On Misspecified ARMA Model Fittings to Some Stationary Processes

Minoru Tanaka

School of Network and Information, Senshu University,
2-1-1, Higasimita, Tama-ku, Kawasaki,
Kanagawa 214-8580, Japan

Abstract

This paper gives discussions on (i) a misspecified ARMA(1,1) model fitting to MA(2) processes, and also on (ii) a misspecified MA(2) model fitting to AR(2) processes. They are mainly concerned a problem for finding a number of locally maximal points of the conditional likelihood function of the models when the sample size tends to infinity. It is detected in the case (i) that the general conditions for MA(2) parameters on which the conditional likelihood function of the ARMA(1,1) model has more than one locally maximal points in the stationary and invertible parameter space. Also in the case (ii) it is seen that the MA(2) model has three locally maximal points in the invertible parameter space if the model is fitted to special AR(2) processes. These results are inspected by simulation.

Key words: *ARMA process; ARMA(1,1) and MA(2) model fitting; conditional likelihood function; locally minimal points; misspecification.*

1. Introduction

When applying a model to a time series, we assume a suitable model since we do not know a true model. If it is judged that the data conforms to the model, we shall analyze specification of a spectrum, future prediction, etc. based on the model. Although many methods related with model selection are studied, we here do not take up the subject of model selection. When attention is seldom paid to the specification of the model which we assume first, isn't there any experience in which parameter estimation of the model did not work? When estimating a parameter with a conditional maximum likelihood method (least-squares method), we experience well that a model changes by an initial value, or not being completed as a fixed value for the parameter of a model. By the way, it is well known that if we fit an ARMA($p+g$, $q+g$) model to the data, there is no unique solution and the maximum likelihood method can show strong dependence on the initial conditions, where a series of data is from an ARMA(p, q) process (see [7]). The problem which we deal with here is not a problem of such an exaggerated fitting model. It is a problem in case the model to fit differs from a true model. If the order of an ARMA model is enlarged enough, it may be not taking into consideration. But if we have a small sample, it will be the problem which may fully arise. The purpose of our research is to investigate on what kind of conditions such a situation arises. Since our research has just still started, we will treat the ARMA(1,1) model and also MA(2) model as an early stage.

This paper is a sequel of the paper [11] last year. It relates to incorrect identification of an ARMA(1,1) model. We treated applying this model to the time series which follows AR(2) process incorrectly. We searched for the conditions of the coefficient parameters of AR(2) process in which two or more maximum points exist in quest of a conditional likelihood function paying attention to the number of the maximum points there. The following graphs of the domain is obtained.

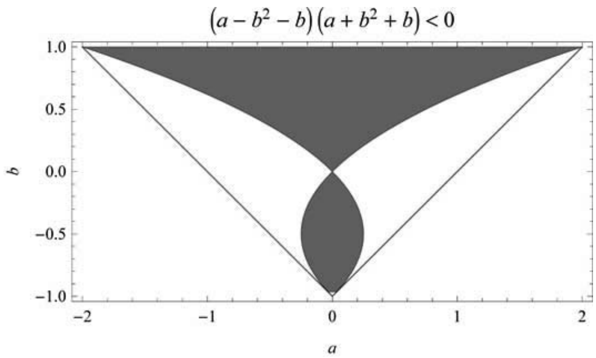


Figure 1. The region of an ARMA(1,1) parameters where more than one locally maximum points exist.

This is also a sequel of the paper "On a moving average time series model fitting" contributed with Mr. Kenji Aoki in 1991 ([12]). It is known that when we fit an MA(1) model to some special time series data which does not follow MA(1) process, the MA(1) parameter does not have an unique Gaussian quasi-maximum likelihood estimator. Tanaka and Huzii [13] have given the conditions of AR(2) parameters on which the MA(1) quasi-likelihood function has more than one local maximal points in the invertible parameter space (-1,1). Furthermore, Tanaka and Aoki [12] gave the region for the AR(2) parameters on which the MA(1) quasi-likelihood function has more than one local maximal points in the parameter space. In this case, maximizing the likelihood function is equivalent to minimizing the following function S(x; a, b) when the data length is large (see [13]). Here x is an MA(1) parameter and a and b are AR(2) parameters.

$$S(x;\, a,\, b) = \frac{1+b-a\,(1-b)\,x-b\,(1+b)\,x^2}{(1-b)\left(1-a^2+2\,b+b^2\right)\left(1-x^2\right)\left(1+a\,x+b\,x^2\right)}. \qquad (1.1)$$

From Tanaka and Huzii [10], we have two minimal points of the function $S(x;a,b) = S(x)$, say. For example, in the case of an AR(2) process with a = -0.1, b = 0.8, the function $S(x)$ has a graph shown in the following figure. In order to have the conditions on which the function has two local minimal points in the parameter space, we should consider the differentiation $DS(x) = 0$. And we specified the case where the solution of the equation $DS(x) = 0$ changed from three to two. That is, the value of the resultant ([5]) was able to formalize the contour line for zero (the bifurcation set). We set the domain D_1 with a deep color surrounded with the curve of the shape of a wedge given in the upper part of Fig. 2. Its boundary is the bifurcation set. It will be seen that the function $S(x;a,b)$ is locally a cusp.

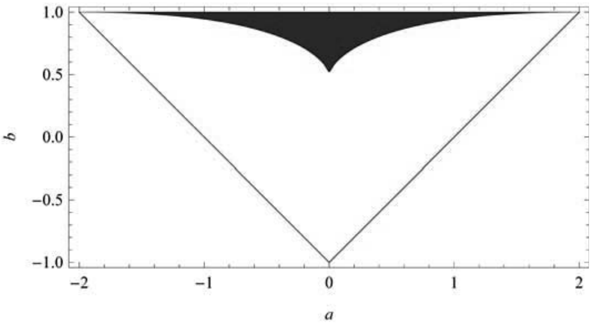


Figure 2. Bifurcation set and the domain for MA(1) model fitting to AR(2) process.

The function S(x) has the two minimum points separated by a maximum within D_1, whereas outside it S(x) has a single minimum, which was given by Prof. Aoki using the concept of the cusp of Catastrophe theory with a potential S(x). It is also seen that the two minimum points are put together and S(x) has only one minimum point at the tip of the wedge (refer to information science research [11], and also [5] and [10] for details).

In this paper, we also consider the ARMA(1,1) model fitting to MA(2) process and study a problem similar to the ARMA(1,1) model fitting to AR(2) processes, and also consider an MA(2) model fitting to AR(2) processes.

