Detecting Abrupt Changes in ARMA Signals

Kevin J. Delaney

Charles W. Therrien*

Abstract

Autoregressive Moving Average (ARMA) models are useful for the short-time description of time series in such diverse fields as speech processing, seismology, and economics. Detecting phonemes, earthquakes or recessions can be viewed as detecting changes in the parameters of the ARMA model. This paper describes a new method for optimal sequential detection of parameter changes in ARMA models. Examples are given of its application.

1 Introduction

This paper describes a new method for detecting changes in time series represented by ARMA models, based upon a method by Nikiforov (Nikiforov, 1986), (Nikiforov and Tikhonov, 1986), (Kushnir, Nikiforov, and Savin, 1983). We first review previous work by Nikiforov, describing a derivation of the sequential change detection method. Following this, we describe the application of Nikiforov's method to AR models and its extension to ARMA models. Finally, we give examples of the algorithm's performance.

2 Sequential change detection method

Hypothesis testing problem This method is based on a Generalized Likelihood Ratio Test (GLRT) which is applied to include cases where the post-change model is unknown. We begin with a hypothesis-testing problem. Under H_0 , the hypothesis that there is no change, the parameter vector $\underline{\theta}$ that describes the model remains equal to the parameter vector $\underline{\theta}_0$ of a known model. Under H_1 , the hy-

pothesis that a model change has occurred, the model changes at some unknown time τ from $\underline{\theta}_0$ to $\underline{\theta}_1 \neq \underline{\theta}_0$:

$$H_0 : \underline{\theta} = \underline{\theta}_0 \quad n = 1, \dots, t$$

$$H_1 : \underline{\theta} = \begin{cases} \underline{\theta}_0 & n = 1, \dots, \tau \\ \underline{\theta}_1 & n = \tau + 1, \dots, t. \end{cases}$$

In order to make a Maximum Likelihood (ML) decision, we form the log-likelihood ratio based upon $\underline{y}_1^t$, the entire received signal up to time t .

$$\ln \Lambda(t) \triangleq \ln \frac{p(\underline{y}_1^t | H_1)}{p(\underline{y}_1^t | H_0)}$$

$$= \ln S(t) - \ln S(\tau)$$

where we have defined:

$$S(t) \triangleq \frac{p_1(\underline{y}_1^t)}{p_0(\underline{y}_1^t)}. \quad (1)$$

The ML estimate of the change time τ is that value of $\hat{\tau}$ that minimizes $\ln S(\hat{\tau})$. Substituting $\hat{\tau}_{ML}$ in place of the unknown τ leads to the GLRT:

$$\ln S(t) - \min_n \ln S(n) \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} h. \quad (2)$$

The likelihood ratio of (1) assumes that we know the value of the parameter vector $\underline{\theta}$ both before and after the change. However, while we know $\underline{\theta}_0$, the model *before* the change, we do not know $\underline{\theta}_1$, the model *after* the change. Using results from Wald (Wald, 1947, pp. 73-77, 199-200) and Roussas (Roussas, 1972, p. 53), Nikiforov derived a sufficient statistic χ_t which does not require knowledge of the post-change model (Nikiforov, 1986, pp. 227-228):

$$\chi_t \triangleq \frac{1}{t} \Delta_t^T(\underline{\theta}_0) F_1^{-1}(\underline{\theta}_0) \Delta_t(\underline{\theta}_0) \quad (3)$$

*Both authors are with the U.S. Naval Postgraduate School, Code EC/Ti, Monterey CA, 93943. For information, contact Prof. C.W. Therrien at (408) 646-3347 or fax (408) 646-2760.

