where $\mathbf{w} = (w_1, w_2, \cdots, w_T)'$.

In the next two sections, we will discuss the approach of applying the Kalman filter technique to find estimates which maximize L_2 and L_3 iteratively and finally maximize L.

3. Recursive Estimation Procedure

In this section, we investigate the application of the Kalman filter to compute the value of the likelihood function, the variance of the noise, and the conditional linear least-square estimates of both $\{w_t\}$ and $\{y_t\}$. Then we discuss the maximum likelihood estimation of the unknown parameters.

The Kalman filter is frequently used in the estimation of the time series models. Mehra (1974) and Caines and Rissanen (1974) use Kalman recursive estimation to compute the exact likelihood of an ARIMA process. Akaike (1973, 1974, 1975) introduce Markovian representation which provides a minimal state space representation for recursive calculation of the likelihood function for a Gaussian ARMA process. Harvey and Phillips (1979) and

Jones (1980) extend the application of the Kalman filter to compute the likelihood of a stationary ARMA process with an error component. Kalman (1960, 1963) shows that the difference between the next available observation and the prediction from the best estimate of the current state is orthogonal to earlier observations.

Different state space forms can be used to model the same ARMA with error process. In this paper we choose the Markovian representation used by Jones (1980). The observation w_t in (3) is expressed in the following data generation equation:

$$w_t = HZ_t + a_t \tag{7}$$

where $H = (1, 0, \cdots, 0)$, Z_t is the $m \times 1$ state vector with its j^{th} element

$$Z_{j,t} = E[y_{t+j-1} | y_s, \quad s \leq t \quad] \quad \text{for } j = 1, \ldots, m$$

in which $m = \max(p, q+1)$, and y_t is defined in (2).

The first element of Z_t is $Z_{1,t}$ satisfying

$$Z_{1,t} = y_t \ .$$

The state vector Z_t can be expressed by the state transition equation

$$Z_t = FZ_{t-1} + G\varepsilon_t \tag{8}$$

where

$$F = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \phi_m & \phi_{m-1} & \cdot & \cdots & \phi_1 \end{pmatrix}$$

and

$$G = (1, g_1, \cdots, g_{m-1})'$$

in which $g_0 = 1$ and

$$g_j = \sum_{k=1}^{j} \phi_k g_{j-k} - \theta_j \quad \text{for } j = 1, \ldots, m-1\,. \tag{9}$$

In (9), $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$.

The state space dynamical system represented by equations (7) and (8) has some important properties. These properties will be discussed next.

Property 1: *The state space dynamical system is stable. This means that all the eigen values (i.e., characteristic roots) of F is inside the unit circle. This can be shown with the observation that the characteristic polynomial of F is given by*

$$\lambda^m - \phi_1 \lambda^{m-1} - \phi_2 \lambda^{m-2} - \cdots - \phi_{m-1}\lambda - \phi_m$$

Thus, the characteristic roots of F is given by the reciprocal of the zeros of $\Phi(B)$. Since we have assumed that the zeros of $\Phi(B)$ lie outside the unit circle, all the eigen values of F lie within the unit circle.

Beside being a stable dynamical system, it has the following properties (see appendix for the proof):

Property 2 *The state space dynamical system is identified in the sense that all its parameters are identified.*

In this recursive procedure, we first calculate a one-step prediction from Equation (8) such that

$$Z_{t+1|t} = F Z_{t|t}$$

where

$$Z_{t+j|t} = E[Z_{j+t} | w_s\,, \quad s \leq t \quad]$$

is the projection of Z_{j+t} on $\{w_s\,,\ s \leq t\}$. From (8), the covariance matrix of this prediction is

$$S_{t+1|t} = FS_{t|t}F' + \sigma_\varepsilon^2 GG'\ . \tag{10}$$

The predicted value of the next observation is

$$w_{t+1|t} = HZ_{t+1|t} = y_{t+1|t}$$

where $w_{t+j|t} = E[w_{j+t}|w_s\,,\ s \leq t\,]$ and $y_{t+j|t} = E[y_{j+t}|w_s\,,\ s \leq t\,]$ are the conditional linear least-square estimates of w_{t+j} and y_{t+j} respectively given w_s for $s \leq t$.

