Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection

Author(s): Chung Chen and George C. Tiao

Source: *Journal of Business & Economic Statistics*, Jan., 1990, Vol. 8, No. 1 (Jan., 1990), pp. 83-97

Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association

Stable URL: https://www.jstor.org/stable/1391755

Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection

Chung Chen
Department of Quantitative Methods, Syracuse University, Syracuse, NY 13244-2130

George C. Tiao
Graduate School of Business, University of Chicago, Chicago, IL 60637

The main purpose of this article is to assess the performance of autoregressive integrated moving average (ARIMA) models when occasional level shifts occur in the time series under study. A random level-shift time series model that allows the level of the process to change occasionally is introduced. Between two consecutive changes, the process behaves like the usual autoregressive moving average (ARMA) process. In practice, a series generated from a random level-shift ARMA (RLARMA) model may be misspecified as an ARIMA process. The efficiency of this ARIMA approximation with respect to estimation of current level and forecasting is investigated. The results of examining a special case of an RLARMA model indicate that the ARIMA approximations are inadequate for estimating the current level, but they are robust for forecasting future observations except when there is a very low frequency of level shifts or when the series are highly negatively correlated. A level-shift detection procedure is presented to handle the low-frequency level-shift phenomena, and its usefulness in building models for forecasting is demonstrated.

KEY WORDS: ARMA models; Estimation of current level; Forecasting; RLARMA models.

1. INTRODUCTION

Consider the general Gaussian (autoregressive moving average) ARMA(p, q) model

$$\varphi(B)z_t = \theta(B)a_t, \tag{1.1}$$

where $\{z_t\}$ is the observable time series, $\varphi(B) = (1 - \varphi_1 B - \cdots - \varphi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$, B is the backshift operator, all the zeros of $\varphi(B)$ and $\theta(B)$ are on or outside the unit circle, and $\{a_t\}$ is a series of iid $N(0, \sigma^2)$ variables. During the past 15 years, this class of models has been widely used to represent real time series in many areas, especially for forecasting in business and economic applications. Model (1.1) assumes that observations are homogeneously generated from the same probabilistic structure. In practice, however, time series data often contain structural changes or discrepant observations. It is critical to investigate to what extent such phenomena can make standard statistical methods less efficient or even invalid.

1.1 An Example

Figure 1 shows the seasonally filtered series of logarithmic monthly variety-store sales from January 1968 to September 1979. The original logarithmic variety-store series was analyzed by Chang and Tiao (1982); see also Bell, Hillmer, and Tiao (1983). In their analysis, the fitted model before adjustment for outliers can be written as

$$(1 + .40B + .27B^2)(1 - B)z_t = a_t, \tag{1.2}$$

where

$$z_t = (1 - .81B^{12})^{-1}(1 - B^{12})y_t, \tag{1.3}$$

$\{y_t\}$ is the observable logarithmic series, and $\{a_t\}$ is supposed to be a series of homogeneous white noise. For simplicity, we shall work with the seasonally filtered series $\{z_t\}$ to motivate the problems considered in this article.

From (1.2), z_t follows an AR(3) model in which the autoregressive polynomial contains a differencing factor $(1 - B)$. On the other hand, Figure 1 shows that a linear trend appears to exist in the data with an obvious level drop occurring in 1976. If this level drop is ignored, one may entertain a model containing a linear-trend term and an AR(3) noise term of the form

$$z_t = c + \omega_0 t + N_t, \tag{1.4}$$

where $(1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3)N_t = a_t$. The estimation results of this model are given in Table 1. Standard model checking based on residuals autocorrelations does not indicate evidence of model inadequacy. The estimated trend coefficient $\hat{\omega}_0 = .24 \times 10^{-4}$ with a t value of .02 is exceedingly small, however, which seems incongruent with the behavior of the series.

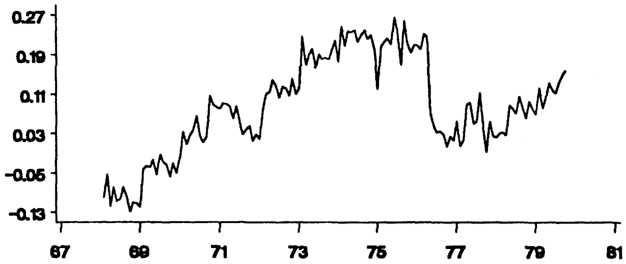

Figure 1. Seasonally Filtered Logarithmic Monthly Variety-Store Sales.

This calculation shows the need for careful examination of the homogeneity assumption underlying Model (1.4). Indeed, applying the level-shift detection procedure considered by Chen (1984), which is summarized later in (4.17) and (4.18), to the residuals of (1.4), we find strong evidence of a level drop occurring at the time point $t^{(1)}$ = April 1976. To adjust for the effect of this unusual event, we may entertain the following modified model:

$$z_t = c + \omega_0 t + \omega_1 LS_t(t^{(1)}) + N_t, \qquad (1.5)$$

where

$$LS_t(t^{(1)}) = 0, \qquad t < t^{(1)}$$
$$= 1, \qquad t \geq t^{(1)}.$$

The estimates of Model (1.5) are also reported in Table 1. It shows that the adjustment of a single level shift leads to appreciable changes in the estimates of the parameters $(\varphi_1, \varphi_2, \varphi_3, c, \omega_0)$ and a large reduction in the estimated variance σ^2. Repeating the level-shift detection procedure based on the residuals of Model (1.5), two additional possible level shifts have been indicated at $t^{(2)}$ = September 1970 and $t^{(3)}$ = January 1972. If all three level shifts are taken into account, the estimation results of Model (1.6) are again in Table 1:

$$z_t = c + \omega_0 t + \sum_{k=1}^{3} \omega_k LS_t(t^{(k)}) + N_t. \qquad (1.6)$$

We see that the estimated linear-trend coefficient $\hat{\omega}_0$ changes markedly from $.24 \times 10^{-4}$ for (1.4) to $.17 \times 10^{-2}$ for (1.6) after the level adjustments. The latter estimate, with a t value of 3.84, is seen to be compatible with the behavior of the series exhibited in Figure 1. Now, examining the zeros of the fitted autoregressive polynomial $(1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3)$, we consider an alternative parameterization $(1 - \varphi_1^* B)(1 - \varphi_2^* B - \varphi_3^* B^2)$ for the noise term N_t, where φ_1^* is the zero of B^{-1} closest to the unit circle. The factorization of fitted results gives for (1.4), $(1 - .97B)(1 + .35B + .24B^2)$; for (1.5), $(1 - .93B)(1 + .48B + .31B^2)$; and for (1.6), $(1 - .89B)(1 + .51B + .34B^2)$. In particular, the estimate of φ_1^* decreases from $\hat{\varphi}_1^* = .97$ for (1.4) to $\hat{\varphi}_1^* = .93$ for (1.5) and then to $\hat{\varphi}_1^* = .89$ for (1.6), and we see that prior to level-shift adjustments the fitted model (1.4) is in reasonably close agreement with the factor $(1 + .40B + .27B^2)(1 - B)$ in (1.2). Although one cannot reject the hypothesis that there is a unit root in the AR polynomials for (1.4), (1.5), and (1.6) (Dickey and Fuller 1979; Said and Dickey 1985), it is clear that the fitted AR part of Model (1.6) shows much less evidence of nonstationarity compared with that of (1.4). This is implying that the differencing operator $(1 - B)$ simply attempts to model the occasional level changes in this series. It should be pointed out that, as reported by Chang and Tiao (1982), the level shift at $t^{(1)}$ = April 1976 was associated with the fact that a major chain store went out of business at that time, but the reasons for the other two indicated shifts are unknown.

The key point of the preceding example is that occasional level shifts in real time series can lead to models of a very different nature, depending on whether these changes are appropriately taken into account. It is also of interest to examine the potential influence of adjusting occasional level shifts on the forecasting performance. For illustration, let t_0 = September 1978 be the forecast origin and observations from October 1978 to September 1979 be reserved for calculating forecast errors. A model of the form (1.2), which contains an exact differencing factor $(1 - B)$, and Models (1.4) and (1.6) are estimated, respectively, based on observations up to t_0. Table 2 lists the 12-period forecast errors. As expected, the forecast errors of (1.2) and (1.4) are in close agreement and are much worse than those of (1.6). For this particular forecast period, we find that adjusting occasional level shifts helps improve forecast accuracy.

Table 1. Time Series Modeling of Variety-Store Series

Model	c	ω_0	φ_1	φ_2	φ_3	$10^3 \times \hat{\sigma}^2$	ω_k	
(1.4)	.134	$.24 \times 10^{-4}$	.62	.10	.23	1.061		
	(.176)	(.0014)	(.08)	(.10)	(.08)			
(1.5)	−.069	.0026	.45	.13	.29	.791	ω_1	−.19
	(.050)	(.0005)	(.08)	(.09)	(.08)			(.03)
(1.6)	−.087	.0017	.38	.12	.30	.694	ω_1	−.19
	(.031)	(.0004)	(.08)	(.09)	(.08)			(.02)
							ω_2	.06
								(.02)
							ω_3	.07
								(.02)

NOTE: Estimated standard errors are in parentheses.

Table 2. Forecast Errors for Variety-Store Series (lead 1 = October 1978)

	Number of lead											
Model	1	2	3	4	5	6	7	8	9	10	11	12
(1.2)	−.023	−.006	.002	−.052	−.009	−.031	−.061	−.046	−.041	−.063	−.077	−.086
(1.4)	−.024	−.007	.004	−.053	−.011	−.033	−.064	−.050	−.045	−.068	−.083	−.092
(1.6)	−.015	.001	−.008	−.041	.003	−.018	−.047	−.030	−.024	−.044	−.057	−.064

1.2 The Random Level-Shift Problem

The preceding example motivates the study of the following problems: (a) What is an appropriate model to describe the phenomena of occasional level shifts in time series? (b) What would be the potential loss of forecasting efficiency by ignoring the information of such level shifts? (c) Is there an efficient way to detect such occasional level shifts? This article attempts to address the issues in (a) and (b). A procedure for (c) will be given with illustrations of its usefulness in model building, but its detailed properties will be discussed elsewhere.

General statistical approaches to problems of occasional stuctural changes so far consist of (a) developing methods to detect changes or outliers at known or unknown time point (e.g., Box and Tiao 1975; Chang and Tiao 1982; Hinkley 1970, 1972; Page 1955, 1957), (b) devising robust procedures for data analysis in the possible presence of structure changes or outliers (Martin 1980), and (c) modeling the nature of the changing behavior or outlying observations and deriving methods according to the proposed models (Box and Tiao 1968; Chernoff and Zacks 1964; Kander and Zacks 1966; Taylor 1980; Yao 1984). In particular, the models considered by Chernoff and Zacks, Taylor, and Yao share the characteristic that the structure of the series may change occasionally and the time periods of structural changes are determined by a stochastic process. The models considered by Chernoff and Zacks and Yao, however, are perhaps too simple to describe usual time series data, and Taylor's model is so flexible that it needs an identification procedure to specify a suitable form of the model through the data to gain any insight from the model.