2. On misspecified ARMA(1,1) model fitting to an MA(2) process

2.1 Definitions and Notations

Let {Z(t)} be a weakly stationary process with E[Z(t)] = 0. {Z(t)} is said to satisfy a autoregressive moving average model of order p and q (ARMA(p, q) model) if {Z(t)} is expressed as

$$(1 - a_1 B - ... - a_p B^p) Z(t) = (1 + b_1 B + ... + b_q B^q) e(t), \quad (2.1)$$

where {e(t)}, t being an integer, consists of independently and identically distributed random variables with E[e(t)] = 0, $E\left[e(t)^2\right] = \sigma^2$, the a_p's and b_q's are constants which are independent of t, and B is the usual backshift operator such that B[e(t)] = e(t-1) and B^k[e(t)] = $B\left[B^{k-1}[e(t)]\right]$ for k =1,2,.. (see, for example, [3], [4]).

Let

$$\phi(B) = 1 - a_1 B - ... - a_p B^p = \prod_{k=1}^{p} (1 - \phi_k B), \quad (2.2)$$

$$\theta(B) = 1 + b_1 B + ... + b_q B^q = \prod_{k=1}^{q} (1 - \theta_k B). \quad (2.3)$$

In our model fitting, it is assumed that $|\phi_h| < 1$, $|\theta_k| \leq 1$ for all $h = 1, 2, \cdots, p$, and $k = 1, 2, \cdots, q$. Let $\Theta = (\phi_1, \ldots, \phi_p, \theta_1, \cdots, \theta q)$ be a $(p+q)$-dimensional unknown parameter, and let $\{F_k(\Theta)\}$ be a sequence of functions of Θ, which are defined in the following way. For t > 0,

$$e(t) = \{\prod_{k=1}^{p} (1 - \phi_k B) \prod_{k=1}^{q} (1 - \theta_k B)^{-1}\} Z(t) = \left\{\sum_{k=1}^{\infty} F_k(\Theta) B^k\right\} Z(t). \qquad (2.4)$$

For evaluating the asymptotic properties of the conditional quasi-maximum likelihood estimators of Θ when the sample size tends to infinity, we should attend to a function

$$S_{p,q}(\Theta) = E\left[e(t)^2\right] \qquad (2.5)$$

$$= \int_{-1/2}^{1/2} \frac{\left|\prod_{k=1}^{p}[1-\phi_k \exp(-2\pi i\omega)]\right|^2}{\left|\prod_{j=1}^{q}[1-\theta_j \exp(-2\pi i\omega)]\right|^2} f_Z(\omega)\, d\omega.$$

The value $\hat{\Theta}$ which minimizes $S_{p,q}(\Theta)$ with respect to Θ should be obtained (see Tanaka and Huzii [10] and also Huzii [5]). The spectrum of an ARMA(p,q) process, $f_Z(\omega)$, is given by

$$f_Z(\omega) = \frac{\sigma^2}{2\pi} \frac{\left|\theta\left(e^{-i\omega}\right)\right|^2}{\left|\phi\left(e^{-i\omega}\right)\right|^2}. \qquad (2.6)$$

AR and MA spectra are special cases of this spectrum when $\theta(x) = 1$ and $\phi(x) = 1$, respectively. Hence if the process {Z(t)} is an ARMA(p,q) process and is correctly fitted by the ARMA(p,q) model, then we have $S_{p,q}(\Theta) = \frac{\sigma^2}{2\pi}$, which is a spectral density of a white noise process.

Let {X(t)} be a weakly stationary process with mean $E[X(t)] = 0$, *known variance* $E\left[X(t)^2\right] = \sigma_X{}^2$ and spectral density $f_X(\omega)$. When we consider an ARMA(p,q) model fitting to this process {X(t)}, then $S_{p,q}(\Theta)$ is expressed as

$$S_{p,q}(\Theta) = \int_{-1/2}^{1/2} \frac{\left|\prod_{k=1}^{p}[1-\phi_k \exp(-2\pi i\omega)]\right|^2}{\left|\prod_{j=1}^{q}[1-\theta_j \exp(-2\pi i\omega)]\right|^2} f_X(\omega)\, d\omega. \qquad (2.7)$$

In this paper, consideration is given to the case when an ARMA(1,1) model is fitted incorrectly to an MA(2) process {X(t)}; $X(t) = (1 + b_1 B + b_2 B^2)\, e(t)$. Here we set the ARMA(1,1) model parameters (x, y) in stead of (ϕ, θ). In this case, $S_{p,q}(\Theta)$ can be derived from (2.7), ignoring the constant term $\frac{\sigma^2}{2\pi}$ which is known, as

$$S_{11}(x, y) = S_{1,1}(x, y \mid b_1, b_2)$$

$$= \frac{1}{1-y^2}\left\{1 + b_1^2 - 2\,y^2\,b_2 - 2\,x^2\,y^2\,b_2 - 2\,x\,y^3\,b_2 + b_2^2 + 2\,x\,(-b_1 + b_1\,b_2) + 2\,y\,(-b_1 + b_1\,b_2)\,2\,x^2\,y\,(-b_1 + b_1\,b_2) + 2\,x\,y^2\,(-b_1 + b_1\,b_2) + x^2\left(1 + b_1^2 + b_2^2\right) + 2\,x\,y\left(1 + b_1^2 - b_2 + b_2^2\right)\right\} \tag{2.8}$$

If we fit the ARMA(1,1) model to a special MA(2) process, the function $S_{11}(x, y)$ has two locally minimal points. For an example of the MA(2) process with $b_1 = 0.0$, $b_2 = 0.6$, we have the following graph of $S_{11}(x, y)$ on the stationary and invertible space of (x, y).

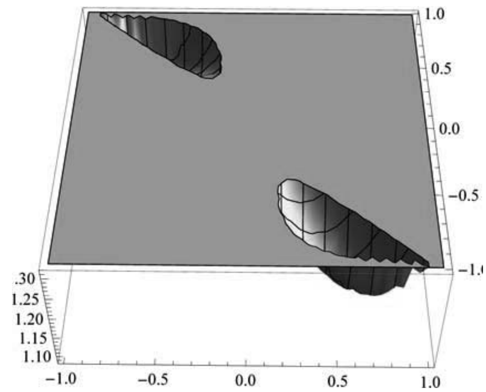


Figure 3. A crosssection of $S_{11}(x, y)$ with $b_1 = 0.0$, $b_2 = 0.6$.

The problem which we consider is investigating the relation between the parameter of the original MA(2) process and the number of the locally minimal point of the conditional likelihood function $S_{11}(x, y)$. Moreover, it is knowing at what rate it happening.