Figure 1: Processing noise-free AR signals

where:

$$\Delta_t(\theta) \triangleq \begin{bmatrix} \frac{d \ln p(y_t | \theta)}{d \theta_0} \\ \frac{d \ln p(y_t | \theta)}{d \theta_1} \\ \vdots \\ \frac{d \ln p(y_t | \theta)}{d \theta_R} \end{bmatrix}$$

$$F_1(\theta) \triangleq E [\Delta_1(\theta) \Delta_1^T(\theta) | \theta].$$

Application to AR models Nikiforov applied the algorithm to detection of AR model changes by deriving Δ_t , F_1 and χ_t for the AR case. By processing the received noise-free signal y through the inverse of the filter which created y , as shown in Figure 1, we produce the white-noise innovations process ϵ . Nikiforov derived an expression for χ_t based upon this inverse-filter innovations process. To summarize Nikiforov's derivation:

$$\Delta_t(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{l=1}^t Y_{l-1} \epsilon_l \\ \frac{1}{\sigma} \sum_{l=1}^t \left(\frac{\epsilon_l^2}{\sigma^2} - 1 \right) \end{bmatrix}$$

$$F_1(\theta_0) = \begin{bmatrix} \frac{R_y}{\sigma^2} & 0 \\ 0^T & \frac{2}{\sigma^2} \end{bmatrix}$$

where we have defined the received signal vector Y_{l-1} as:

$$Y_{l-1} \triangleq \begin{bmatrix} y[l-1] \\ y[l-2] \\ \vdots \\ y[l-P] \end{bmatrix}.$$

3 Extension to noisy ARMA signals

3.1 Derivation

State space approach We have reformulated the change-detection method by allowing the creating filter to be of the more general ARMA type, and for y

Figure 2: Processing noisy ARMA signals

Figure 3: Kalman model

to be contaminated with additive noise. Rather than process y through an inverse filter, as in Nikiforov's noise-free case, we employ a Kalman filter to produce the residual process ϵ , as shown in Figure 2. From the structure of Figure 3, we can directly write the Kalman model in state-space form as:

$$\underline{q}[n+1] = \Phi \underline{q}[n] + \underline{v}[n]$$

$$y[n] = \underline{c}^T \underline{q}[n] + w[n]$$

where $\underline{v}[n]$ is the driving noise, $w[n]$ is the additive noise and:

$$\underline{q}[n] \triangleq \begin{bmatrix} q[n] \\ q[n-1] \\ \vdots \\ q[n-P+1] \end{bmatrix}$$

$$\Phi \triangleq \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_P \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\underline{c} \triangleq [b_0 \ b_1 \ \cdots \ b_Q \ 0 \ \cdots \ 0]^T$$

$$\underline{v}[n] \triangleq \begin{bmatrix} v[n] \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$R_v \triangleq E [\underline{v}[n] \underline{v}^T[n]]$$

$$= \begin{bmatrix} \sigma_v^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}$$

$$\sigma_w^2 \triangleq E [w^2].$$

The above expressions are valid for the case in which $Q + 1 \leq P$. If $Q + 1 > P$, we rearrange the model to retain more of the history of the state $\underline{q}$ than is required for the AR portion. In this case, $\underline{q}[n]$, Φ , and $\underline{c}$ have alternate definitions; we do not show them here as they do not change the subsequent derivation.

We complete the transformation of the model into Kalman form by computing an estimate of the state $\underline{q}$ through the recursive equations (Haykin, 1989, pp. 349-362):

$$\epsilon[n] = y[n] - \underline{c}^T \hat{\underline{q}}[n] \quad (4)$$

$$\hat{\underline{q}}[n+1] = \Phi \hat{\underline{q}}[n] + \underline{g}_n \epsilon[n] \quad (5)$$

$$\sigma_{\epsilon[n]}^2 = \underline{c}^T K_n \underline{c} + \sigma_\omega^2 \quad (6)$$

$$\underline{g}_n = \frac{\Phi K_n \underline{c}}{\sigma_{\epsilon[n]}^2} \quad (7)$$

$$K_{n+1} = \left[\Phi - \underline{g}_n \underline{c}^T \right] K_n \left[\Phi - \underline{g}_n \underline{c}^T \right]^T + R_v + \sigma_\omega^2 \underline{g}_n \underline{g}_n^T. \quad (8)$$

Computing Δ_t Since χ_t is based upon the gradient of the log-likelihood function with regard to the model parameters, we begin by computing the log-likelihood function of the signal.