In this article, we focus on the phenomena of occasional level shifts in time series. It is known that stationary ARMA models can represent series that homogeneously fluctuate around a constant level and nonstationary ARMA models can describe series in which the level changes from one time period to another. When the level of the series changes only occasionally, as illustrated by the preceding example, the usual ARMA model may not be appropriate. Now occasional level shifts may be regarded as interventions with fixed effects at unknown time points; this approach was taken by Hinkley (1970, 1972). An alternative way is to assume that the effect and the time point of the level shift are both subject to some probability rules, as considered by Chernoff and Zacks (1964) and Yao (1984). The random-effect formulation is intuitively appealing because economic and environmental time series are affected by many unusual events, the occurrence and the impact of which may be described by probability laws. Furthermore, it provides a convenient framework to assess the performance of standard time series method on series with level shifts by evaluating the contributions of the level shifts to forecast (or level estimation) mean squared error (MSE). In Section 2, a random level-shift ARMA model is introduced. The connection between this model and standard autoregressive integrated moving average (ARIMA) model-building procedures will be discussed in Section 3. In Section 4, the efficiency of an ARIMA approximation to the random level-shift ARMA model is investigated. Section 5 outlines some possible generalizations.

2. THE RANDOM LEVEL–SHIFT ARMA MODEL

A time series $\{y_t\}$ is said to follow a random level-shift ARMA (RLARMA) model with order (p, q), if the stochastic structure of $\{y_t\}$ is defined as

$$y_t = \mu_t + x_t, \tag{2.1}$$

where $\mu_t = j_t\eta_t + \mu_{t-1}$ $(t = 2, 3, \ldots, \mu_1)$ is an unknown initial level of the process; $\{x_t\}$ follows a stationary Gaussian ARMA (p, q) model such that $\phi(B)x_t = \theta(B)a_t$, $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$, and $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$; all the zeros of $\phi(B)$ and $\theta(B)$ are outside the unit circle; and $\{a_t\}$ are iid $N(0, \sigma^2)$. It is also assumed that $\{\eta_t\}$ are iid $N(0, \kappa\sigma^2)$; $\{j_t\}$ are iid such that $\Pr(j_t = 1) = \lambda$ and $\Pr(j_t = 0) = 1 - \lambda$; and $\{a_t\}$, $\{\eta_t\}$, and $\{j_t\}$ are mutually independent.

A random level-shift ARMA process $\{y_t\}$ can be regarded as the sum of a level process $\{\mu_t\}$ and a stationary ARMA process $\{x_t\}$. In the level process $\{\mu_t\}$, the random variable j_t is an indicator variable for the occurrence or nonoccurrence of level shift, the frequency of the level shift is determined by λ, and the magnitude of the shifts are given by the random variables η_t. If $\lambda = 0$, the level of $\{y_t\}$ stays unchanged and the structure of $\{y_t\}$ reduces to that of a stationary ARMA(p, q) process. If $\lambda = 1$, the level changes from period to period. In this case, $\{y_t\}$ is the sum of a Gaussian random-walk process and the stationary process $\{x_t\}$ and hence follows a nonstationary ARIMA$(p, 1, q^*)$ model, where $q^* = \max(p, q + 1)$. When $0 < \lambda < 1$, the degree of nonstationarity of the $\{y_t\}$ process is between that of a stationary ARMA(p, q) process and a nonstationary ARIMA(p, 1, q^*) process. In this sense, the RLARMA process can be used to represent series that exhibit nonstationarity between zero differencing and first-order differencing.

3. RLARMA MODEL AND STANDARD ARIMA MODEL-BUILDING PROCEDURE

In this article, we shall not be concerned with direct estimation of the parameters of the RLARMA model (2.1). Our main focus will be on assessing the performance of an ARIMA approximation to the RLARMA model for forecasting and estimation of current level. For the model in (1.1), widely used specification procedure proposed by Box and Jenkins (1976) is (a) to difference y_t as many times as is needed to produce stationarity and then (b) to identify the resulting stationary ARMA process. The necessity of differencing the series is often signalled by the slowly decaying behavior of the sample autocorrelation function (SACF). The following theorem gives the asymptotic properties of the SACF for series generated from an RLARMA process.

Theorem 1. Let $\{y_t\}$ ($t = 1, \ldots, T$) be observations of an RLARMA(p, q) process, as defined in (2.1). If the probability of level shift λ is strictly positive, then for any positive integer l

$$\frac{\sum_{t=1}^{T-l} y_t y_{t+l}}{\sum_{t=1}^{T} y_t^2} \xrightarrow{p} 1 \quad \text{as } T \longrightarrow \infty. \tag{3.1}$$

The theorem follows from the results of lemma 2.6 and corollary 2.6 of Tiao and Tsay (1983).

To illustrate the pattern of SACF for finite samples,

Table 3. Mean and Standard Deviation (SD) of SACF of an RLARMA (0, 0) Process: 500 Iterations

	Lag											
	1	2	3	4	5	6	7	8	9	10	11	12
T = 200, κ = 9												
λ = .05												
Mean	.84	.81	.79	.77	.75	.73	.71	.69	.67	.65	.63	.62
SD	.15	.16	.16	.17	.18	.18	.18	.19	.19	.20	.20	.20
λ = .02												
Mean	.67	.65	.63	.62	.60	.58	.56	.55	.53	.52	.50	.49
SD	.26	.26	.26	.26	.26	.26	.26	.25	.26	.25	.25	.25
λ = .01												
Mean	.48	.46	.45	.44	.43	.41	.40	.39	.38	.36	.35	.34
SD	.33	.33	.32	.32	.31	.31	.30	.30	.29	.29	.28	.28
T = 200, κ = 25												
λ = .05												
Mean	.92	.89	.86	.84	.81	.79	.76	.74	.72	.69	.67	.65
SD	.08	.09	.11	.12	.13	.14	.15	.16	.16	.17	.18	.18
λ = .02												
Mean	.81	.79	.76	.74	.72	.69	.67	.65	.63	.61	.59	.57
SD	.20	.20	.21	.21	.21	.22	.23	.23	.24	.24	.24	.24
λ = .01												
Mean	.60	.58	.56	.54	.53	.51	.50	.48	.47	.45	.44	.43
SD	.35	.35	.35	.34	.33	.34	.33	.33	.32	.32	.31	.31
T = 1,000, κ = 9												
λ = .05												
Mean	.91	.90	.89	.89	.88	.87	.86	.86	.85	.84	.83	.83
SD	.08	.09	.09	.10	.10	.11	.11	.12	.12	.13	.13	.14
λ = .02												
Mean	.93	.92	.92	.91	.91	.90	.90	.89	.88	.88	.87	.87
SD	.06	.06	.06	.07	.07	.07	.08	.08	.08	.09	.09	.09
λ = .01												
Mean	.87	.87	.86	.86	.85	.85	.84	.84	.83	.83	.82	.82
SD	.11	.11	.11	.11	.11	.12	.12	.12	.12	.12	.13	.13
T = 1,000, κ = 25												
λ = .05												
Mean	.96	.95	.94	.94	.93	.92	.91	.90	.89	.89	.88	.87
SD	.04	.04	.05	.06	.06	.07	.07	.08	.09	.09	.10	.10
λ = .02												
Mean	.97	.96	.96	.95	.95	.94	.93	.93	.92	.92	.91	.91
SD	.03	.03	.03	.04	.04	.04	.05	.05	.05	.06	.06	.06
λ = .01												
Mean	.94	.93	.93	.92	.92	.91	.91	.90	.90	.89	.88	.88
SD	.06	.06	.06	.06	.07	.07	.07	.07	.08	.08	.08	.08

Table 3 gives the means and standard deviations of the SACF from a simulation experiment based on an RLARMA(0, 0) process with selected values of λ, κ, and T. The averaged SACF behaves as one would expect from Theorem 1. The convergence rate of the sample autocorrelations to 1 depends on the magnitude of $\lambda\kappa$. When $\lambda\kappa$ is small, the sample autocorrelations may not be near 1, but the slowly decreasing pattern always exists. This pattern of SACF usually leads to differencing the series. Applying $\nabla = (1 - B)$ to both sides of Equation (2.1) yields

$$w_t = j_t\eta_t + \nabla x_t, \tag{3.2}$$

where $w_t = \nabla y_t$. The differenced series $\{w_t\}$ is the sum of a stationary ARMA process $\{\nabla x_t\}$ and a stationary process $\{j_t\eta_t\}$; hence it is stationary but non-Gaussian. Let $\rho_w(l)$ be the autocorrelation function (ACF) of series $\{w_t\}$. It is easy to see that

$$\rho_w(l) = c\rho_{\nabla x}(l), \qquad l = 1, 2, \ldots, \tag{3.3}$$

where $c = (1 + \lambda\kappa\sigma^2/\mathrm{var}(\nabla x_t))^{-1}$. This shows that the ACF of a differenced RLARMA series, $\{w_t\}$, will have the same pattern as that of a stationary ARMA process, $\{\nabla x_t\}$, except for the initial dropoff from $\rho(0)$ to $\rho(1)$. Hence, following the Box–Jenkins iterative-modeling procedure, an RLARMA process may easily be misspecified as an ARIMA process.

As an illustration, series A with 150 observations is generated by the model

$$y_t = \mu_t + x_t; \qquad \mu_t = \mu_{t-1} + j_t\eta_t; \qquad x_t = a_t + .6a_{t-1}; \tag{3.4}$$

a_t, j_t, and η_t are as defined in (2.1); $\lambda = .02$; $\kappa = 9$; $\sigma^2 = 1$; and the initial level μ_1 is set to be $\mu_1 = 3$. Figure 2 shows the plot of the series, exhibiting two level shifts at $t_1 = 48$ and $t_2 = 113$. The SACF of the original and the differenced series are given in Table 4. The SACF of the original series dies out slowly, and that of the differenced series gives insignificant values after lag 2. With the preceding patterns of SACF's, one would tentatively suggest an ARIMA(0, 1, 2) model for the original series. The estimated model for series A is

$$(1 - B)y_t = (1 - \underset{(.08)}{.35}B - \underset{(.09)}{.33}B^2)a_t, \qquad \hat{\sigma} = 1.03, \tag{3.5}$$

where the values in parentheses are the associated standard errors.

Portmanteau tests on the SACF of the residuals does not show evidence of inadequacy of ARIMA(0, 1, 2).

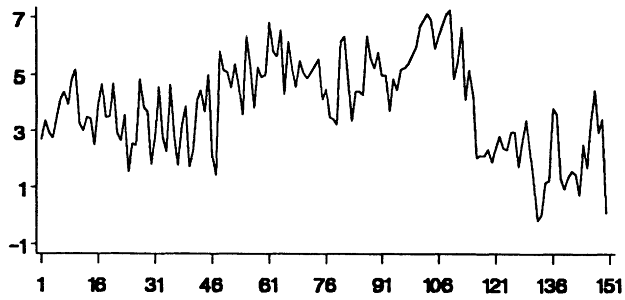

Figure 2. Time Series Plot of Series A.

It could be interpreted that the ARMA model is able to explain the behavior of the first two moments of a suitably differenced RLARMA process. It is then natural to ask (a) whether the RLARMA formulation provides any advantage over the ARIMA model for the purpose of data analysis and (b) whether, if one starts with the ARMA modeling, it is possible to discover the random level-shift ARMA structure.