In order to investigate the minimal point of the function $S_{11}(x, y)$, it is first necessary to consider the admissible parameter space (Ω_2) of MA(2) process with parameters b_1 and b_2, where

$$\Omega_2 = \{(b_1,b_2);\ 0 \le (b_2+b_1+1)(b_2+b_1-1),\ -2 \le b_1 \le 2,\ -1 \le b_2 \le 1\}. \tag{2.9}$$

The locally minimal and maximal points satisfy simultaneously the following two equations,

$$\frac{\partial\, S_{11}(x, y)}{\partial_x} = 0, \tag{2.10}$$

$$\frac{\partial\, S_{11}(x, y)}{\partial_y} = 0. \tag{2.11}$$

We shall solve the equations as following. The equation (2.10) is equivalent to

$$-x - y + b_1 + 2\,x\,y\,b_1 + y^2\,b_1 - x\,b_1^2 - y\,b_1^2 + y\,b_2 + 2\,x\,y^2\,b_2 + y^3\,b_2 - b_1\,b_2 - 2\,x\,y\,b_1\,b_2 - y^2\,b_1\,b_2 - x\,b_2^2 - y\,b_2^2 = 0. \tag{2.12}$$

Then we have

$$x = \left(-y + b_1 + y^2 b_1 - y b_1^2 + y b_2 + y^3 b_2 - b_1 b_2 - y^2 b_1 b_2 - y b_2^2\right) / \left(1 - 2 y b_1 + b_1^2 - 2 y^2 b_2 + 2 y b_1 b_2 + b_2^2\right). \quad (2.13)$$

Also the equation (2.11) is equivalent to the following equation,

$$x + y + x^2 y + x y^2 - b_1 - x^2 b_1 - 4 x y b_1 - y^2 b_1 - x^2 y^2 b_1 + x b_1^2 + y b_1^2 + x^2 y b_1^2 + x y^2 b_1^2 - x b_2 - 2 y b_2 - 2 x^2 y b_2 - 4 x y^2 b_2 + x y^4 b_2 + b_1 b_2 + x^2 b_1 b_2 + 4 x y b_1 b_2 + y^2 b_1 b_2 + x^2 y^2 b_1 b_2 + x b_2^2 + y b_2^2 + x^2 y b_2^2 + x y^2 b_2^2 = 0 \quad (2.14)$$

From (2.12) and (2.13), we have

$$(-b_1 - y b_2 + b_1 b_2)\left(-y b_1 + b_1^2 + y^2 b_1^2 - y b_1^3 + b_2 - 2 y^2 b_2 + 2 y b_1 b_2 + 3 y^3 b_1 b_2 - b_1^2 b_2 - 4 y^2 b_1^2 b_2 + y b_1^3 b_2 + 2 y^4 b_2^2 - 2 y b_1 b_2^2 - 3 y^3 b_1 b_2^2 + b_1^2 b_2^2 + y^2 b_1^2 b_2^2 + b_2^3 - 2 y^2 b_2^3 + y b_1 b_2^3\right) = 0 \quad (2.15)$$

In general, it is very difficult to solve the equation, but to know the number of the real solutions it is sufficient to consider the resultant of the polynomial

$$f(y) = (-b_1 - y b_2 + b_1 b_2) \left(-y b_1 + b_1^2 + y^2 b_1^2 - y b_1^3 + b_2 - 2 y^2 b_2 + 2 y b_1 b_2 + 3 y^3 b_1 b_2 - b_1^2 b_2 - 4 y^2 b_1^2 b_2 + y b_1^3 b_2 + 2 y^4 b_2^2 - 2 y b_1 b_2^2 - 3 y^3 b_1 b_2^2 + b_1^2 b_2^2 + y^2 b_1^2 b_2^2 + b_2^3 - 2 y^2 b_2^3 + y b_1 b_2^3\right). \quad (2.16)$$

Since the derivative of the function $f(y)$ is given by

$$\frac{\partial}{\partial y} f(y) = b_1^2 - 2 y b_1^3 + b_1^4 + 6 y b_1 b_2 - 4 b_1^2 b_2 - 12 y^2 b_1^2 b_2 + 12 y b_1^3 b_2 - 2 b_1^4 b_2 - b_2^2 + 6 y^2 b_2^2 - 8 y b_1 b_2^2 - 20 y^3 b_1 b_2^2 + 5 b_1^2 b_2^2 + 30 y^2 b_1^2 b_2^2 - 12 y b_1^3 b_2^2 + b_1^4 b_2^2 - 10 y^4 b_2^3 + 8 y b_1 b_2^3 + 20 y^3 b_1 b_2^3 - 4 b_1^2 b_2^3 - 12 y^2 b_1^2 b_2^3 + 2 y b_1^3 b_2^3 - b_2^4 + 6 y^2 b_2^4 - 6 y b_1 b_2^4 + b_1^2 b_2^4, \quad (2.17)$$

the resultant of the two polynomials (2.16) and (2.17) on y is given as

$$\begin{aligned} R(b_1, b_2) = &-1024 (1 + b_1 - b_2)^2 b_2^{11} (-1 + b_1 + b_2)^2 (-b_1 - b_2 + b_1 b_2)^2 (-b_1 + b_2 + b_1 b_2)^2 \left(1 + b_1^2 + b_2^2\right)^2 \\ &\left(b_1^8 + 12 b_1^6 b_2 + 4 b_1^8 b_2 + 48 b_1^4 b_2^2 + 50 b_1^6 b_2^2 + 4 b_1^8 b_2^2 + 64 b_1^2 b_2^3 + 240 b_1^4 b_2^3 + 84 b_1^6 b_2^3 - 4 b_1^8 b_2^3 + 544 b_1^2 b_2^4 + \right. \\ &357 b_1^4 b_2^4 + 78 b_1^6 b_2^4 - 10 b_1^8 b_2^4 + 512 b_2^5 + 448 b_1^2 b_2^5 + 636 b_1^4 b_2^5 + 64 b_1^6 b_2^5 - 4 b_1^8 b_2^5 + 1632 b_1^2 b_2^6 + 510 b_1^4 b_2^6 + \\ &78 b_1^6 b_2^6 + 4 b_1^8 b_2^6 + 1536 b_2^7 + 768 b_1^2 b_2^7 + 636 b_1^4 b_2^7 + 84 b_1^6 b_2^7 + 4 b_1^8 b_2^7 + 1632 b_1^2 b_2^8 + 357 b_1^4 b_2^8 + \\ &\left. 50 b_1^6 b_2^8 + b_1^8 b_2^8 + 1536 b_2^9 + 448 b_1^2 b_2^9 + 240 b_1^4 b_2^9 + 12 b_1^6 b_2^9 + 544 b_1^2 b_2^{10} + 48 b_1^4 b_2^{10} + 512 b_2^{11} + 64 b_1^2 b_2^{11}\right). \end{aligned} \quad (2.18)$$