Because the innovations process ϵ is linearly derived from the signal y , we may use the log-likelihood function of ϵ to compute χ_t . Since $\epsilon[l]$ is a white-noise process, we write:

$$\begin{aligned} p(\epsilon) &= \frac{1}{2\pi^{t/2} \left(\prod_{l=1}^t \sigma_{\epsilon[l]}^2 \right)^{1/2}} e^{-\frac{1}{2} \sum_{l=1}^t \frac{\epsilon^2[l]}{\sigma_{\epsilon[l]}^2}} \\ \ln p(\epsilon) &= -\frac{t}{2} \ln 2\pi - \frac{1}{2} \sum_{l=1}^t \ln \sigma_{\epsilon[l]}^2 - \frac{1}{2} \sum_{l=1}^t \frac{\epsilon^2[l]}{\sigma_{\epsilon[l]}^2} \\ \frac{d \ln p(\epsilon)}{d\theta_i} &= \sum_{l=1}^t \frac{1}{\sigma_{\epsilon[l]}^2} \left(\frac{1}{2} \left(\frac{\epsilon^2[l]}{\sigma_{\epsilon[l]}^2} - 1 \right) \frac{d\sigma_{\epsilon[l]}^2}{d\theta_i} - \epsilon[l] \frac{d\epsilon[l]}{d\theta_i} \right) \end{aligned} \quad (9)$$

where θ_i is any one of the $P + Q + 1$ elements of the model parameter vector $\underline{\theta}$, which we here define as:

$$\underline{\theta} \triangleq \left[-a_1 \quad -a_2 \quad \cdots \quad -a_P \quad b_0 \quad b_1 \quad \cdots \quad b_Q \right]^T.$$

In order to use (9) to compute $\frac{d \ln p(\epsilon)}{d\theta_i}$ and therefore $\Delta_t(\underline{\theta}_0)$, we need expressions for the derivatives $\frac{d\sigma_{\epsilon[l]}^2}{d\theta_i}$ and $\frac{d\epsilon[l]}{d\theta_i}$. We can compute these derivatives recursively along with the recursive computations of (4) -

(8). The result is the following expression for the i^{th} element of $\Delta_t(\underline{\theta}_0)$:

$$\sum_{l=1}^t \frac{1}{\sigma_{\epsilon[l]}^2} \left(\frac{1}{2} \left(\frac{\epsilon^2[l]}{\sigma_{\epsilon[l]}^2} - 1 \right) \frac{d\sigma_{\epsilon[l]}^2}{d\theta_i} + \epsilon[l] \left(\frac{d\underline{c}^T}{d\theta_i} \hat{\underline{q}}[l] + \underline{c}^T \frac{d\hat{\underline{q}}[l]}{d\theta_i} \right) \right) \quad (10)$$

for each value of i : $i = 1, 2, \dots, R$, where $R \triangleq P + Q + 1$.

Computing F_1 Since $F_1(\underline{\theta}_0) \triangleq E_0 [\Delta_1(\underline{\theta}_0) \Delta_1^T(\underline{\theta}_0)]$ then, defining $F_1(i, j)$ as the $(i, j)^{\text{th}}$ element of $F_1(\underline{\theta}_0)$, and using (10) at $t = 1$:

$$\begin{aligned} F_1(i, j) &= \frac{1}{2\sigma_{\epsilon[1]}^4} \frac{d\sigma_{\epsilon[1]}^2}{d\theta_i} \frac{d\sigma_{\epsilon[1]}^2}{d\theta_j} + \frac{1}{\sigma_{\epsilon[1]}^2} \left(\frac{d\underline{c}^T}{d\theta_i} E_0 [\hat{\underline{q}}[1] \hat{\underline{q}}^T[1]] \frac{d\underline{c}}{d\theta_j} \right. \\ &\quad + \underline{c}^T E_0 \left[\frac{d\hat{\underline{q}}[1]}{d\theta_i} \hat{\underline{q}}^T[1] \right] \frac{d\underline{c}}{d\theta_j} + \frac{d\underline{c}^T}{d\theta_i} E_0 \left[\hat{\underline{q}}[1] \frac{d\hat{\underline{q}}[1]}{d\theta_j} \right] \underline{c} \\ &\quad \left. + \underline{c}^T E_0 \left[\frac{d\hat{\underline{q}}[1]}{d\theta_i} \frac{d\hat{\underline{q}}[1]}{d\theta_j} \right] \underline{c} \right). \end{aligned} \quad (11)$$

We can compute the indicated expected values in a recursive manner.