4. PERFORMANCE OF THE ARIMA APPROXIMATION

Statistical models are built to explain phenomena that involve uncertainty. The choice of suitable models usually depends on the purpose of the practitioners. In Section 3, it was shown that a random level-shift ARMA model may often be misspecified as an ARIMA model. Inferential procedures—for instance, forecasting—based on this misspecified ARIMA model may be regarded as an approximation to those corresponding to the RLARMA process. In this section the efficiency of this approximation is investigated by comparing the MSE's of forecasts and those of estimating the current level of the process. We assume first that all of the parameters in (2.1) are known, but this assumption will be relaxed later.

4.1 Minimum Mean Squared Error Estimates of y_{T+l} and μ_T

Let $\mathbf{y}_T = (y_1, \ldots, y_T)'$ be the $T \times 1$ vector of observations of an RLARMA model (2.1) and $\mathbf{J} = (j_1, \ldots, j_T)'$ be the $T \times 1$ vector of indicator variables for level shifts. In practice, $\mathbf{J}$ is usually unknown. For estimating (forecasting) the future observations y_{T+l} ($l = 1, 2, \ldots$) and the current level μ_T based on $\mathbf{y}_T$, the minimum MSE (MMSE) estimates are the conditional expectations $E(y_{T+l} \mid \mathbf{y}_T)$ and $E(\mu_T \mid \mathbf{y}_T)$. We shall denote the errors as

$$\begin{aligned} e(l) &= \mu_T - E(\mu_T \mid \mathbf{y}_T), & l &= 0 \\ &= y_{T+l} - E(y_{T+l} \mid \mathbf{y}_T), & l &> 0, \end{aligned} \tag{4.1}$$

Table 4. Sample ACF of Series A and $\{\nabla y_t\}$

	l																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\rho_y(l)$	.72	.60	.60	.53	.50	.53	.55	.47	.42	.44	.39	.32	.27	.22	.21	.22	.20	.17
$\rho\nabla_y(l)$	−.24	−.26	.17	−.11	−.11	.01	.16	−.03	−.15	.13	.03	−.04	.04	−.06	−.08	.06	.00	.08

and the corresponding MSE of $e(l)$ as

$$\mathrm{MSE}(l) = Ee(l)^2, \tag{4.2}$$

where the expectation is taken over the joint distribution of $(\mu_T, y_{T+l}, \mathbf{y}_T)$. When $l > 0$, $e(l)$ is the lead-l forecast error and $e(0)$ is the error of estimating the current level. Now it will be useful to decompose $e(l)$ as the sum

$$e(l) = e_1(l) + e_0(l), \tag{4.3}$$

where

$$\begin{aligned} e_1(l) &= \mu_T - E(\mu_T \mid \mathbf{y}_T, \mathbf{J}), && l = 0 \\ &= y_{T+l} - E(y_{T+l} \mid \mathbf{y}_T, \mathbf{J}), && l > 0, \\ e_0(l) &= E(\mu_T \mid \mathbf{y}_T, \mathbf{J}) - E(\mu_T \mid \mathbf{y}_T), && l = 0 \\ &= E(y_{T+l} \mid \mathbf{y}_T, \mathbf{J}) - E(y_{T+l} \mid \mathbf{y}_T), && l > 0, \end{aligned}$$

and $E(\cdot \mid \mathbf{y}_T, \mathbf{J})$ are expectations conditional on $(\mathbf{y}_T, \mathbf{J})$. Using the fact that

$$\begin{aligned} E[e_1(l)e_0(l)] &= E\{E[e_1(l)e_0(l) \mid \mathbf{y}_T, \mathbf{J}]\} \\ &= E\{e_0(l)E[e_1(l) \mid \mathbf{y}_T, \mathbf{J}]\} = 0, \end{aligned}$$

we can then write $\mathrm{MSE}(l)$ as

$$\mathrm{MSE}(l) = V_1(l) + V_0(l), \qquad l \geq 0, \tag{4.4}$$

where $V_1(l) = \sum_{\mathbf{J}} p(\mathbf{J})E[e_1^2(l) \mid \mathbf{J}]$ and $V_0(l) = E[e_0^2(l)]$. Note that in (4.4) the expectation $E[e_1^2(l) \mid \mathbf{J}]$ is taken over the joint distribution of $(\mu_T, y_{T+l}, \mathbf{y}_T)$, conditional on $\mathbf{J}$, and $p(\mathbf{J})$ is the distribution of $\mathbf{J}$. On the other hand, the expectation $E[e_0^2(l)]$ is taken with respect to the joint distribution of $(\mathbf{y}_T, \mathbf{J})$. As will be shown in Appendix A, where explicit expressions of $V_1(l)$ and $V_0(l)$ are derived, the main reason for the decomposition in (4.3) and (4.4) is that $E[e_1^2(l) \mid \mathbf{J}]$ can be readily calculated, but determination of $E[e_0^2(l)]$ is considerably more burdensome. Note that, since $V_0(l) \geq 0$, we can regard $V_1(l)$ as a lower bound for $\mathrm{MSE}(l)$. Therefore, in our comparison of $\mathrm{MSE}(l)$ with the MSE from ARIMA approximations to be presented later in Section 4.3, we shall mainly use simulation estimates of the lower bound $V_1(l)$. In a few cases, simulation estimates of $V_0(l)$ will also be given to illustrate the magnitude of $V_0(l)$.

4.2 The MSE of the ARIMA Approximation

From the discussion in Section 3, we see that the RLARMA model (2.1) may be misspecified as a Gaussian ARIMA model following standard model-building procedures. The reason is that these procedures assume normality of the observations and consider mainly the covariance structure of $\{y_t\}$ and that of the differenced series $\{w_t\}$. To see this misspecification clearly, note that in (2.1) the level component μ_t is such that for $t \geq 2$, $\mu_t = \mu_{t-1} + j_t\eta_t$, where the $j_t\eta_t$'s are independent and have zero means and common variance $\kappa\lambda\sigma^2$. If we assume instead that

$$\mu_t = \eta_t^* + \mu_{t-1}, \qquad t \geq 2, \tag{4.5}$$

where $\{\eta_t^*\}$ are iid with zero means, common variance $\kappa\lambda\sigma^2$, and Gaussian, then y_t will follow a standard Gaussian ARIMA model. The covariance structure of the vector series $\{y_t, \mu_t\}$ corresponding to the model in (2.1) and that corresponding to the replacement (4.5) are exactly the same, however. Thus the misspecification of an RLARMA model by an ARIMA model concerns only the distributional differences between $\{\eta_t^*\}$ in (4.5) and $\{j_t\eta_t\}$ in (2.1). Now with the replacement (4.5), minimum MSE estimates of y_{T+l} and μ_T based on $\mathbf{y}_T$ will be linear in $\mathbf{y}_T$. We can alternatively regard our comparison of the efficiency of forecasting y_{T+l} and estimating μ_T between the RLARMA model and the ARIMA approximation as a study of the loss in the efficiency of linear estimation.

Specifically, for estimating (forecasting) y_{T+l} based on $\mathbf{y}_T$, note first that from (2.1) and (4.5), the overall ARIMA model for y_t takes the form

$$\phi(B)(1 - B)y_t = \alpha(B)c_t, \tag{4.6}$$

where $\alpha(B) = (1 - \alpha_1 B - \cdots - \alpha_{q^*}B^{q^*})$, $q^* = \max(p, q + 1)$, $\{c_t\}$ are iid $N(0, \sigma_c^2)$, and σ_c^2 and $\alpha(B)$ can be obtained from the relation

$$\frac{\alpha(B)\alpha(F)\sigma_c^2}{\phi(B)\phi(F)\sigma^2} = \kappa\lambda + (1 - B)(1 - F)\frac{\theta(B)\theta(F)}{\phi(B)\phi(F)}, \tag{4.7}$$

where $F = B^{-1}$ (e.g., see Cleveland and Tiao 1976; Hillmer and Tiao 1982). In particular, for $\lambda > 0$, $\alpha(B) \neq 0$ for $|B| = 1$. For large T, the results of Box and Jenkins (1976) can be used to obtain the optimal forecast for y_{T+l}, $\hat{y}_T(l)$, whose MSE is

$$\begin{aligned} \mathrm{MSE}_{\mathrm{ARMA}}(l) &= E(y_{T+l} - \hat{y}_T(l))^2 \\ &= \sigma_c^2 \sum_{s=0}^{l-1} \overline{\psi}_s^2, \qquad \overline{\psi}_0 = 1, \qquad l > 0, \end{aligned} \tag{4.8}$$

where the $\overline{\psi}_i$'s satisfy the relation $\phi(B)(1 - B)(1 + \overline{\psi}_1 B + \overline{\psi}_2 B^2 + \cdots) = \alpha(B)$.

Next, for estimating the current level μ_t from (2.1) and (4.5), we may employ the signal-extraction theory for a nonstationary ARIMA model given, for example, by Box, Hillmer, and Tiao (1976). Substituting the future y's with the forecast values based on (4.6) and $\{y_1, \ldots, y_T\}$, it can be shown that the best linear estimate of current level is

$$\hat{\mu}_T = \frac{\alpha(1)\phi(B)}{\phi(1)\alpha(B)}\, y_T. \tag{4.9}$$

As shown by Beveridge and Nelson (1981), the right side is, in fact, the limit $\lim_{l \to \infty} \hat{y}_T(l)$, and they have regarded this as an estimate of the permanent level of y_T. It is shown in Appendix B that the MSE of $\hat{\mu}_T$, denoted as $\mathrm{MSE}_{\mathrm{ARMA}}(0)$, is

$$\begin{aligned} \mathrm{MSE}_{\mathrm{ARMA}}(0) &= E(\mu_T - \hat{\mu}_T)^2 \\ &= \sigma_c^2 \left[\frac{\alpha(1)}{\phi(1)}\right]^2 \left[\frac{\phi'(1)}{\phi(1)} - \frac{\alpha'(1)}{\alpha(1)}\right], \end{aligned} \tag{4.10}$$

where $\phi'(1)$ and $\alpha'(1)$ are, respectively, the first derivatives of $\phi(B)$ and $\alpha(B)$ with respect to B evaluated at $B = 1$. The results in (4.8) and (4.10) can now be used to compare with the MSE(l) in (4.4).

Note that in the Gaussian ARIMA framework, decomposition of an overall model for y_t into a nonstationary component model for μ_t and a stationary component model for x_t is, in general, not unique (e.g., see Tiao and Hillmer 1978); also the estimation of μ_t will depend on the component model assumed. For the present problem, we are investigating the efficiency of the ARIMA approximation to the RLARMA model defined in (2.1); here the random-walk formulation (4.5) for μ_t seems to be the natural one to use, since, as mentioned earlier, it leads to an identical covariance structure of (y_t, μ_t) for both the RLARMA and the ARIMA models. One may raise the broader issue of the uniqueness of the RLARMA component structure, but this topic is beyond the scope of this article. The uniqueness issue discussed here only affects estimation of current level and has no impact on forecasting.