From the Catastrophe theory, a number of locally minimum points of $S_{11}(x, y)$ on Ω_2 for MA(2) process with parameters (b_1, b_2) is explained by considering a change for the sign of the resultant $R(a,b)$. If the two polynomials (2.16) and (2.17) have common zeros, the resultant must be vanished. Hence we consider the conditions for $R(b_1, b_2)= 0$ on Ω_2. Since the polynomial $\left(1 + b_1^2 + b_2^2\right)^2$ in (2.18) is always positive on Ω_2, it is sufficient to consider the zeros of the polynomial such that

$$
\begin{aligned}
G_1(b_1, b_2) = {} & (1 + b_1 - b_2)(-1 + b_1 + b_2)(-b_1 - b_2 + b_1 b_2)(-b_1 + b_2 + b_1 b_2) \\
& \left(b_1^8 + 12 b_1^6 b_2 + 4 b_1^8 b_2 + 48 b_1^4 b_2^2 + 50 b_1^6 b_2^2 + 4 b_1^8 b_2^2 + 64 b_1^2 b_2^3 + 240 b_1^4 b_2^3 + 84 b_1^6 b_2^3 - \right. \\
& 4 b_1^8 b_2^3 + 544 b_1^2 b_2^4 + 357 b_1^4 b_2^4 + 78 b_1^6 b_2^4 - 10 b_1^8 b_2^4 + 512 b_2^5 + 448 b_1^2 b_2^5 + 636 b_1^4 b_2^5 + \\
& 64 b_1^6 b_2^5 - 4 b_1^8 b_2^5 + 1632 b_1^2 b_2^6 + 510 b_1^4 b_2^6 + 78 b_1^6 b_2^6 + 4 b_1^8 b_2^6 + 1536 b_2^7 + 768 b_1^2 b_2^7 + \\
& 636 b_1^4 b_2^7 + 84 b_1^6 b_2^7 + 4 b_1^8 b_2^7 + 1632 b_1^2 b_2^8 + 357 b_1^4 b_2^8 + 50 b_1^6 b_2^8 + b_1^8 b_2^8 + 1536 b_2^9 + \\
& \left. 448 b_1^2 b_2^9 + 240 b_1^4 b_2^9 + 12 b_1^6 b_2^9 + 544 b_1^2 b_2^{10} + 48 b_1^4 b_2^{10} + 512 b_2^{11} + 64 b_1^2 b_2^{11}\right).
\end{aligned}
\tag{2.19}
$$

Then we have the following graph for a contour of $G_1(b_1, b_2) = 0$ on Ω_2.

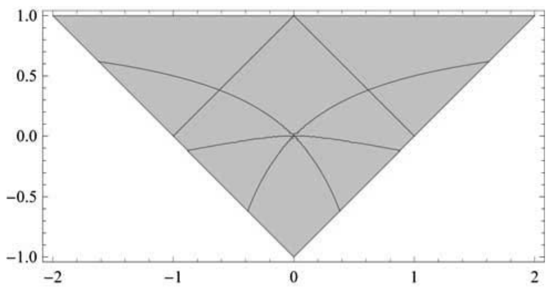


Figure 4. A contour line of $G_1(b_1, b_2) = 0$ on Ω_2.

It turns out that the function $S_{11}(x, y)$ has the two minimum points in a domain (D_2) of a portion with a deep color surrounded with the curve in Figure.5, where

$$D_2 = \{(b_1, b_2) \in \Omega_2 ;\ (1 + b_1 - b_2)(-1 + b_1 + b_2)(-b_1 - b_2 + b_1 b_2)(-b_1 + b_2 + b_1 b_2) < 0\}. \tag{2.20}$$

Also we define the (bifurcation) set

$$B_2 = \{(b_1, b_2) \in \Omega_2 ;\ (1 + b_1 - b_2)(-1 + b_1 + b_2)(-b_1 - b_2 + b_1 b_2)(-b_1 + b_2 + b_1 b_2) = 0\}. \tag{2.21}$$

When numerical integration is performed by using Mathematica (Ver.7), it turns out that the area of this domain D_2 is about 2.490 square, and the rate to the parameter space of a lower triangle is 62.3% exactly. This means that if we incorrectly fit the ARMA(1,1) model to a series of an MA(2) process, then there are 62.3% probability for existing two locally maximum likelihood estimators on the stationary and invertible parameter space.

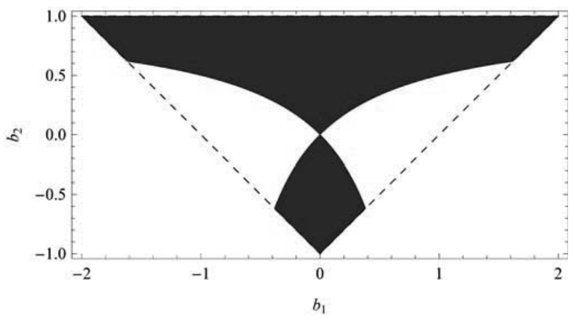


Figure 5. The domain D_2 in Ω_2.

We next determine the property of $S_{11}(x, y)$ at every point in D_2 by considering only one point within each of the domains.

2.2. Illustrations and Simulation study

2.2.1. Illustrations

By varying the MA(2) parameters, b_1 and b_2, continuously and staying inside of D_2, for example, going from position P_1 to P_2 in Fig.6, the system remains in a stable equilibrium that is the function $S_{11}(x, y)$ has two minima. However, if a and b are changed so that the bifurcation set B_2 is transversed, something unusual happens. To see this, start in position P_2 of Fig.6, where the system is in a stable equilibrium. Moving parallel to the b_1-axis toward position P_3, when the position is reached, the system becomes unstable the and the function $S_{11}(x, y)$ has only one minima. There the system is stable again and remains so while moving onward to position P_4. In position P_5 inside of D_2, it is also seen that the function $S_{11}(x, y)$ has two minima.

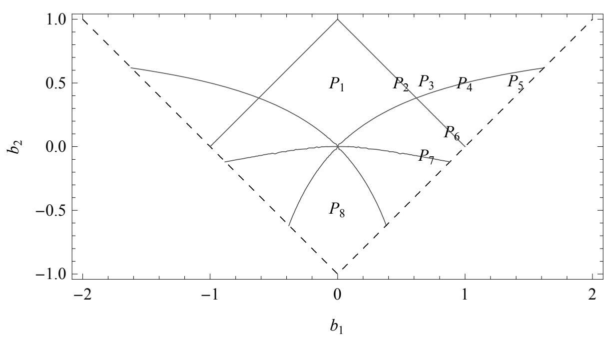


Figure 6. Selected MA(2)-parameters (b_1, b_2) of positions P_1- P_8.

[1] position P_1 ; b_1= 0.0 and b_2 = 0.5. In this case, $S_{11}(x, y)$ has two locally minimum points on the parameter space Ω_2 at {x = -0.601501, y = 0.831254} and {x = 0.601501, y = -0.831254} shown in Fig.2.2.1.

[2] position P_2 ; b_1 = 0.5 and b_2 = 0.5. In this case, $S_{11}(x, y)$ has only one locally minimum on the parameter space Ω_2 at {x = 0.820194, y = -0.859612} shown in Fig.2.2.2.

[3] position P_3 ; b_1 = 0.7 and b_2 = 0.5. In this case, $S_{11}(x, y)$ has two locally minimum at {x = 0.896162, y = -0.907935} and {x = -0.398676, y = 0.90415} shown in Fig.2.2.3.