Using the next observation, the state vector estimate and its covariance matrix are updated by

$$Z_{t+1|t+1} = Z_{t+1|t} + \Delta_{t+1}[w_{t+1} - w_{t+1|t}\,] \tag{11}$$

and

$$S_{t+1|t+1} = S_{t+1|t} - \Delta_{t+1}HS_{t+1|t} \tag{12}$$

respectively, where

$$\Delta_t = S_{t|t-1}H'\{HS_{t|t-1}H' + \sigma_{at}^2\}^{-1}\ .$$

One may refer to Meinhold and Singpurwalla (1983) for the proof of Equations (11) and (12). They use a Bayesian approach to explain these equations. In this Kalman filtering process, the quantity $\xi_{t|t-1}$ is defined as

$$\xi_{t|t-1} = w_t - w_{t|t-1}$$

with variance $\nu_{t|t-1}$. Since

$$E(\xi_{t|t-1}) = 0\,,$$

we have

$$\nu_{t|t-1} = E({\xi_{t|t-1}}^2)\ . \tag{13}$$

Kailath (1968) shows that $\xi_{t|t-1}$ satisfies

$$E(\xi_{t|t-1}\xi_{s|s-1}) = [HS_{t|t-1}H' + \sigma_{at}^2]\delta_{t,s} \tag{14}$$

where $\delta_{t,t} = 1$ and $\delta_{t,s} = 0$ for $t \neq s$. This result enables us to transform the log-likelihood function of $y_1, \cdots, y_T$ into

$$l = \frac{1}{2}\sum_{t=i}^{T}\left[-\log 2\pi - \log \nu_{t|t-1} - \frac{{\xi_{t|t-1}}^2}{\nu_{t|t-1}}\right] . \tag{15}$$

Maximizing (15) is equivalent to maximizing L_3 in (6). There are various ways of finding the initial values of $Z_{0|0}$ and $S_{0|0}$. A good discussion of the methods is found in Harvey (1989).

Denote by $\eta = (\Phi, \Theta, \sigma_\varepsilon^2)'$ the set of unknown parameters. One may apply the Newton-Raphson method or the method of scoring to obtain the maximum likelihood estimate. If $\tilde{\eta}$ is the value of the parameter vector η at the previous iteration, the Newton-Raphson method gets the new estimate $\hat{\eta}$ such that

$$\hat{\eta} = \tilde{\eta} - \ddot{l}(\tilde{\eta})^{-1}\dot{l}(\tilde{\eta}) , \tag{16}$$

and the method of scoring gets

$$\hat{\eta} = \tilde{\eta} - E[\ddot{l}(\tilde{\eta})^{-1}]\dot{l}(\tilde{\eta}) , \tag{17}$$

where l is the likelihood function defined in (15) with its scoring function $\dot{l}$ and its Hessian matrix $\ddot{l}$. In this paper we use the notation $\dot{x}$ and $\ddot{x}$ to represent respectively the first and the second derivatives of x with respect to η for any x. Applying Equation (16) or Equation (17) iteratively, one will obtain the maximum likelihood estimate η^* of η.

For the asymptotic distribution of the estimate η^*, Dunsmuir (1983) has shown that under some regularity conditions

$$\sqrt{T}(\eta^* - \eta) \xrightarrow{D} N\left(0, I(\eta)^{-1}\right) ,$$

and $\ddot{L}(\eta^*)/2$ converges in mean square to the information matrix, $I(\eta)$. He also shows that $\ddot{L}(\hat{\eta})/2$ can be replaced by the estimate of

$$\frac{1}{2T}\sum_{t=i}^{T}\left[\frac{\dot{\nu}_{t|t-1}\dot{\nu}'_{t|t-1}}{{\nu_{t|t-1}}^2} + \frac{2\dot{\xi}_{t|t-1}\dot{\xi}'_{t|t-1}}{\nu_{t|t-1}}\right] . \tag{18}$$

One can check that under the assumptions of this model setting, the regularity conditions are satisfied and hence Equation (18) can be used to estimate the asymptotic variance of η^*. For computing this variance, one has to find $\nu_{t|t-1}$, $\dot{\nu}_{t|t-1}$ and $\dot{\xi}_{t|t-1}$. The variance $\nu_{t|t-1}$ can be obtained by (13). This leaves $\dot{\xi}_{t|t-1}$ and $\dot{\nu}_{t|t-1}$ to be estimated. For computing the maximum likelihood estimate η^* by applying Equation (16) or (17), first one has to estimate the scoring function $\dot{l}$ and the Hessian matrix $\ddot{l}$. It is easy to show that the scoring function is in terms of $\xi_{t|t-1}$, $\nu_{t|t-1}$, and their first derivatives. From Equation (18), the elements in the Hessian matrix are also in terms of these variables. Hence, we have to estimate $\dot{\xi}_{t|t-1}$ and $\dot{\nu}_{t|t-1}$ to obtain the maximum likelihood estimates of η. We will discuss how to estimate the derivatives $\dot{\xi}_{t|t-1}$ and $\dot{\nu}_{t|t-1}$ in the next section.