4.3 Efficiency of the ARIMA Approximation to the RLARMA Model

To illustrate the loss in efficiency of the ARIMA approximation to the RLARMA model, we consider the following simple model:

$$y_t = \mu_t + \frac{a_t}{(1 - \phi B)}, \qquad \mu_1 = 0$$

$$\mu_t = \mu_{t-1} + j_t\eta_t, \qquad t = 2, 3, \ldots, \qquad (4.11)$$

where $\{a_t\}$, $\{j_t\}$, and $\{\eta_t\}$ are defined in (2.1). Chernoff and Zacks (1964) considered methods of estimating the current level for a special case of this model ($\phi = 0$). The purpose of our study is to investigate the performance of the ARIMA approximation to this RLARMA model in terms of both forecasting and current-level estimation. Let $w_t = \nabla y_t$; it is straightforward to see that

$$\operatorname{var}(w_t) = \left(\lambda\kappa + \frac{2}{(1 + \phi)}\right)\sigma^2$$

$$\rho_w(1) = \frac{\phi - 1}{2 + \lambda\kappa(1 + \phi)}, \qquad (4.12)$$

and $\rho_w(l) = \phi\rho_w(l - 1)$ for $l \geq 2$. From the ACF of w_t, the ARIMA approximation to $\{y_t\}$ would be the model

$$(1 - \phi B)\nabla y_t = (1 - \alpha B)c_t, \qquad (4.13)$$

where α is a constant parameter and $\{c_t\}$ would be regarded as iid $N(0, \sigma_c^2)$. The quantities α and σ_c^2 can be determined in terms of the parameters in (4.11) by

$$\alpha^2 - m\alpha + 1 = 0, \qquad (1 + \phi\kappa\lambda)/\alpha = \sigma_c^2/\sigma^2, \qquad (4.14)$$

where $m = ((1 + \phi^2)\lambda\kappa + 2)/(1 + \phi\lambda\kappa)$ and $0 < \alpha \leq 1$.

The MSE of the ARIMA approximation $\mathrm{MSE}_{\mathrm{ARMA}}(l)$ can be easily obtained from (4.8) and (4.10). From (4.4) and applying the weak law of large numbers, a consistent estimate of $V_1(l)$ can be obtained by calculating $\sum E[e_1^2(l) \mid \dot{\mathbf{J}}]/N_1$, where the summation is taken over a random sample of $\dot{\mathbf{J}}$'s from the distribution (A.1) and $E[e_1^2(l) \mid \dot{\mathbf{J}}]$ is obtained for each $\dot{\mathbf{J}}$ from (A.8). The length of the observed series is set equal to 100; that is, $T = 100$. After some preliminary experimentation over various values of λ's, we choose the number of samples N_1 to be 1,000. An upper bound for the loss of efficiency by using the ARIMA forecasts is obtained by calculating the ratio

$$R(l) = \mathrm{MSE}_{\mathrm{ARMA}}(l)/V_1(l), \qquad l = 0, 1, \ldots. \qquad (4.15)$$

For forecasting comparisons, the $R(l)$'s have been calculated for $\lambda = (.01, .02, .05, .08, .10, .15)$, $\phi = (-.9, -.5, .0, .5, .9)$, $\kappa = (9.0, 25.0)$, and $l = 1, \ldots, 10$. Only partial results are reported here to show the pattern of relative forecasting performances. Specifically, Table 5 gives $R(l)$'s when $\phi = .0$ and $\kappa = 25.0$ for various values of λ and l, and Figure 3 shows plots of $R(l)$ ($l = 1, \ldots, 10$) for $\lambda = (.01, .05, .15)$, $\phi = (-.9, -.5, .0, .5, .9)$, and $\kappa = 25$. The loss in MSE's by using the ARIMA approximation varies with the frequency of level shift λ, the autoregressive coefficient ϕ of the noise term, and the lead time l. In general, the loss in efficiency decreases as λ is increased, and it becomes very small for large l. For short-term forecasts (low values of l), the potential maximum loss by using the ARIMA approximation is about 25% when $\phi \geq 0$. Although this loss can be substantially higher for $l = 1$ when $\phi < 0$, it declines rather precipitously for $l \geq 1$. We remark that for $\kappa = 9$, the same pattern of $R(l)$ is found and, as would be expected, the loss is uniformly smaller than the cases reported here. For instance, for ($\phi = .0$, $\lambda = .01$), $R(1)$ is 1.173 for $\kappa = 9$ versus 1.267 for $\kappa = 25$; for ($\phi = -.5$, $\lambda = .01$), it is 1.341 for $\kappa = 9$ versus 1.487 for $\kappa = 25$; for ($\phi = .5$, $\lambda = .01$), it is 1.087 for $\kappa = 9$ versus 1.105 for $\kappa = 25$. In sum, our results suggest that ARIMA approximation performs quite well for forecasting except when the frequency of shift λ is very small and/or the autoregressive coefficient ϕ has a large negative value.

Accuracy of the Upper Bound R(l) and Current Level Estimation Loss. From (4.4), the exact loss of efficiency is

$$\mathrm{Eff}(l) = \mathrm{MSE}_{\mathrm{ARMA}}(l)/(V_1(l) + V_0(l))$$
$$= R(l) \times \frac{V_1(l)}{V_1(l) + V_0(l)}. \qquad (4.16)$$

To assess the sharpness of the upper bound $R(l)$ of the exact loss, evaluation of $V_0(l)$ is required. From (A.13), $V_0(l)$ may be estimated by the average of $e_0^2(l)$ over, say, N_1 random drawings of $(\mathbf{J}, \mathbf{y}_T)$ from the joint distribution $P(\mathbf{J})P(\mathbf{y}_T \mid \mathbf{J})$ given in (A.1) and (A.11). Now,

Table 5. $R(l)$ for ϕ = .0 and κ = 25 as a Function of λ and l

	λ					
l	.01	.02	.05	.08	.10	.15
1	1.267	1.256	1.202	1.154	1.135	1.096
	(.0085)	(.0072)	(.0045)	(.0043)	(.0040)	(.0039)
2	1.224	1.195	1.133	1.095	1.081	1.055
	(.0076)	(.0063)	(.0042)	(.0040)	(.0039)	(.0039)
3	1.192	1.157	1.099	1.069	1.058	1.039
	(.0069)	(.0057)	(.0041)	(.0039)	(.0039)	(.0039)
4	1.169	1.132	1.079	1.054	1.045	1.030
	(.0064)	(.0053)	(.0040)	(.0039)	(.0039)	(.0039)
5	1.150	1.114	1.066	1.044	1.037	1.024
	(.0060)	(.0050)	(.0039)	(.0039)	(.0039)	(.0039)
6	1.136	1.100	1.056	1.038	1.031	1.020
	(.0057)	(.0048)	(.0039)	(.0039)	(.0039)	(.0039)
7	1.124	1.089	1.049	1.033	1.027	1.018
	(.0054)	(.0046)	(.0039)	(.0039)	(.0039)	(.0038)
8	1.113	1.080	1.044	1.029	1.024	1.016
	(.0052)	(.0045)	(.0039)	(.0039)	(.0039)	(.0038)
9	1.015	1.073	1.039	1.026	1.021	1.014
	(.0051)	(.0044)	(.0039)	(.0039)	(.0039)	(.0038)
10	1.097	1.067	1.036	1.023	1.019	1.013
	(.0050)	(.0043)	(.0039)	(.0039)	(.0038)	(.0038)

NOTE: Numbers in parentheses are the standard deviations of the estimates, calculated by the jackknife method.

to obtain $e_0(l)$, we see from (A.12) and (A.13) that computational complexity arises due to the need to calculate $\Sigma_{\mathbf{J}} P(\dot{\mathbf{J}} \mid \mathbf{y}_T)\mathbf{1}'_{\dot{m}}\hat{\boldsymbol{\eta}}_{\dot{m}}$ for $l = 0$ and $\Sigma_{\mathbf{J}} P(\dot{\mathbf{J}} \mid \mathbf{y}_T)(\mathbf{1}_{\dot{m}} - W'_{\dot{\mathbf{J}}}\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_{\dot{m}}$ for $l > 0$. To approximate these two summations, for each realization of the $(\mathbf{J}, \mathbf{y}_T)$ vector, we form weighted averages of $\mathbf{1}'_{\dot{m}}\hat{\boldsymbol{\eta}}_{\dot{m}}$ and $(\mathbf{1}_{\dot{m}} - W'_{\dot{\mathbf{J}}}\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_{\dot{m}}$ over a set S of $\dot{\mathbf{J}}$'s instead of all the possible $\dot{\mathbf{J}}$'s. Since our main interest is to investigate the situation for small values of λ ($\lambda \le .05$), we have adopted the following rule in constructing the set S: S is the union of three subsets, S_1, S_2, and S_3. S_1 consists of 1,000 $\mathbf{J}$'s randomly generated from distribution (A.1), which represents a typical collection of $\dot{\mathbf{J}}$'s. S_2 consists of $T - 1$ points of a single level shift, $S_2 = \{\mathbf{J} \mid \mathbf{J}'\mathbf{J} = 1 \text{ and } j_1 = 0\}$. The choice of this reflects our interest in situations of low λ. The set of S_3 is chosen as follows. For any $\mathbf{J}$, let $t^{(1)} < \cdots < t^{(m)}$ represent the positions of the m change points. Define the k neighborhood of $\mathbf{J}$, $NB(\mathbf{J}; k)$, as a set of $\dot{\mathbf{J}}$'s such that $\dot{\mathbf{J}}'\dot{\mathbf{J}} = m$ and the m change points of $\dot{\mathbf{J}}$, $\dot{t}^{(1)} < \cdots < \dot{t}^{(m)}$, are such that $|t^{(i)} - \dot{t}^{(i)}| \le k$ ($i = 1, \ldots, m$). For any realized $\mathbf{J}$, if $m \le 4$, S_3 is chosen as $NB(\mathbf{J}; 2)$ and when $m > 4$, S_3 consists of a random sample of 500 $\mathbf{J}$'s from $NB(\mathbf{J}; 2)$. The main reason for this choice of S_3 is that when the generated $\mathbf{J}$ consists of $m > 0$ change points, the conditional distribution $p(\dot{\mathbf{J}} \mid \mathbf{y}_T)$ tends to have most of its probabilities concentrated in a small neighborhood of $\mathbf{J}$. Once the S is obtained, the required conditional expectations are approximated by calculating $W \Sigma_s p(\dot{\mathbf{J}})p(\mathbf{y}_T \mid \dot{\mathbf{J}})\mathbf{1}'_{\dot{m}}\hat{\boldsymbol{\eta}}_{\dot{m}}$ and $W \Sigma_s p(\dot{\mathbf{J}})p(\mathbf{y}_T \mid \dot{\mathbf{J}})(\mathbf{1}_{\dot{m}} - W'_{\dot{\mathbf{J}}}\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_{\dot{m}}$, where $W^{-1} = \Sigma_s p(\dot{\mathbf{J}})p(\mathbf{y}_T \mid \dot{\mathbf{J}})$.