[4] position P_4 ; b_1 = 1.0 and b_2 = 0.5 (lies in B2). In this case, $S_{11}(x, y)$ has only one locally minimum at {x = -0.387582, y = 0.790048} shown in Fig.2.2.4.

[5] position P_5 ; b_1 = 1.4 and b_2 = 0.5. In this case, $S_{11}(x, y)$ has only one locally minimum on the parameter space Ω_2 at {x = -0.36349, y = 0.675553} shown in Fig.2.2.5.

[6] position P_6 ; b_1 = 0.9 and b_2 = 0.1. In this case, $S_{11}(x, y)$ has no locally minimum points on the parameter space Ω_2 shown in Fig.2.2.6.

[7] position P_7 ; b_1 = 0.7 and b_2 = -0.085687, which is on the line. In this case, $S_{11}(x, y)$ has only one locally minimum points on the parameter space Ω_2 at {x = 0.129372, y = 0.569795} shown in Fig.2.2.7.

[8] position P_8 ; b_1 = 0.0 and b_2 = -0.5. In this case, $S_{11}(x, y)$ has two locally minimum at {x = -0.765121, y = 0.653491} and {x = 0.765121, y = -0.653491} shown in Fig.2.2.8.

The following figures give cross-sectional images of $S_{11}(x, y)$ with the parameters (b_1, b_2) of positions P_1- P_8, respectively.

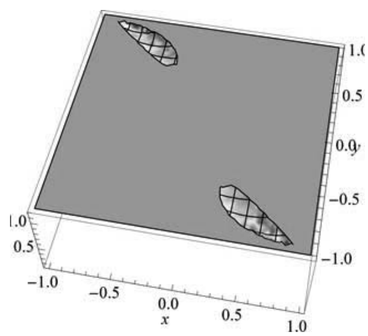


Figure 2.2.1. $S_{11}(x, y)$ with b_1 = 0.0 and b_2 = 0.5.

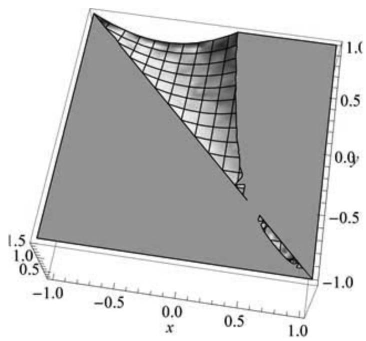


Figure 2.2.2. $S_{11}(x, y)$ with b_1 = 0.5 and b_2 = 0.5.

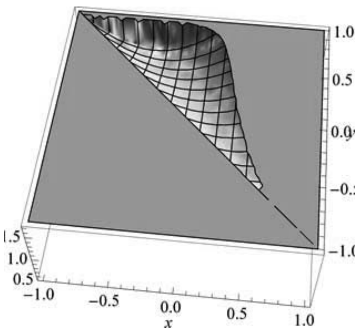


Figure 2.2.3. $S_{11}(x, y)$ with b_1 = 0.7 and b_2 = 0.5.

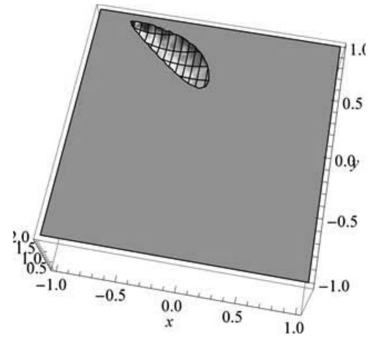


Figure 2.2.4. $S_{11}(x, y)$ withb_1 = 1.0 and b_2 = 0.5.

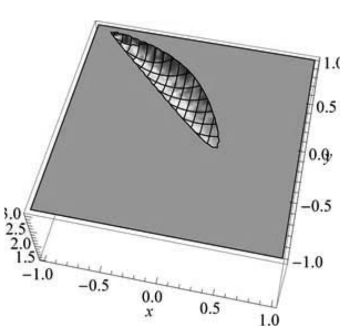


Figure 2.2.5. $S_{11}(x, y)$ with b_1 = 1.4 and b_2 = 0.5.

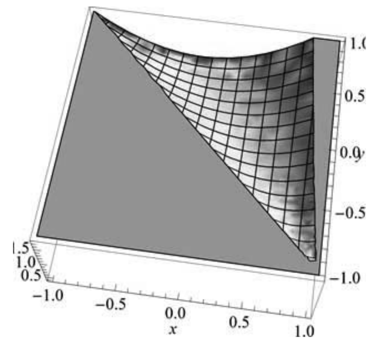


Figure 2.2.6. $S_{11}(x, y)$ with b_1 = 0.9 and b_2 = 0.1.

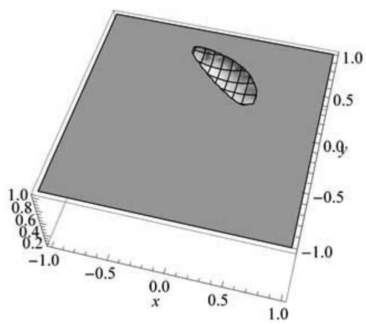


Figure 2.2.7. $S_{11}(x, y)$ with b_1 = 0.7 and b_2 = -0.08.

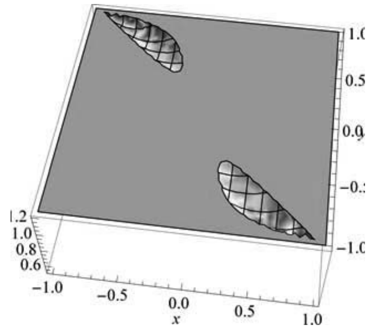
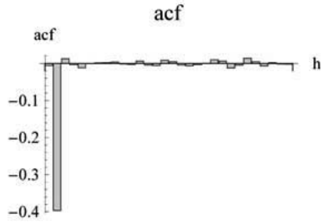


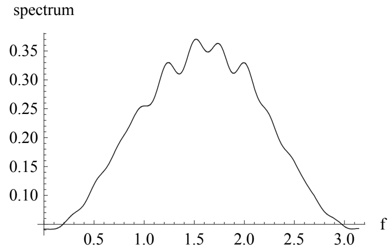
Figure 2.2.8. $S_{11}(x, y)$ with b_1 = 0.0 and b_2 = -0.5.

2.2.2. Computer simulation

We generate a time series of length n = 40,000 from the MA(2) models which are discussed above (1), ... , (8), where the noise is generated from the normal distribution with mean 0 and variance 1. Then we fit an ARMA(1,1) model to each of the time series using the conditional maximum likelihood method with initial values of parameters for the arguments (x, *y*) of the model. The calculations below are supported by the computer software *Mathematica* (Ver.7) and an application software ([7]).