4. Estimating the Derivatives and Maximum Likelihood Estimation

In this section we will discuss the recursive estimation procedure for the derivatives $\dot{\xi}_{t+1|t}$ and $\dot{\nu}_{t+1|t}$. Then, we will discuss the maximum likelihood estimates for the unknown parameters and their asymptotic covariance matrix.

The derivatives $\dot{\xi}_{t+1|t}$ and $\dot{\nu}_{t+1|t}$ can be expressed as

$$\dot{\xi}_{t+1|t} = -\dot{w}_{t+1|t} = -\dot{y}_{t+1|t} = -H\dot{Z}_{t+1|t}$$

and

$$\dot{\nu}_{t+1|t} = H\dot{S}_{t+1|t}H'$$

where $\dot{Z}_{t+1|t}$ and $\dot{S}_{t+1|t}$ can be expressed as

$$\dot{Z}_{t+1|t} = F\dot{Z}_{t|t} + \dot{F}Z_{t|t}$$

and

$$\dot{S}_{t+1|t} = F\dot{S}_{t|t}F' + 2\dot{F}\dot{S}_{t|t}F' + 2\sigma_\varepsilon^2\dot{G}G' \ . \tag{19}$$

Using the next observation, the estimates can be updated by

$$\dot{Z}_{t+1|t+1} = \dot{Z}_{t+1|t} + \dot{\Delta}_{t+1}\xi_{t+1|t} + \Delta_{t+1}\dot{\xi}_{t+1|t}$$

and

$$\dot{S}_{t+1|t+1} = \dot{S}_{t+1|t} - \dot{\Delta}_{t+1}H\dot{S}_{t+1|t} - \Delta_{t+1}H\dot{S}_{t+1|t}$$

respectively, where

$$\begin{aligned}\dot{\Delta}_{t+1} \quad = \quad & -S_{t+1|t}H'\{HS_{t+1|t}H' + \sigma_{at}^2\}^{-2}(H\dot{S}_{t+1|t}H') \\ & +\dot{S}_{t+1|t}H'\{HS_{t+1|t}H' + \sigma_{at}^2\}^{-1} \ .\end{aligned}$$

Applying the same principles in the estimation of $Z_{0|0}$ and $S_{0|0}$, we can find the initial estimates of $\dot{Z}_{0|0}$ and $\dot{S}_{0|0}$. Then, one can apply the Kalman filter recursive estimation precedure to find the estimate of η. The estimate of $w_{t|t}$ which is equal to $HZ_{t|t}$ can also be found in the process. Replacing w_t by $w_{t|t}$ in Equation (5), one can get a new estimate of σ_{et}^2 and consequently get a new estimate of σ_{at}^2. Substituting the new estimate of σ_{at}^2 into equations for the recursive estimation procedure, one can obtain the new estimate of η. This iterative procedure is to find the estimates which maximize L_2 and L_3 iteratively and finally the estimates will converge to the maximum likelihood estimates which maximize as L defined in (4).

5. Example

We illustrate the applicatin of the innovation transformation to maximum likelihood estiamtion with the modeling of "George" robot data, $\{z_{ij}\}$, obtained from Bill Fulkerson, Deere and Company. Wong and Miller (1990) use the same dataset and we will compare the result using the approach in this paper with the result in Wong and Miller (1990). John Deere markets a repeatability test unit designed to mearsure the ability of a robot arm to return to a designed point. For each time t, five equally spaced measurements are recorded in a very short time such that the five measurements seem to be simultaneous. The data are measured in inches.