Based on this S, $e^2(l)$ may be estimated for any given $\mathbf{y}_T$ and $\mathbf{J}$, and hence $V_0(l)$ may be estimated by taking the average of $e^2(l)$ over a sample of N_1 drawings of $(\mathbf{J}, \mathbf{y}_T)$. Again we choose $N_1 = 1{,}000$ and $T = 100$. Table 6 gives results of the ratio $V_1(l)/(V_0(l) + V_1(l))$ for κ

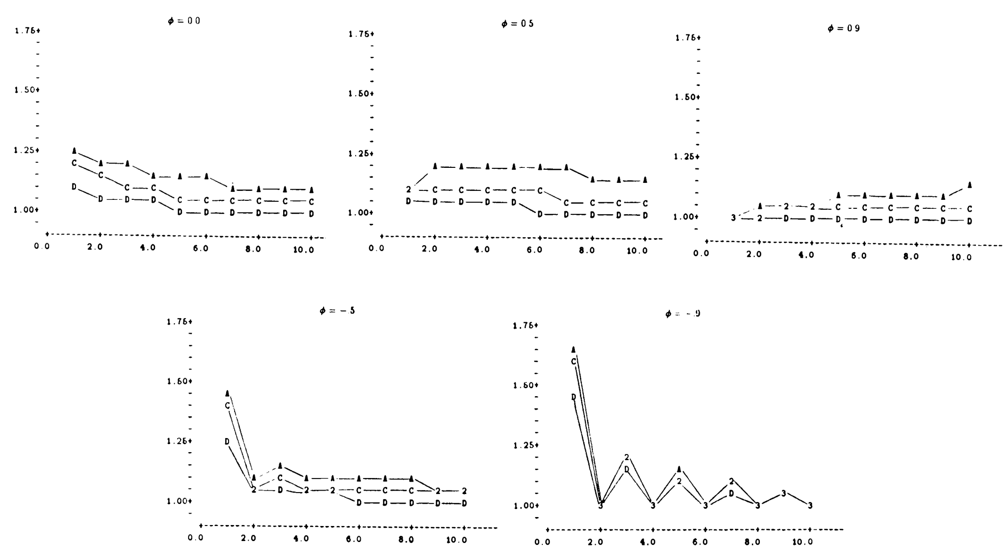

Figure 3. Plots of $R(l)$ When Parameters Are Known: κ = 25; A: λ = .01; C: λ = .05; D: λ = .15.

Table 6. $V_1(l)/(V_0(l) + V_1(l))$: $\kappa = 25$, $T = 100$

	$\phi = .0$			$\phi = -.5$			$\phi = .5$		
l	$\lambda = .01$	$\lambda = .02$	$\lambda = .05$	$\lambda = .01$	$\lambda = .02$	$\lambda = .05$	$\lambda = .01$	$\lambda = .02$	$\lambda = .05$
0	.3814	.5610	.7489	.5254	.5052	.4608	.3408	.4100	.6004
1	.9466	.9567	.9774	.9691	.9578	.9336	.9181	.9192	.9557
2	.9549	.9667	.9850	.9883	.9844	.9774	.9138	.9203	.9607
10	.9799	.9883	.9959	.9929	.9926	.9917	.9454	.9722	.9850

= 25 and selective values of (ϕ, λ, l). It shows that for the comparisons of forecasting performances the upper bound $R(l)$ ($l \geq 1$) will be close to the exact ratio Eff(l) for measuring the loss by using the ARIMA approximation. The entries for $l = 0$ in Table 6 imply that the upper bound $R(0)$ will substantially overestimate the loss in efficiency for estimating the current level μ. It is for this reason that the $R(0)$'s have not been included in Table 5. The estimated loss of efficiency Eff(0) is given in Table 7 for $\kappa = 25$ and selected values of (ϕ, λ). Although because of computational complexity we have only estimated Eff(0) for a much smaller number of combinations of values of (ϕ, λ), it is clear from Table 7 that, in contrast to the forecasting comparison shown in Table 5 and Figure 3, a very substantial gain in efficiency will be obtained by knowing the true model when estimating the current level. Moreover, this gain is higher when the noise term is negatively correlated and there is a general tendency for Eff(0) to decrease as λ is increased for fixed ϕ. Notice that the results are obtained from approximation and the accuracy is not high enough to establish the property that Eff(0) is uniformly decreasing in λ for fixed ϕ. The overall magnitude of Eff(0) is our major concern, however.

Discussion. When $\lambda = 1.0$, the model in (4.11) becomes exactly a nonstationary ARIMA(1, 1, 1) model. The findings reported in Table 3 and Figure 3 show the remarkable degree of robustness of the ARIMA(1, 1, 1) model to departure from $\lambda = 1.0$ for forecasting future observations. The ARIMA approximation, however, becomes inadequate when the probability λ of level change is very small or the series are negatively correlated.

On the other hand, Table 7 shows that the ARIMA approximation is clearly inadequate for estimating the current level of an RLARMA model. This finding suggests that for a component model of the form $y_t = \mu_t + x_t$, where x_t follows a stationary stochastic model, estimation of μ_t depends rather critically on the true distributional structure of μ_t. In particular, a Gaussian random-walk type of approximation for μ_t such as that given in (4.5) will lead to very inefficient results if in fact μ_t only changes occasionally.

Table 7. Eff(0) for $\kappa = 25$ and $T = 100$

	λ		
ϕ	.01	.02	.05
−.5	5.013	4.094	2.618
.0	3.303	3.043	2.935
.5	1.998	1.847	1.940

Now nonstationary models of nearly the forms $(1 - B)^d\mu_t = \eta_t^*$ for $d = 1, 2$ or $(1 - B^s)\mu_t = \eta_t^*$ for $s = 4$, 12 have been employed to represent the underlying stochastic structures for the trend and seasonal components of commonly used seasonal-adjustment procedures such as the Census X-11 method (e.g., see Burridge and Wallis 1984; Cleveland and Tiao 1976), model-based approach (Hillmer and Tiao 1982), and Kalman-filtering method (Engle 1976). Since the main interest in these adjustment procedures is in estimating the components rather than forecasting, our findings suggest that the procedures could be very sensitive to occasional level shifts in the series, and appropriate statistical analysis should be performed on the data to ascertain if such shifts indeed occur.

4.4 A Level-Shift Detection Procedure

A full analysis of the RLARMA process (2.1) would involve specification of the model for x_t and joint estimation of the parameters (λ, κ) for the level process μ_t, as well as those for the x_t process. For series exhibiting occasional level shifts, accurate estimates of (λ, κ) would require a very large sample. The simulation results in Section 4.3 suggest that, at least for forecasting purposes, one needs only be concerned with the situation in which λ is very small or x_t is highly negatively autocorrelated, but in these cases the sample would contain much information concerning the time points and the magnitudes of possible level changes. The reasons are: (a) when λ is very small, the changes are infrequent so that there will tend to be a large number of observations available between changes, and (b) it is well known that negative autocorrelations increase the efficiency in estimating the mean of observations. Now for series of moderate length with very infrequent level shifts of unknown magnitude at unknown time points, there is little to choose between a random-effect formulation such as that given by the RLARMA model in (2.1) and a fixed-effect formulation for practical purposes. A potentially useful modeling procedure to improve forecast performance might consist of (a) constructing an ARIMA model for the observed series $\{y_t\}$, (b) detecting the possible level shifts (i.e., estimating $\mathbf{J}$ by appropriate analysis of the residuals, and (c) remodeling the data and making forecasts based on the estimated information of $\mathbf{J}$.

Specifically, let $\mathbf{y}_T = (y_1, \ldots, y_T)'$ be a set of observations and this series be empirically described by a general ARMA model $\varphi(B)y_t = \theta(B)a_t$. Let $\hat{\mathbf{a}}_T = (\hat{a}_1, \ldots, \hat{a}_T)'$ be the estimated residuals of the ARMA fitting to $\mathbf{y}_T$. Based on the fitted model and the residuals, we can define the following statistics:

$$\hat{\tau}_i = \hat{\mathbf{a}}_T'\Pi\mathbf{L}(i)/\sqrt{\hat{\sigma}^2(\Pi\mathbf{L}(i))'\Pi\mathbf{L}(i)}, \qquad (4.17)$$

where $\hat{\sigma}^2$ is the residual variance; $\mathbf{L}(i) = (LS_1(i), \ldots, LS_T(i))'$ is a $T \times 1$ vector to indicate the occurrence of a level shift at $t = i$; and Π is a $T \times T$ lower triangular matrix with elements π_{ij} such that $\pi_{ij} = 0$ if $j > i$, $\pi_{ij} = 1$ if $j = i$, and otherwise $\pi_{ij} = -\pi_{i-j}$, where $(1 - \pi_1 B - \pi_2 B^2 - \cdots) = \varphi(B)\theta(B)^{-1}$. Notice that $\hat{\tau}_i$ is the likelihood ratio statistic for testing the hypothesis of no level shift versus that of a fixed level shift at $t = i$. Chen (1984) studied the properties and the performance of a level-shift detection procedure based on

$$\tau_{\max} > C, \qquad (4.18)$$

where $\tau_{\max} = \max|\tau_i|$ and C is a positive constant. Typically, we choose $C = 2.8$ for high sensitivity level-shift detection, $C = 3.0$ for moderate sensitivity, and $C = 3.3$ for low sensitivity. It can be shown that the statistic $\hat{\tau}_i$ is measuring difference of levels before and after time period i. Consequently, the power of this statistic will be less when i is near to either the beginning or the ending of the sample. Sampling properties of this statistic and its relationship with the RLARMA model will be discussed in another article. In practice, a more efficient approach is to apply the method in an iterative fashion. Once a level shift is identified, the residuals are adjusted by removing the effect of this level shift. Then a new set of τ_i statistics as defined in (4.17) may be computed based on the adjusted residuals. The iteration stops when $\tau_{\max}$ is smaller than or equal to C.

In Section 1.1, the statistic (4.18) was applied to the filtered variety-store sales data. To further illustrate the potential usefulness of the method, we apply the iterative procedure as described with $C = 2.8$ to series A generated by the RLARMA model in (3.4). The procedure successfully identifies two level shifts at $t_1 = 48$ and $t_2 = 113$. To incorporate this information in the model-building process, we consider the fixed-effect model with level shifts at known time points $t_1 = 48$ and $t_2 = 113$ in (4.19) as an alternative to Model (3.4) and obtain the following estimates (numbers in parentheses are standard errors of the estimates):

$$\begin{aligned} y_t &= \omega_1 LS_{1t}(48) + \omega_2 LS_{2t}(11) \\ &\quad + (1 - \theta_1 B - \theta_2 B^2)/(1 - B)a_t \\ &= \underset{(.34)}{-3.18}LS_{1t}(48) + \underset{(.32)}{2.10}LS_{2t}(113) \\ &\quad + (1 - \underset{(.08)}{.51}B - \underset{(.08)}{.50}B^2)/(1 - B)a_t, \quad \hat{\sigma} = .994. \end{aligned} \qquad (4.19)$$

Examining the estimated ARMA coefficients, there is a near cancellation between the MA and the AR parts. This leads us to consider a more parsimonious model,

$$\begin{aligned} y_t &= \omega_0 + \omega_1 LS_{1t}(48) + \omega_2 LS_{2t}(113) + (1 - \theta_1^* B)a_t \\ &= \underset{(.21)}{3.31} - \underset{(.30)}{3.18}LS_{1t}(48) + \underset{(.28)}{2.10}LS_{2t}(113) \\ &\quad + (1 + \underset{(.07)}{.49}B)a_t, \qquad \hat{\sigma} = .993. \end{aligned} \qquad (4.20)$$

Note that the fitted noise structure $(1 + .49B)a_t$ is close to the structure of x_t in (3.4). This example shows that the procedure not only detects the level shifts but also helps identify partially the true underlying structure of the process.