Decision rule In analogy with (2), the decision rule based upon χ_t is of the form:

$$L(\chi_t) = \min_n L(\chi_n) \begin{matrix} H_1 \\ > h. \\ H_0 \end{matrix} \quad (12)$$

To use this rule to accept or reject the change hypothesis H_1 , we must compute the log-likelihood ratio based upon the non-central χ^2 statistic χ_t . This log-likelihood ratio $L(\chi_t)$ is (Jackson, 1960, p. 351), (Patnaik, 1949, p. 204):

$$L(\chi_t) = -t \frac{\lambda_1^2}{2} + \ln \left({}_0F_1 \left(\frac{r}{2}; \frac{t\lambda_1^2 \chi_t}{4} \right) \right) \quad (13)$$

where ${}_0F_1(c; x)$ is the hypergeometric function (Jackson and Bradley, 1961, p. 519):

$${}_0F_1(c; x) \triangleq 1 + \frac{x}{c} + \cdots + \frac{x^n}{c(c+1) \cdots (c+n-1)n!} + \cdots \quad (14)$$

As an infinite series, the hypergeometric function can take thousands of term to converge; rather than compute it directly we use a polynomial approximation we have derived.

Figure 4: Stochastic ARMA signal undergoes model change.

Figure 5: Signal in 3 dB noise.

4 Example

In Figure 4, we see a stochastic signal which undergoes a model change at the time indicated; Figure 5 shows the same signal in 3 dB additive white noise. After computing the statistic χ_t from (3) and its log-likelihood ratio from (13), we detect the model change as shown in Figures 6 and 7.

References

- [1] Haykin, S. S., *Modern Filters*. Macmillan, New York, 1989.
- [2] Jackson, J. E., "A Short Table of the Hypergeometric Function ${}_0F_1(c; x)$ ", *Sankhyā*, Vol. 22, Parts 3 & 4, June 1960, pp. 351-356.

Figure 6: χ_t statistic responds to model change.

Figure 7: Change detected by Generalized Likelihood Ratio Test.

- [3] Jackson, J. E., and Bradley, R. A., "Sequential χ^2 - and T^2 -Tests and Their Application to an Acceptance Sampling Problem," *Technometrics*, Vol. 3, No. 4, November 1961, pp. 519-534.
- [4] Kushnir, A. F., Nikiforov, I. V., and Savin, I. V., "Statistical Adaptive Algorithms for Automatic Detection of Seismic Signals, I. One-Dimensional Case," *Computational Seismology*, Vol. 15, 1983, pp. 145-154.
- [5] Nikiforov, I. V., "Sequential Detection of Changes in Stochastic Systems," *Detection of Abrupt Changes in Signals and Dynamical Systems*, ed. M. Basseville and A. Beneveniste, Springer-Verlag, Berlin, 1986, pp. 218-258.
- [6] Nikiforov, I. V., and Tikhonov, I. N., "Application of Change Detection Theory to Seismic Signal Processing," *Detection of Abrupt Changes in Signals and Dynamical Systems*, ed. M. Basseville and A. Beneveniste, Springer-Verlag, Berlin, 1986, pp. 356-373.
- [7] Patnaik, P. B., "The Non-Central χ^2 and F -Distributions and Their Applications," *Biometrika*, Vol. 36, 1949, pp. 202-232.
- [8] Roussas, G. C., *Contiguity of Probability Measures: Some Applications in Statistics*. Cambridge University Press, London, 1972.
- [9] Wald, A., *Sequential Analysis*. John Wiley & Sons, Inc., New York, 1947.