The data, $\{z_{tj}\}$ for $j = 1, \cdots, 5$, are repeated time series measurements of the "George" Robot arm's positions which are assumed to satisfy

$$z_{t,j} = y_t + e_{t,j}, \qquad j = 1, 2, \cdots, 5, \quad \text{and} \quad t = 1, 2, \cdots, 206$$

where $\{y_t\}$ is defined in Session 2 and $e_{t,j}$ is independently distributed as $N(0, \sigma^2_{et})$ in which σ^2_{et} is allowed to vary over time. Define $w_t = \sum_{j=1}^{5} z_{t,j}/5$ and $a_t = \sum_{j=1}^{5} a_{t,j}/5$. Then $\{y_t\}$ satisfies

$$w_t = y_t + a_t$$

where $\{a_t\}$ independently distributed as $N(0, \sigma^2_{at})$ with $\sigma^2_{at} = \sigma^2_{et}/5$ for each t. We first study the variances of the error components. Define the estimated error component $\hat{e}_{tj} = z_{t,j} - w_t$ for each t.

We choose $i = 1$ and use $F = \hat{\sigma}^2_{ei}/\hat{\sigma}^2_{ej}$ to test the hypothesis $H_0 : \sigma^2_{ei} = \sigma^2_{ej}$ for $j > i$. For all the j such that H_0 is not rejected, we further test the hypothesis H'_0 : equality of the variances, see p377 in Lehman (1991) for the test statistic. For all the j such that H'_0 is not rejected, we form Λ_1. We then choose the smallest i and repeat the process to form Λ_2 and so on. Following this procedure, we find that the standard deviation of the error components e_{tj} are estimated to be is 0.0001446 for all the periods except for the periods in the following table:

S.D. ($\times 10^3$) of the Error Component and the Corresponding Time Periods (t)

t	$\hat{\sigma}_{ej}$	t	$\hat{\sigma}_{ej}$	t	$\hat{\sigma}_{ej}$	t	$\hat{\sigma}_{ej}$	t	$\hat{\sigma}_{ej}$	t	$\hat{\sigma}_{ej}$
13	.0000	27	.3050	33	.2510	40	.3701	53	.0447	59	.0447
60	.2510	64	.3701	72	.5727	74	.2828	82	.0447	90	.0447
105	.2490	145	.0447	174	.0447	176	.0447	189	.3507	198	.0447

The index set $\Lambda = \{1, 2, \cdots, T\}$ is supposed to be partitioned into $\{\Lambda_i\}$ such that σ^2_{et} is constant for each t over Λ_i and is different from those in other Λ_j. The standard error of the error components e_{tj} is 0.0001446 for all the periods except for a few data.

Wong and Miler (1990) find that both the series $\{y_t\}$ follows an ARIMA(0,1,1) model. We use their finding as initial estimate. Based on the technique discussed in this paper, we find that

$$(1 - B)y_t = (1 - \theta B)\varepsilon_t$$

where ε_t is $N(0, \sigma^2_{\varepsilon_t})$ with $\hat{\sigma}_{\varepsilon_t} = 0.0001936$ and $\hat{\theta} = -0.3825$ with standard error 0.07879. The ACF and PACF of the innovations are:

ACF and PACF of the Innovations

Lag	1	2	3	4	5	6	7	8
ACF	0.017	.030	.003	.031	.017	-.006	-.067	-.037
PACF	0.017	.030	.002	.030	.016	-.009	-.068	-.036

The standard error of $\hat{\theta}$ in our approach is 0.07879 which is much smaller than the standard error, 0.1282, obtained by the "hybrid" estimation technique in Wong and Miller (1990). The disadvantage of using our approach is to require the normality assumption while the approach in Wong and Miller (1990) does not. However, we do not reject that both the error component estimates and the innovations are normally distributeed. Hence, the model in this paper is preferred in this example.

6. Summary

In this paper we have analysed the repeated time series model where the fundamental components of the series follows a stationary ARMA process. The model allows the error component variances to change over time. Furthermore, the number of repitition is also allowed to change over time. The Kalman filter technique is used to obtain the maximum likelihood estimates of the parameters of the model. It is shown that the model is identified. It is easy to extend the results to include the missing observation case. One can also extend this to the case in which the signal component is non-stationary and non-linear.

The disadvantage of using our approach is to require the normality assumption. However, one may easily incorporate the Bayesian technique, see Matsumura, et al (1990) and Wong and Bian (2000), or the modified maximum likelihood estimation approach, see Tiku and Wong (1998) and Tiku et al (1999a,b,2000), to our approach to release the normality assumption. Further Extension includes application of the model to investment decisions, see Thompson and Wong (1991, 1996), Wong and Li (1999), Manzur, et al (1999) Wong et al (2001).