4.5 A Study of ARMA Approximation to RLARMA When the Parameters Are Estimated

The results of Section 4.3 are obtained assuming that the parameters of the RLARMA model (4.11) and those of its ARIMA approximation (4.13) are known. In reality, the model and the parameters are determined from the nature of the data. The following Monte Carlo experiment is designed to investigate, in finite-sample situations, the relative performance of ARIMA model building with and without incorporating the level-shift detection procedure (4.17) and (4.18) when the data are generated from RLARMA models.

Step 1. For given values of λ, κ, and ϕ, 200 observations of z_t are generated from an RLARMA (1, 0) model in (4.11).

Step 2. Standard Box–Jenkins techniques are applied to the first 190 observations of the simulated series to build an adequate ARIMA (or ARMA) model. To reduce the judgmental effect in the process of model building, however, three models are sequentially considered to fit the first 190 observations of the simulated series. They are AR(1), ARIMA(1, 1, 0), and ARIMA(1, 1, 1). We take the position that a simpler model is maintained unless there is strong evidence against it. Model adequacy is tested by the Box–Ljung Q statistics with 12 lags, Q_{12}, at the 1% level of significance. If both AR(1) and ARIMA(1, 1, 0) models are rejected, then the ARIMA(1, 1, 1) model will be fitted to the data.

Step 3. Lead-1 to lead-10 out of sample forecasts and the corresponding forecast errors are calculated at origin 190 based on the model specified in Step 2.

Step 4. The level-shift detection procedure (4.18) is applied in an iterative fashion to the residuals of the fitted results in Step 2. In this study, the critical value C is chosen as 2.8. When level shifts are identified, an intervention model incorporating level shifts with AR(1) noise will be employed to estimate the magnitudes of the shifts as well as the time series parameter in a manner similar to (4.20). Lead-1 to lead-10 out-sample forecasts are computed based on this modified model, assuming that the effect of level shift will persist to the future.

Step 5. Repeat Steps 1–4 500 times for each set of λ, κ, and ϕ. Mean squared forecast errors based on ARIMA (or ARMA) models and AR(1) models incorporated with the level-shift information in Step 4 are computed, respectively.

Results. The averages of the Q_{12}'s of the residuals-sample ACF over the 500 iterations for various combinations of λ and ϕ range from 7.52 to 13.83 for the final model in Step 2 and from 9.06 to 12.75 for model fittings in Step 4. On the average, the results do not indicate any serious inadequacy in model specification. Table 8 reports the average π weights of the fitted ARIMA (or ARMA) models for Step 2, as well as those from the intervention models for Step 4. The average π weights without adjusting the level shifts match fairly well with the theoretical ones calculated from (4.13) and (4.14), which also indicates that model specification has been conducted properly. We also found that, after adjusting for the occasional level shifts, the estimated π weights are very close to the true structure of the underlying process. Figure 4 presents the ratios of mean squared forecast errors from ARIMA (or ARMA) models over those from models incorporating level shifts. The overall patterns are largely consistent with those obtained in Figure 3, corresponding to the cases of known parameters. In most cases, the magnitudes of $R(l)$'s are about the same except for the case of (λ = .01, ϕ = .5), in which the improvement of forecast accuracy is even higher when the parameters are estimated. With critical value set to C = 2.8, about 75.4% of level shifts are correctly identified, and the average number of false level shifts found is 267 for each combination of λ and ϕ over the 500 iterations. To further illustrate the comparison, we compute the ratios of mean squared forecast errors when parameters are estimated to that when parameters are known. The results of ARIMA approximation and intervention model (RLARMA consideration) are reported in Figures 5 and 6, respectively. For the same ϕ, the magnitudes of MSE ratios in both approaches are about the same; this suggests that the loss of accuracy when the parameters

Table 8. Theoretical and Estimated π Weights: κ = 25,500 Iterations

	ϕ = .9		ϕ = .5		ϕ = .0		ϕ = −.5		ϕ = −.9	
i	$\hat{\pi}_i$	π_i	$\hat{\pi}_i$	π_i	$\hat{\pi}_i$	π_i	$\hat{\pi}_i$	π_i	$\hat{\pi}_i$	π_i
					ARIMA models					
λ = .01										
1	.959	.944	.742	.710	.388	.390	−.002	.042	−.332	−.256
	(.216)		(.272)		(.286)		(.316)		(.340)	
2	−.008	.003	.002	.061	.158	.238	.394	.519	.628	.809
	(.220)		(.166)		(.126)		(.219)		(.339)	
3	.012	.002	.030	.048	.085	.145	.142	.238	.167	.288
	(.091)		(.031)		(.066)		(.106)		(.144)	
4	−.004	.002	.022	.038	.050	.088	.068	.109	.071	.103
	(.088)		(.020)		(.041)		(.063)		(.077)	
λ = .02										
1	.972	.957	.760	.770	.475	.500	.145	.186	−.150	−.097
	(.158)		(.181)		(.195)		(.220)		(.237)	
2	−.002	.003	.054	.062	.221	.250	.497	.558	.797	.881
	(.154)		(.123)		(.095)		(.121)		(.164)	
3	.010	.002	.039	.045	.015	.125	.160	.175	.173	.173
	(.041)		(.030)		(.050)		(.090)		(.140)	
4	−.000	.002	.028	.033	.060	.063	.073	.055	.069	.034
	(.031)		(.017)		(.034)		(.056)		(.070)	
λ = .05										
1	.965	.974	.819	.853	.624	.656	.363	.394	.097	.129
	(.092)		(.092)		(.130)		(.158)		(.169)	
2	.011	.002	.079	.052	.225	.226	.517	.542	.862	.896
	(.089)		(.065)		(.083)		(.082)		(.061)	
3	.008	.002	.030	.034	.072	.078	.074	.058	.008	−.026
	(.039)		(.032)		(.047)		(.088)		(.136)	
4	−.000	.002	.019	.022	.034	.027	.027	.006	.023	.001
	(.029)		(.020)		(.025)		(.034)		(.036)	
					Intervention models					
λ = .01	.879	.9	.503	.5	.042	.0	−.497	−.5	−.894	−.9
	(.039)		(.079)		(.132)		(.064)		(.032)	
λ = .02	.869	.9	.471	.5	−.012	.0	−.498	−.5	−.893	−.9
	(.042)		(.064)		(.072)		(.064)		(.033)	
λ = .05	.848	.9	.438	.5	−.036	.0	−.507	−.5	−.893	−.9
	(.057)		(.076)		(.079)		(.067)		(.035)	

NOTE: π_i is the theoretical value and $\hat{\pi}_i$ is the mean estimate of it. Numbers in parentheses are the standard deviations of the estimates.

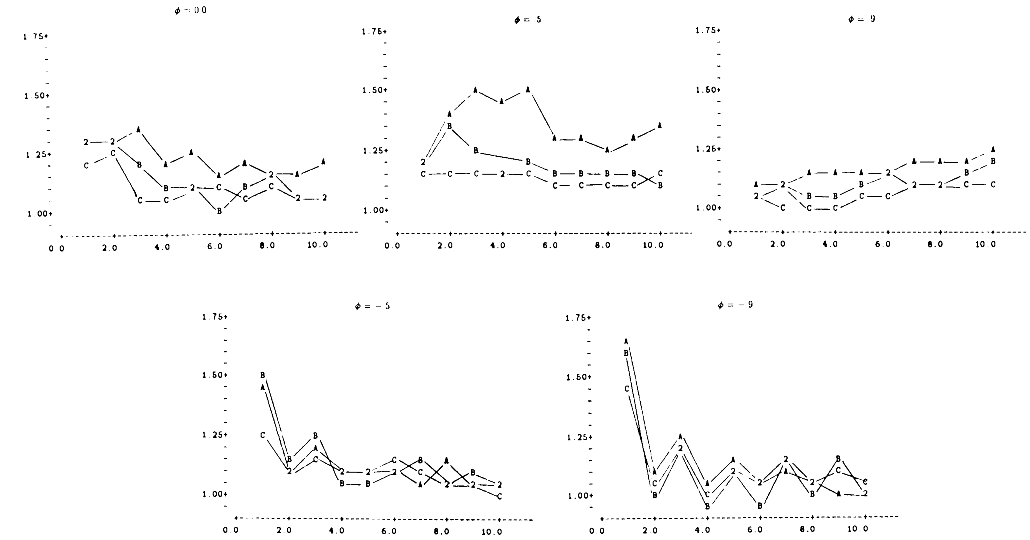

Figure 4. Plots of R(l) When Parameters Are Estimated: κ = 25; A: λ = .01; B: λ = .02; C: λ = .05.

are estimated is about the same in both approaches. There is no obvious pattern in Figures 5 and 6 with respect to λ. In both cases there is a slight increasing trend of MSE's with respect to the number of forecast leads, which seems to indicate that the short-run forecasts suffer less due to the uncertainty of estimated parameters. Further studies are called for to understand the performance of the proposed detection procedure and the sampling properties of the estimates in the intervention model.

5. GENERALIZATIONS

The idea of occasionally changing level in stationary process can be extended to the situations that the homogeneous component $\{x_t\}$ in (2.1) is nonstationary. Two potentially useful models are (a) the observed series is the sum of a level process $\{\mu_t\}$ and a nonstationary ARMA process N_t and (b) a suitably differenced series follows an RLARMA process. The first generalization can be defined explicitly as

$$y_t = \mu_t + N_t, \qquad t = 1, \ldots, T, \tag{5.1}$$

where $U_d(B)N_t = x_t$, $\{\mu_t\}$ and $\{x_t\}$ are as defined in (2.1), and $U_d(B)$ is a polynomial of degree d, having all of its zeros lying on the unit circle. In particular, if $\{N_t\}$ is a random-walk process, the model becomes a discretized version of the compound events models considered by Press (1967) and Clark (1973). That type of model has been a candidate for describing the behavior of security prices. The second possible generalization may be formulated as $U_d(B)y_t = \mu_t + x_t$ $(t = 1, \ldots, T)$ or

$$y_t = \frac{\mu_t}{U_d(B)} + \frac{x_t}{U_d(B)}, \tag{5.2}$$

where $\{x_t\}$ and $\{\mu_t\}$ are as defined in (2.1) and $U_d(B)$ is that given in (5.1). This class of model may be used to

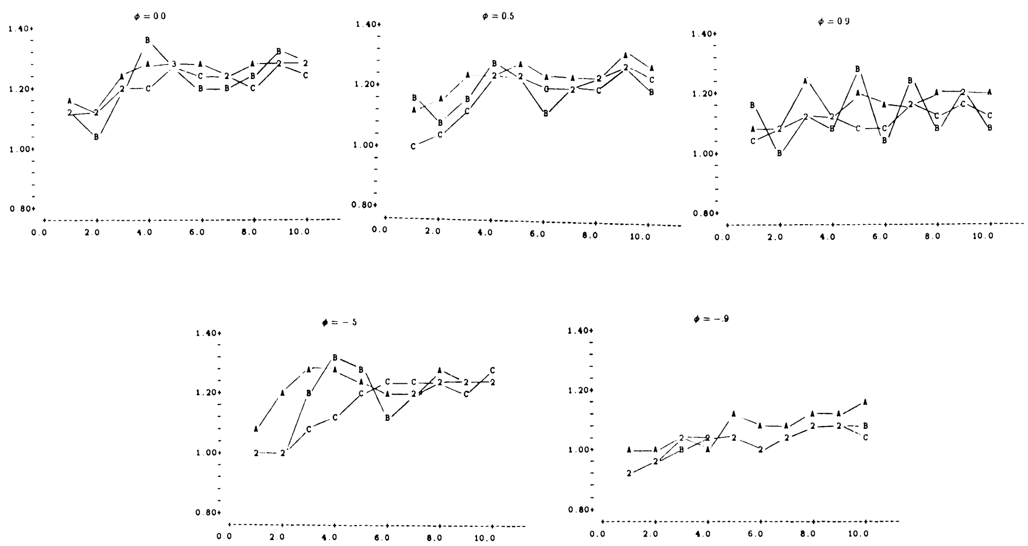

Figure 5. Ratios of MSE's: Estimated Parameters Versus Known Parameters, ARIMA Approximation: κ = 25; A: λ = .01; B: λ = .02; C: λ = .05.