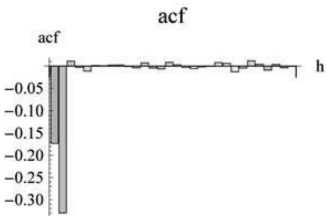
(1) Case when MA(2) process with parameters (b_1, b_2) = (0.0, 0.5).

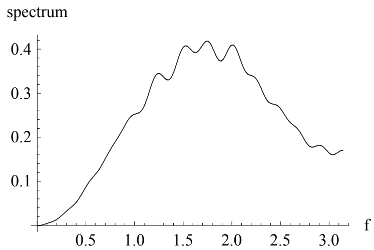




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.5, y = -0.5) are provided as the arguments of ARMA(1,1) model. Then we have ARMA(1,1) model with {x = 0.604353}, {y = -0.829897} as the conditional maximum likelihood estimate of the model. On the other hand, different initial values (x = -0.5, y = 0.5) lead to another model, ARMA model with {x = -0.598163}, {y = 0.828965}. Therefore we can have two conditional maximum likelihood estimates of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.0, 0.5), which corresponds to the discussion (1) in 2.2.1 and also Figure 2.2.1.

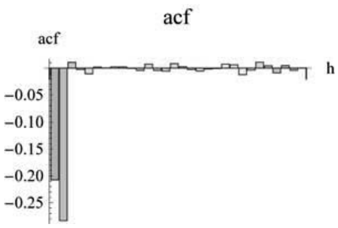
(2) Case when MA(2) process with parameters $(b_1, b_2) = (0.5, 0.5)$.

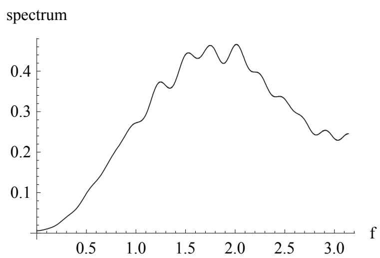




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.82, y = -0.86) are provided as the arguments of ARMA(1,1). Then we have an ARMA model with {x = 0.817475}, {y = -0.854429} as the conditional maximum likelihood estimate of an ARMA(1,1) model. On the other hand, different initial values (x = -0.5, y = 0.5) lead to another model, ARMA model with {x = -0.396277}, {y = 0.997437}, this is almost on the boundary of the domain. Therefore we can have only one conditional maximum likelihood estimate of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.5,0.5), which corresponds to the discussion (2) in 2.2.1 and also Figure 2.2.2.

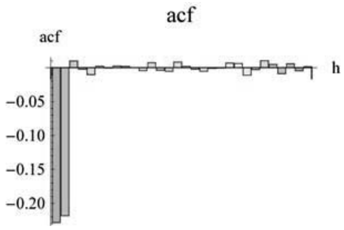
(3) Case when MA(2) process with parameters $(b_1, b_2) = (0.7, 0.5)$.

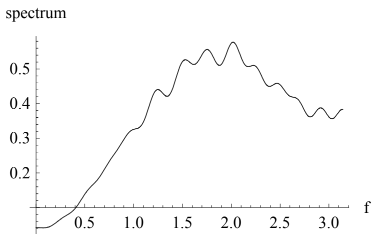




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.9, y = -0.9) are provided as the arguments of ARMA(1,1). Then we have an ARMA model with {x = 0.883103}, {y = -0.893064} as the conditional maximum likelihood estimate of an ARMA(1,1) model. On the other hand, different initial values (x = -0.5, y = 0.5) lead to another model, ARMA model with {x = -0.393588}, {y = 0.90174}. Therefore we can have two conditional maximum likelihood estimates of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.7, 0.5), which corresponds to the discussion (3) in 2.2.1 and also Figure 2.2.3.

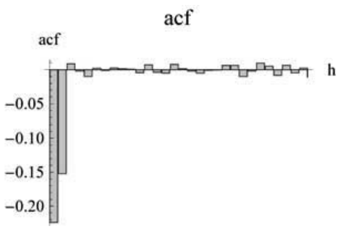
(4) Case when MA(2) process with parameters $(b_1, b_2) = (1.0, 0.5)$.

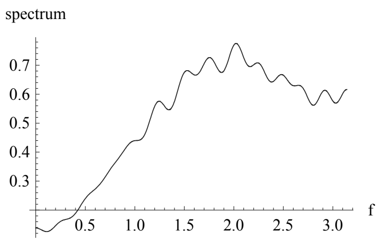




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.5, y = -0.5) are provided as the arguments of ARMA(1,1). Then we have an ARMA model with {x = -0.3821}, {y = 0.787593} as the conditional maximum likelihood estimate of an ARMA(1,1) model. On the other hand, different initial values (x = -0.5, y = 0.5) lead to the same model, ARMA model with {x = -0.382131}, {y = 0.787618}. Therefore we can have only one conditional maximum likelihood estimate of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (1.0, 0.5), which corresponds to the discussion (4) in 2.2.1 and also Figure 2.2.4.

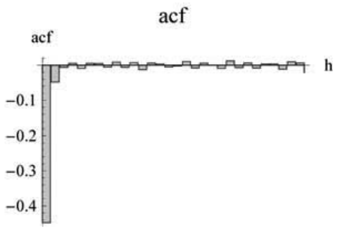
(5) Case when MA(2) process with parameters $(b_1, b_2) = (1.4, 0.5)$.

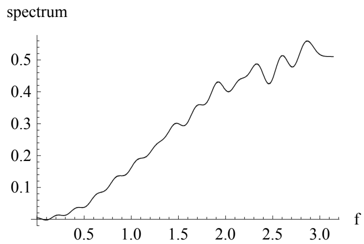




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = -0.5, y = 0.5) are provided as the arguments of ARMA(1,1). Then we have an ARMA model with {x = -0.354893}, {y = 0.670486} as the conditional maximum likelihood estimate of an ARMA(1,1) model. On the other hand, different initial values (x = -0.5, y = 0.5) lead to the same model, ARMA model with {x = 0.35491}, {y = 0.670501. Therefore we can have only one conditional maximum likelihood estimate of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (1.4, 0.5), which corresponds to the discussion (5) in 2.2.1 and also Figure 2.2.5.

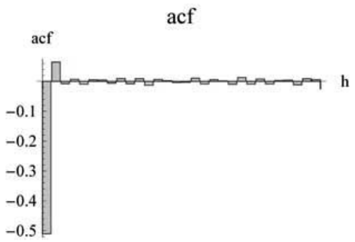
(6) Case when MA(2) process with parameters (b_1, b_2) = (0.9, 0.1).

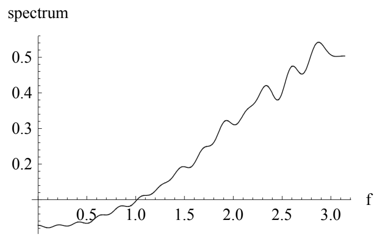




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.5, y = -0.5) are provided as the arguments of ARMA(1,1). Then we have an ARMA model with {x = -0.0951825}, {y = 0.994696} as the conditional maximum likelihood estimate of an ARMA(1,1) model, and a different initial value (x = -0.5, y = 0.5) lead to the same model, ARMA model with {x = -0.0951822}, {y = 0.994696}, this is almost on the boundary of the domain. Therefore we have no conditional maximum likelihood estimate of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.9, 0.1), which corresponds to the discussion (6) in 2.2.1 and also Figure 2.2.6.