References

Akaike, H., 1973. Maxium likelihood identification of Gaussian autoregressive moving average models. Biometrika 60, 255-265.

Akaike, H., 1974. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics 26, 363-387.

Akaike, H., 1975. Markovian representation of stochastic processes by canonical variables. SIAM J. Control 13, 162-173.

Akaike, H., 1978. Covariance matrix computation of the state variable of a stationary gaussian process. Research Memorandum No. 139, The Institute of Statistical Mathematics, Togyo.

Anderson, T.W., 1978. Repeated measurement on autoregressive processes. Journal of the American Statistical Association 73, 371-378.

Azzalini, A., 1981. Replicated observations of low order autoregressive time series. Journal of Time Series Analysis 2, 63-70.

Caines, P.E., Rissanen, J., 1974. Maximum likelihood estimation of parameters in multivariate gaussian stochastic processes. IEEE Transactions on information theory. IT-20, 102-104.

Dunsmuir, W., 1983. Large sample properties of estimation in time series observed at unequally spaced times. Brillinger, D., Fienberg, J., Gani, J., Hartigan, J., Krickeberg, K., Time Series Analysis of Irregularly Observed Data. New York: Springer, 58-77.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Harvey, A.C., Phillips, G.D., 1979. Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika 66, 49-58.

Jones, R.H., 1980. Maximum likelihood fitting of ARMA models to time series with miss observations. Technimetrics 22, 389-395.

Kailath, T., 1968. An innovations approach to least-squares estimation part I: linera filtering in additive white noise. IEEE Transactions on Automatic Control. AC-13, 6, 646-660.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans ASME, J. Basic Engineering. 82, 34-35.

Kalman, R.E., 1963. New methods in Wiener filtering theory. Bogdanoff, J.L. and Kozin, F. *Procceding of first symposium on engineering application of random function theory and probability.* New York: Wiley.

Kendall, M.G., 1944. On autoregressive time series. Biometrika. 33, 105-122.

Lehmann, E.L., 1991. Testing statistical hypotheses. Wadsworth & Brooks.

Manzur, M., Wong, W.K., and Chau, I.C., 1999. Measuring international competitiveness : experience from East Asia. Applied Economics 31, 1383-1391.

Matsumura, E.M., Tsui, K.W., and Wong, W.K., 1990. An Extended Multinomial-Dirichlet Model for Error Bounds for Dollar-Unit Sampling. Contemporary Accounting Research 6, No 2-I, p485-500.

Mehra, R.K., 1974. Identification in control econometrics; similarities and differences. Annals of Economics and Social Measurement 3, 21-47.

Meinhold, R.J., Singpurwalla, N.D., 1983. Understanding the Kalman filter. The American Statistician 37, 2, 123-7.

Quenouille, M.H., 1947. A large-sample test for the goodness of fit of autoregressive schemes. Journal of Royal Statistical Society A 110, 123-129.

Thompson H.E., Wong, W.K., 1991. On the unavoidability of 'scientific' judgement in estimating the cost of capital. Managerial and Decision Economics 12, 27-42.

Thompson H.E., Wong, W.K., 1996. Revisiting 'Dividend Yield Plus Growth' and Its Applicability. Engineering Economist 41, No. 2, 123-147.

Tiku, M.L., Wong, W.K., 1998. Testing for unit root in AR(1) model using three and four moment approximations. Communications in Statistics: Simulation and Computation 27 (1), 185-198.

Tiku, M.L., Wong, W.K., Bian, G., 1999a. Estimating Parameters in Autoregressive Models in Non-normal Situations: symmetric Innovations. Communications in Statistics: Theory and Methods 28(2), 315-341.

Tiku, M.L., Wong, W.K., Bian, G., 1999b. Time series models with asymmetric innovations, Communications in Statistics: Theory and Methods 28, no. 6, 1331–1360.

Tiku, M.L., Wong, W.K., Vaughan, D.C., Bian, G., 2000. Time series models with nonnormal innovations: symmetric location–scale distributions. Journal of Time Series Analysis 21, No. 5, 571-596.

Walker, A.M., 1960. Some consequences of superimposed error in time series analysis. Biometrika 47, 33-43.

Wong W.K., Bian, G., 2000. Robust Bayesian Inference in Asset Pricing Estimation. Journal of Applied Mathematics & Decision Sciences 4(1), 65-82.