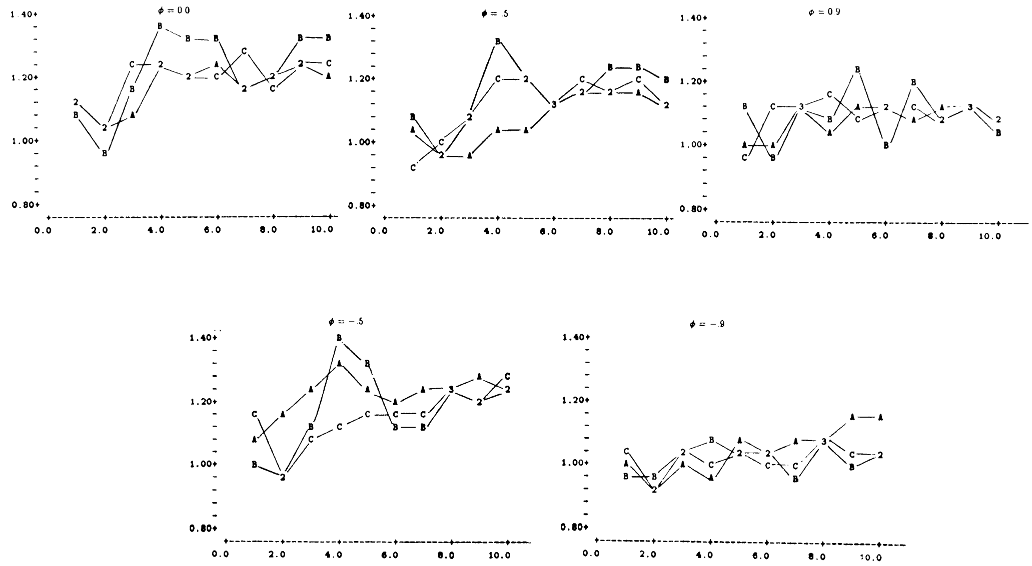

Figure 6. Ratios of MSE's: Estimated Parameters Versus Known Parameters, RLARMA Consideration: κ = 25; A: λ = .01; B: λ = .02; C: λ = .05.

represent series with various types of abrupt structural changes. For instance, if $U_d(B) = (1 - B)$ in (5.2), then $\{y_t\}$ may reveal occasional changes in trend. Notice that (5.1) and (5.2) are very different in terms of the effect of the changing μ_t. For example, if $U_d(B) = 1 - B$, then (5.1) implies y_t has random level shifts and nonstationary error, whereas (5.2) implies that y_t has random linear-trend shifts and nonstationary error. Moreover, $(1 - B)y_t$ has random point outliers and stationary error under (5.1) and random level shifts and stationary error under (5.2). Identifying these two models requires careful analysis of both the original series and suitably differenced data. In general, the preceding models can be employed to represent series with behavior of nonstationarity between d differencing and $(d + 1)$ differencing. It would be of interest to investigate the efficiency of ARIMA approximations to such models and for low frequency shift in μ_t (small values of λ) and to extend the level-shift detection procedure in (4.17) and (4.18) to cover these cases.

ACKNOWLEDGMENT

We thank the referees and an associate editor for their helpful comments.

APPENDIX A: EXPLICIT EXPRESSIONS OF $V_1(l)$ AND $V_0(l)$

We here give explicit expressions for $V_1(l)$ and $V_0(l)$ in (4.4). From (2.1), the distribution $p(\mathbf{J})$ is

$$p(\mathbf{J}) = \prod_{t=1}^{T} \lambda^{j_t}(1 - \lambda)^{1-j_t}. \tag{A.1}$$

For notational convenience, we now assume that the processes $\{\mu_t\}$ and $\{x_t\}$ start at $t = 1$. Specifically, we suppose that, for $t = 1, 2, \ldots,$

$$\mu_t = \sum_{s=1}^{t} j_s \eta_s, \qquad x_t = \sum_{s=0}^{t-1} \psi_s a_{t-s}, \quad \psi_0 = 1, \tag{A.2}$$

where the ψ_i's can be obtained from the relation $\phi(B)\psi(B) = \theta(B)$ with $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots)$. From (2.1) and (A.2), we can write

$$\begin{pmatrix} \mathbf{y}_T \\ \mu_T \\ y_{T+l} \end{pmatrix} = \begin{pmatrix} UD & \mathbf{0} \\ \mathbf{J}' & 0 \\ \mathbf{J}' & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta} \\ r_1 \end{pmatrix} + \begin{pmatrix} \Psi & \mathbf{0} \\ \mathbf{0}' & 0 \\ \boldsymbol{\psi}^{(l)\prime} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ r_2 \end{pmatrix}, \tag{A.3}$$

where $\mathbf{0}$ is a $T \times 1$ null vector, Ψ and U are $T \times T$ lower triangular matrices such that

$$\Psi = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \psi_1 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \psi_{T-1} & \ldots & \psi_1 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \ldots & 1 & 1 \end{pmatrix},$$

D is a $T \times T$ diagonal matrix with diagonal elements $(j_1, \ldots, j_T)$, $\boldsymbol{\psi}^{(l)} = (\psi_{T+l-1}, \ldots, \psi_l)'$, $r_1 = \sum_{s=1}^{l} j_{T+s}\eta_{T+s}$, and $r_2 = \sum_{s=0}^{l-1} \psi_s a_{T+l-s}$. Since $\{j_t\}$, $\{\eta_t\}$, and $\{a_t\}$ are independent, it follows that, conditional on $\mathbf{J}$, $(\mathbf{y}_T, \mu_T, y_{T+l})$ are jointly normal and their covariance matrix is

$$\operatorname{cov}(\mathbf{y}_T, \mu_T, y_{T+l} \mid \mathbf{J}) = \sigma^2 A, \tag{A.4}$$

where

$$A = \kappa \begin{pmatrix} UDU' & U\mathbf{J} & U\mathbf{J} \\ \mathbf{J}'U' & m & m \\ \mathbf{J}'U' & m & m + \lambda l \end{pmatrix} + \begin{pmatrix} \Psi\Psi' & \mathbf{0} & \Psi\boldsymbol{\psi}^{(l)} \\ \mathbf{0}' & 0 & 0 \\ \boldsymbol{\psi}^{(l)\prime}\Psi' & 0 & v_l \end{pmatrix},$$

$m = \mathbf{J}'\mathbf{J}$ is the number of level shifts, and $v_l = \sum_{i=0}^{T+l-1} \psi_i^2$. It follows that, for the $V_1(l)$ component in (4.4),

$$E[e_1^2(l) \mid \mathbf{J}] = \sigma^2 \times \{\kappa[m - \kappa(U\mathbf{J})'R(U\mathbf{J})],\ l = 0\}$$
$$= \sigma^2 \times \{\kappa(\lambda l + m) + v_l - (\kappa U\mathbf{J} + \Psi\boldsymbol{\psi}^{(l)})' \times R(\kappa U\mathbf{J} + \Psi\boldsymbol{\psi}^{(l)}),\ l > 0\}, \quad \text{(A.5)}$$

where $R = (\kappa UDU' + \Psi\Psi')^{-1}$.

For actual computation of (A.5), the main burden lies in evaluating the $T \times T$ matrix R. Note that the rank of UDU' is m, the number of nonzero values in $\mathbf{J}$ or the number of level shifts that occurred in the observational period. For small m, computation of R can be drastically simplified by employing the identity $(I_n + PQ)^{-1} = I_n - P(I_{n'} + QP)^{-1}Q$, where P is an $n \times n'$ and Q an $n' \times n$ matrix. Specifically, letting H be the $T \times T$ lower triangular matrix

$$H = \Psi^{-1}U = \begin{pmatrix} 1 & 0 & \dots & 0 \\ h_1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{T-1} & \dots & h_1 & 1 \end{pmatrix}$$
$$= (\mathbf{h}_1, \dots, \mathbf{h}_T), \quad \text{(A.6)}$$

where the elements h_i's can be obtained from the relation $\theta(B)(1 - B)H(B) = \phi(B)$ with $H(B) = (1 + h_1B + h_2B^2 + \cdots)$. If $t_1, \dots, t_m$ denote the positions where a level shift occurs, let $W_\mathbf{J} = (\mathbf{h}_{t_1}, \dots, \mathbf{h}_{t_m})$ be a $T \times m$ matrix such that $j_{t_1} + \cdots + j_{t_m} = m$. An alternative expression for R is

$$R = \Psi'^{-1}(I + \kappa HDH')^{-1}\Psi^{-1}$$
$$= \Psi'^{-1}(I + \kappa W_\mathbf{J}W_\mathbf{J}')^{-1}\Psi^{-1}. \quad \text{(A.7)}$$

Then Equation (A.5) may be simplified as

$$E[e_1^2(l) \mid \mathbf{J}] = \sigma^2 \times \{\kappa \mathbf{1}_m' M_\mathbf{J} \mathbf{1}_m,\ l = 0\}$$
$$= \sigma^2 \times \{\kappa\lambda l + \sum_{i=0}^{l-1} \psi_i^2 + \kappa(\mathbf{1}_m - W_\mathbf{J}'\boldsymbol{\psi}^{(l)})'M_\mathbf{J} \times (\mathbf{1}_m - W_\mathbf{J}'\boldsymbol{\psi}^{(l)}),\ l > 0\}, \quad \text{(A.8)}$$

where $\mathbf{1}_m$ is an $m \times 1$ vector of ones, $M_\mathbf{J} = (I + \kappa W_\mathbf{J}'W_\mathbf{J})^{-1}$ is an $m \times m$ matrix, and $\kappa M_\mathbf{J}W_\mathbf{J}'W_\mathbf{J} = I - M_\mathbf{J}$. In (A.8), $t_1, \dots, t_m$ are the positions at which a level shift occurs. Expressions (A.1) and (A.8) can now be employed to evaluate the component $V_1(l)$ in (4.4).