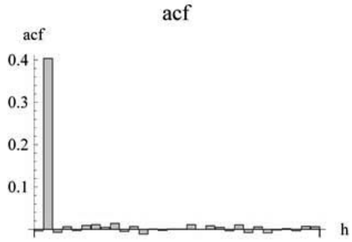
(7) Case when MA(2) process with parameters (b_1, b_2) = (0.7, -0.086).

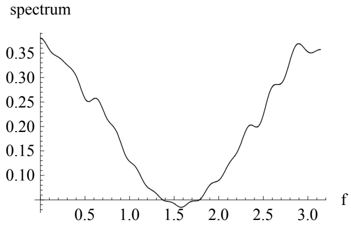




These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.5, y = -0.5) are provided as the arguments of ARMA model(1,1). Then we have an ARMA model with {x = 0.134752}, {y = 0.566116} as the conditional maximum likelihood estimate of an ARMA(1,1) model. Also, different initial values (x = -0.5, y = 0.5) lead to the same ARMA model with {x = 0.134751}, {y = 0.566117}. Therefore we can have only one conditional maximum likelihood estimate of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.7, -0.086), which corresponds to the discussion (7) in 2.2.1 and also Figure 2.2.7.

(8) Case when MA(2) process with parameters (b_1, b_2) = (0.0, -0.5).



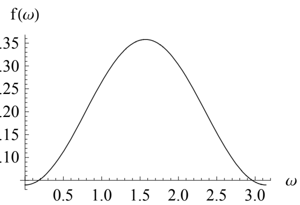


These are plots of the sample auto-correlation function and the sample spectrum. We estimate the ARMA(1,1) model parameters using the conditional maximum likelihood method with some different initial parameter values. The initial parameter values (x = 0.75, y = -0.65) are provided as the arguments of ARMA(1,1) model. Then we have an ARMA model with {x = 0.766094}, {y = -0.650514} as the conditional maximum likelihood estimate of the model. On the other hand, different initial values (x = -0.75, y = 0.65) lead to another ARMA model with {x = -0.774099}, {y = 0.664496}. Therefore we can have two conditional maximum likelihood estimates of an ARMA(1,1) model when we fit the ARMA(1,1) model to the MA(2) process with the parameters (0.0, -0.5), which corresponds to the discussion (8) in 2.2.1 and also Figure 2.2.8.

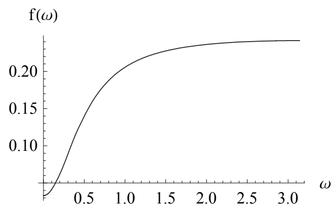
3. Averaging model of all fitted models

Isn't there any method of approximating the true model (process) which generated the data from two or more of the incorrect-identified models? We propose a new method (averaging model) by use of the estimated ARMA(1,1) models from the example treated in Chapter 2. The concept for the model averaging is given in bayesian model averaging (Lunn, Jackson, Best, Thomas and Spiegelhalter [9]). They said that Bernardo and Smith [2] showed decision-theoretically this provides optimal prediction or estimation under an "M-closed" situation, in which the true process is among the list of candidate models. Our situation is an "M-open", in which the true process is not there any more. In this section we shall only make a suggestion since the theoretical discussion seems to be very difficult for us.

(1) MA(2) model with b_1= 0.0 and b_2 = 0.5. In this case, we have two ARMA(1,1) models with parameters {x = -0.601501, y = 0.831254} and {x = 0.601501, y = -0.831254}. The true spectral density function of the MA(2) process and spectral densities of the fitted ARMA(1,1) models are



Spectrum of MAmodel[{0.0, 0.5}, 1.0] (True)

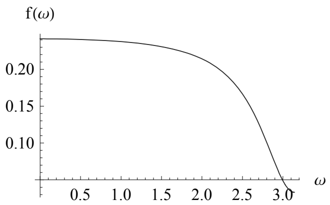


Spectrum of ARMAmodel[{-0.601}, { 0.831}, 1.2]

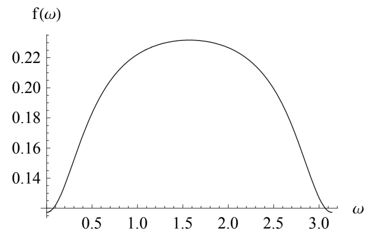
Therefore, we define what compounded the spectrum of two applied models (average). It turns out that this reproduces the feature which the original spectrum has. We also define as follows the model which combined two models (average). When the transfer function of the two ARMA(1,1) models is weight averaged, it turns out that this serves as a transfer function of an ARMA(2,2) model. The weight of a weighted average uses the reciprocal of noise variance (in this case, since both two models have equal variance, it serves as an arithmetic average).

$$\frac{1}{2}\left(\frac{1+0.831254\,B}{1-0.601501\,B}+\frac{1-0.831254\,B}{1+0.601501\,B}\right) = \frac{1.+0.5\,B^2}{1.+0.\,B-0.361803\,B^2} \tag{3.1}$$

Thus the averaging model is an ARMA(2,2) model with parameters {0.0, -0.361803} and {0.0, 0.5}.



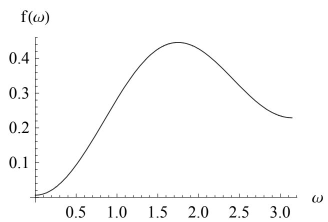
Spectrum of ARMAmodel[{0.601}, { -0.831}, 1.2]



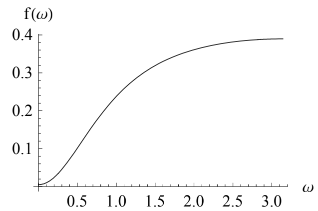
Spectrum of ARMAmodel[{0.0, 0.361803}, {0.0, -0.5}, 1.2]

The averaging spectrum expresses well the feature of the spectrum of a true model (MA(2) process).

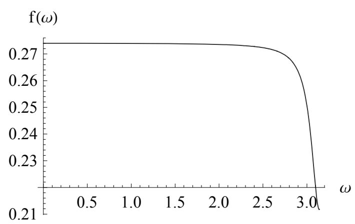
(2) MA(2) model with $b_1 = 0.7$ and $b_2 = 0.5$. In this case, $S_{11}(x, y)$ has two locally minimum at {x = 0.896162, y = -0.907935} and {x = -0.398676, y = 0.90415} shown in Fig.3.3.