Wong W.K., Chew, B.K., Sikorski, D., 2001, Can P/E ratio and bond yield be used to beat stock markets? Multinational Finance Journal 5, 59-86.

Wong, W.K., Li, C.K., 1999. A note on convex stochastic dominance theory. Economics Letters 62, 293-300.

Wong, W.K., Miller, R.B., 1990. Repeated Time Series Analysis of ARIMA-NOISE Models. Journal of Business and Economic Statistics 8, 243-250.

APPENDIX

Proof of property 2:

From the dynamical system represented by equations (7) and (8) we can find observationally equivalent dynamical system using the transformation $Z_t^e = T^{-1}Z_t$. The equivalent system is given by

$$
\begin{aligned}
w_t &= H^e Z_t^e + a_t \\
Z_t^e &= F^e Z_{t-1}^e + G^e \eta_t
\end{aligned}
$$

where $F^e = T^{-1}FT$, $G^e = T^{-1}G$ and $H^e = HT$. Thus, for the model to be identified, there must be enough a priori restrictions on H, F and G so that the only allowed transformation matrix T is an identity matrix. Let T_i^r and T_j^c denote the ith row and the jth column of the matrix T respectively. Similarly, let I_i^r and I_j^c denote the ith row and the jth column of an $m \times m$ identity matrix. It is clear from the model that the matrices H and H^e are both restricted to be equal to I_1^r. Thus we have

$$I_1^r = H^e = HT = I_1^r T = T_1^r$$

Therefore, the first row of the transformation matrix T is equal to the first row of an identity matrix. According to the model both the matrices F and F^e must be such that their first upper off diagonal elements must be equal to unity. Furthermore, all other elements except the lasr row must be zero. The relationship between these two matrices are given by

$$FT = TF^e \, .$$

TF^e is equal to

$$\begin{pmatrix} \Phi_m^e T_{1m} & T_{11} + \Phi_{m-1}^e T_{1m} & T_{12} + \Phi_{m-2}^e T_{1m} & \cdots & T_{1,m-1} + \Phi_1^e T_{1m} \\ \Phi_m^e T_{2m} & T_{21} + \Phi_{m-1}^e T_{2m} & T_{22} + \Phi_{m-2}^e T_{2m} & \cdots & T_{2,m-1} + \Phi_1^e T_{2m} \\ \Phi_m^e T_{3m} & T_{31} + \Phi_{m-1}^e T_{3m} & T_{32} + \Phi_{m-2}^e T_{3m} & \cdots & T_{3,m-1} + \Phi_1^e T_{3m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \Phi_m^e T_{m-1,m} & T_{m-1,1} + \Phi_{m-1}^e T_{m-1,m} & T_{m-1,2} + \Phi_{m-2}^e T_{m-1,m} & \cdots & T_{m-1,m-1} + \Phi_1^e T_{m-1,m} \\ \Phi_m^e T_{mm} & T_{m1} + \Phi_{m-1}^e T_{mm} & T_{m2} + \Phi_{m-2}^e T_{mm} & \cdots & T_{m,m-1} + \Phi_1^e T_{mm} \end{pmatrix}$$

and

$$FT = \begin{pmatrix} T_{21} & T_{22} & T_{23} & \dots & T_{2m} \\ T_{31} & T_{32} & T_{33} & \dots & T_{3m} \\ T_{41} & T_{42} & T_{43} & \dots & T_{4m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ T_{m1} & T_{m2} & T_{m3} & \dots & T_{mm} \\ \hat{\Phi}' T_1^c & \hat{\Phi}' T_2^c & \hat{\Phi}' T_3^c & \dots & \hat{\Phi}' T_m^c \end{pmatrix}$$

where

$$\hat{\Phi} = \begin{pmatrix} \Phi_m & \Phi_{m-1} & \Phi_{m-2} & \dots & \Phi_2 & \Phi_1 \end{pmatrix}'$$

Since $T_1^r = I_1^r$, the first row of TF^e must be equal to I_2^r. This implies that the first row of FT, which is the second row of matrix T, is equal to I_2^r. Similarly, it is clear that the second row of TF^e is equal to I_3^r which implies that the third row of matrix T is equal to I_3^r. Following the similar argument, is can be shown that the ith row of T is equal to the ith row of an $m \times m$ identity matrix for $i = 1, 2, ..., m$. This proves that the equivalent transformation is identity. Note that from the elements of the matrices F and G the parameters of the original model are exactly identified. Thus, the model is identified.