Turning now to the component $V_0(l)$ in (4.4), we first consider the terms of $e_0(l)$ in (4.3). From (A.3) and (A.4), it is readily shown that $E(\mu_T \mid \mathbf{y}_T, \mathbf{J}) = \mathbf{1}_m'\hat{\eta}_m$ and

$$E(y_{T+l} \mid \mathbf{y}_T, \mathbf{J}) = \boldsymbol{\psi}^{(l)\prime}\mathbf{r} + (\mathbf{1}_m - W_\mathbf{J}'\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_m, \quad \text{(A.9)}$$

where $\hat{\boldsymbol{\eta}}_m = \kappa M_\mathbf{J}W_\mathbf{J}'\mathbf{r}$, $\mathbf{r} = \Pi\mathbf{y}_T$, and $\Pi = \Psi^{-1}$ is a $T \times T$ lower triangular matrix. Note that

$$E(\mu_T \mid \mathbf{y}_T) = \sum_\mathbf{J} p(\mathbf{J} \mid \mathbf{y}_T)E(\mu_T \mid \mathbf{y}_T, \mathbf{J})$$
$$E(y_{T+l} \mid \mathbf{y}_T) = \sum_\mathbf{J} p(\mathbf{J} \mid \mathbf{y}_T)E(y_{T+l} \mid \mathbf{y}_T, \mathbf{J}), \quad \text{(A.10)}$$

where $p(\mathbf{J} \mid \mathbf{y}_T)$ is the conditional distribution of $\mathbf{J}$ given $\mathbf{y}_T$. Moreover, by Bayes's theorem, we find that

$$p(\mathbf{J} \mid \mathbf{y}_T) = p(\mathbf{J})p(\mathbf{y}_T \mid \mathbf{J}) \Big/ \sum_\mathbf{J} p(\mathbf{J})p(\mathbf{y}_T \mid \mathbf{J}), \quad \text{(A.11)}$$

where $p(\mathbf{y}_T \mid \mathbf{J}) = (2\pi)^{-T/2}|M_\mathbf{J}|^{1/2}\sigma^{-T}\exp\{-(1/2\sigma^2)(\mathbf{r}'\mathbf{r} - \mathbf{r}'W_\mathbf{J}\hat{\boldsymbol{\eta}}_m)\}$.

It follows that

$$e_0(l) = \mathbf{1}_m'\hat{\boldsymbol{\eta}}_m - \sum_{\dot{\mathbf{J}}} p(\dot{\mathbf{J}} \mid \mathbf{y}_T)\mathbf{1}_{\dot{m}}'\hat{\boldsymbol{\eta}}_{\dot{m}}, \qquad l = 0$$
$$= (\mathbf{1}_m - W_\mathbf{J}'\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_m - \sum_{\dot{\mathbf{J}}} p(\dot{\mathbf{J}} \mid \mathbf{y}_T)(\mathbf{1}_{\dot{m}} - W_{\dot{\mathbf{J}}}'\boldsymbol{\psi}^{(l)})'\hat{\boldsymbol{\eta}}_{\dot{m}}, \qquad l > 0, \quad \text{(A.12)}$$

and hence

$$V_0(l) = \int \sum_\mathbf{J} p(\mathbf{J})p(\mathbf{y}_T \mid \mathbf{J})e_0^2(l)\, d\mathbf{y}_T. \quad \text{(A.13)}$$

From (4.4), (A.1), (A.8), (A.12), and (A.13), it is clear that exact evaluation of either $V_1(l)$ or $V_0(l)$ will be extremely burdensome for large T, because there are 2^T elements in the sample space of $\mathbf{J}$. Since $E[e_1^2(l) \mid \mathbf{J}]$ is available from (A.8), however, estimates of $V_1(l)$ can be readily obtained by performing simulation experiments with respect to $\mathbf{J}$ only. Moreover, we see from (A.12) and (A.13) that considerably more complex simulation experiments will be required to estimate the component $V_0(l)$.

APPENDIX B: PROOF OF THE MSE$_{\text{ARMA}}$(0) IN (4.10)

From (2.1) and (4.9), we have that

$$\mu_T - \hat{\mu}_T = \left(1 - \omega\frac{\phi(B)}{\alpha(B)}\right)\mu_T - \omega\frac{\theta(B)}{\alpha(B)}a_T, \quad \text{(B.1)}$$

where $\omega = \alpha(1)/\phi(1)$. Since $\{\mu_t\}$ and $\{a_t\}$ are independent, the covariance generating function of $(\mu_T - \hat{\mu}_T)$ is

$$G(B, F) = \kappa\lambda\sigma^2(1 - B)^{-1}(1 - F)^{-1} \times \left(1 - \omega\frac{\phi(B)}{\alpha(B)}\right)\left(1 - \omega\frac{\phi(F)}{\alpha(F)}\right) + \omega^2\sigma^2\frac{\theta(B)\theta(F)}{\alpha(B)\alpha(F)}. \quad \text{(B.2)}$$

From (4.7), by setting $B = F = 1$, we obtain $\kappa\lambda\sigma^2 = \sigma_c^2\omega^2$, and hence in (B.2)

$$\sigma^2\frac{\theta(B)\theta(F)}{\alpha(B)\alpha(F)} = \sigma_c^2(1 - B)^{-1}(1 - F)^{-1}\left[1 - \omega^2\frac{\phi(B)\phi(F)}{\alpha(B)\alpha(F)}\right].$$

It follows that (B.2) can be written as

$$G(B, F) = \omega^2\sigma_c^2(1 - B)^{-1}(1 - F)^{-1} \times \left[2 - \omega\frac{\phi(B)}{\alpha(B)} - \omega\frac{\phi(F)}{\alpha(F)}\right]. \quad \text{(B.3)}$$

Let

$$\omega \frac{\phi(B)}{\alpha(B)} = \sum_{s=0}^{\infty} \beta_s B^s, \tag{B.4}$$

and note that $\sum_{s=0}^{\infty} \beta_s = \omega\phi(1)/\alpha(1) = 1$. Thus

$$\begin{aligned} G(B, F) &= \omega^2\sigma_c^2(1 - B)^{-1}(1 - F)^{-1} \\ &\quad \times \left[2(1 - \beta_0) - \sum_{s=1}^{\infty} \beta_s(B^s + F^s)\right] \\ &= \omega^2\sigma_c^2(1 - B)^{-1}(1 - F)^{-1} \\ &\quad \times \sum_{s=1}^{\infty} \beta_s(2 - B^s - F^s). \end{aligned} \tag{B.5}$$

Since

$$(1 - B)^{-1}(1 - F)^{-1}(2 - B^s - F^s) = F^{s-1}\left(\sum_{j=0}^{s-1} B^j\right)^2, \tag{B.6}$$

it follows that

$$G(B, F) = \omega^2\sigma_c^2 \sum_{s=1}^{\infty} \beta_s F^{s-1}\left(\sum_{j=0}^{s-1} B^j\right)^2. \tag{B.7}$$

The variance or $\mathrm{MSE}_{\mathrm{ARMA}}(0)$ of $(\mu_T - \hat{\mu}_T)$ is the constant term in $G(B, F)$, which is

$$\mathrm{MSE}_{\mathrm{ARMA}}(0) = \omega^2\sigma_c^2 \sum_{s=1}^{\infty} s\beta_s.$$

From (B.4), $\sum_{s=1}^{\infty} s\beta_s$ is the derivative of $\omega\phi(B)/\alpha(B)$ with respect to B evaluated at $B = 1$, and hence (4.10) follows.

[*Received November 1986. Revised May 1989.*]

REFERENCES

Bell, W. R., Hillmer, S. C., and Tiao, G. C. (1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," in *Applied Time Series Analysis of Economic Data* (Economic Research Report 5), ed. A. Zellner, Washington, DC: U.S. Bureau of the Census, pp. 74–100.

Beveridge, S., and Nelson, C. R. (1981), "A New Approach to Decomposition of Economic Time Series Into Permanent and Transistory Components With Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 14, 151–174.

Box, G. E. P., Hillmer, S., and Tiao, G. C. (1976), "Analysis and Modeling of Seasonal Time Series," in *Seasonal Analysis of Economic Time Series*, Washington, DC: U.S. Bureau of the Census, pp. 309–334.

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

Box, G. E. P., and Tiao, G. C. (1968), "A Bayesian Approach to Some Outlier Problems," *Biometrika*, 55, 119–129.

——— (1975), "Intervention Analysis With Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 20, 70–79.

Burridge, P., and Wallis, K. F. (1984), "Unobserved-Components Models for Seasonal Adjustment Filters," *Journal of Business and Economic Statistics*, 2, 350–359.

Chang, I., and Tiao, G. C. (1982), "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report 8, University of Chicago, Graduate School of Business.

Chen, C. (1984), "On a Random Level Shift Time Series Model and a Diagnostic Method," unpublished Ph.D. thesis, University of Wisconsin–Madison, Dept. of Statistics.

Chernoff, H., and Zacks, S. (1964), "Estimating the Current Mean of a Normal Distribution Which Is Subject to Change in Time," *The Annals of Mathematical Statistics*, 35, 999–1018.

Clark, P. K. (1973), "A Subordinated Stochastic Process Model With Finite Variance for Speculative Prices," *Econometrica*, 41, 135–156.

Cleveland, W. S., and Tiao, G. C. (1976), "Decomposition of Seasonal Time Series: A Model for the Census X-11 Program," *Journal of the American Statistical Association*, 71, 581–586.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Engle, R. F. (1976), "Estimating Structural Models of Seasonality," in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, DC: U.S. Bureau of the Census, pp. 281–297.

——— (1982), "An ARIMA-Model-Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, 77, 63–70.

Hinkley, D. V. (1970), "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika*, 57, 1–17.

——— (1972), "Time Order Classification," *Biometrika*, 59, 509–523.

Kander, Z., and Zacks, S. (1966), "Test Procedure for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points," *The Annals of Mathematical Statistics*, 37, 1196–1210.

Martin, R. D. (1980), "Robust Estimation of Autoregressive Models," in *Directions in Time Series*, eds. D. R. Brillinger and G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics, pp. 228–254.

Page, E. S. (1955), "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523–526.

——— (1957), "On Problems in Which a Change in a Parameter Occurs at an Unknown Point," *Biometrika*, 44, 248–252.

Press, J. S. (1967), "A Compound Events Model for Security Prices," *Journal of Business*, 40, 317–335.

Said, S. E., and Dickey, D. A. (1985), "Hypothesis Testing in ARIMA(p, 1, q) Models," *Journal of the American Statistical Association*, 80, 369–374.

Taylor, S. J. (1980), "Conjectured Models for Trends in Financial Prices, Tests and Forecasts," *Journal of the Royal Statistical Society*, Ser. A, 143, 338–362.

Tiao, G. C., and Hillmer, S. C. (1978), "Some Consideration of Decomposition of a Time Series," *Biometrika*, 65, 497–502.

Tiao, G. C., and Tsay, R. S. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models," *The Annals of Statistics*, 11, 856–871.

Yao, Y. (1984), "Estimation of a Noisy Discrete-Time Step Function: Bayes and Empirical Bayes Approaches," *The Annals of Statistics*, 12, 1434–1447.