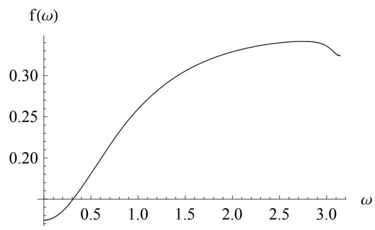
Spectrum of MAmodel[{0.7, 0.5}, 1.0] (True)



Spectrum of ARMAmodel[{-0.398}, {0.904}, 1.2]



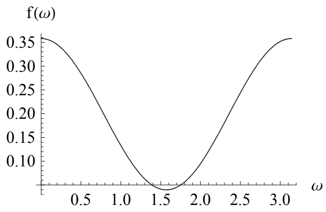
Spectrum of ARMAmodel[{0.896}, {-0.908}, 1.2]



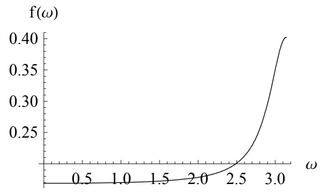
Spectrum of the averaging Spectrum

The equalization (averaging) spectrum seems to express well the feature of the spectrum of a true model (MA(2) process) rather than the spectrum of the ARMA(1,1) model except for the position of a peak.

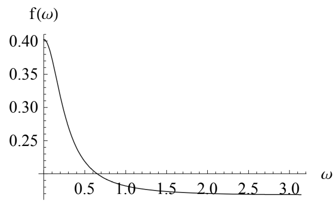
(3) MA(2) model with b_1 = 0.0 and b_2 = -0.5. In this case, $S_{11}(x, y)$ has two locally minimum at {x = -0.765121, y = 0.653491} and {x = 0.765121, y = -0.653491}.



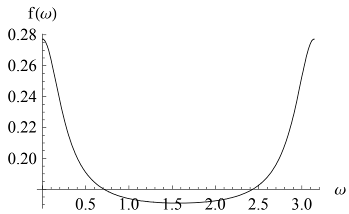
Spectrum of MAmodel[{0.0, -0.5}, 1.0] (True)



Spectrum of ARMAmodel[{-0.765}, {0.659}, 1.2]



Spectrum of ARMAmodel[{0.765}, {-0.659}, 1.2]



Spectrum of Averaging Spectrum

We can say that the equalization (averaging) spectrum expresses well the feature of the spectrum of a true MA(2) process rather than each spectrum of the ARMA(1,1) models.

4. On misspecified MA(2) model fitting to an AR(2) process

When the incorrect-identified model is applied, how many the misspecified models are presumed? Although the ARMA(1,1) model had been considered until now, even when a true model was which of AR(2) and MA(2), the model obtained with the conditional maximum likelihood method was at most two. It is imagined that the number of the models presumed changes by the model to fit and also by the true process. Here we shall pay attention to MA(2) model. Furthermore, we assume that the time series applied to the model follows AR(2) process. Since calculation is very complicated and generalities are not made, we consider a special case only. These contents serve as extension of the paper before fitting MA(1) model to AR(2) process. We note saying to how many the model which locally maximizes a conditional likelihood function appears. Although it was a maximum of two until now in the case of this MA(2) model fitting, the example in which three models appear is found. And we can confirm the fact in simulation with the case of a large sample.

We consider the case when an MA(2) model is fitted incorrectly to an AR(2) process {X(t)}, $(1 - a_1 B - a_2 B^2)$ X(t) = e(t). We set the MA(2) model parameters (x, y). In this case, $S_{p,q}(\Theta)$ can be derived from (2.7) as

$$S_2(x, y) = S_2(x, y; a_1, a_2) = \frac{f(x,y)}{g(x,y)},$$

where $f(x, y) = 1 - y - x a_1 - x y a_1 + y a_1^2 - y^2 a_1^2 + a_2 - x^2 a_2 - y a_2 - x^2 y a_2 + x a_1 a_2 - x y^2 a_1 a_2 - y a_1^2 a_2 + y^2 a_1^2 a_2 - x^2 a_2^2 - x^2 y a_2^2 - y^2 a_2^2 + y^3 a_2^2 + x y a_1 a_2^2 + x y^2 a_1 a_2^2 - y^2 a_2^3 + y^3 a_2^3$,

$$g(x, y) = (1 + x - y)(1 + y)(-1 + x + y)(-1 + a_2)(1 - a_1 + a_2)(1 + a_1 + a_2)\left(1 + x a_1 - y a_1^2 + x^2 a_2 + 2 y a_2 - x y a_1 a_2 + y^2 a_2^2\right). \quad (4.1)$$

Fallowing to the previous section, we have tried to analysis the locally minimum points of the S_2(x, y). But it is very difficult to solve the general equations such that

$$\frac{\partial S_2(x, y)}{\partial_x} = 0, \quad (4.2)$$

$$\frac{\partial S_2(x, y)}{\partial_y} = 0. \quad (4.3)$$

Here we present a special example in which the function $S_2(x, y)$ has three locally minimal points on the invertible parameter space. We have the following graph of a crosssection of the $S_2(x, y)$ if the fitted model is an AR(2) process whose parameters are $a_1 = 0.0$ and $a_2 = 0.95$.

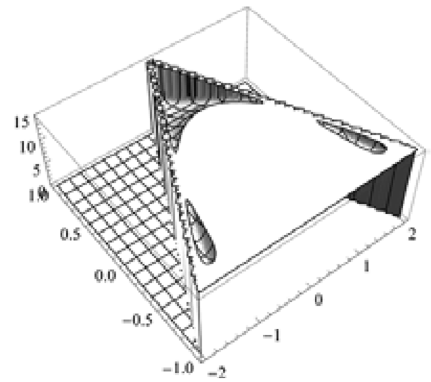


Figure 4.1. A crosssection of $S_2(x, y)$ when $a_1 = 0.0$ and $a_2 = 0.95$.

In order to investigate the minimal point of the function $S_2(x, y)$, it is first necessary to consider its locally minimal points on the admissible parameter space (Ω_{2A}) of AR(2) process with parameters a_1 and a_2, where

$$\Omega_{2A} = \{(a_1, a_2);\ 0 \le (a_1 + a_2 + 1)(a_2 - a_1 + 1),\ -2 \le a_1 \le 2,\ -1 \le a_2 \le 1\}. \quad (4.4)$$

The locally minimal and maximal points satisfy the following equations,

$$-1.8x + 3.8x^3 - 1.805x^5 + 5.79xy - 3.8x^3y - 1.805x^5y - 3.99xy^2 - 3.4295x^3y^2 - 3.249xy^3 + 3.4295x^3y^3 + 5.04949xy^4 - 1.80049xy^5 = 0 \tag{4.5}$$

$$-1.9 + 3.805x^2 - 1.71x^4 + 2.19y + 3.8x^2y - 5.415x^4y + 5.30525y^2 - 7.5905x^2y^2 - 5.14425x^4y^2 - 5.9705y^3 - 6.4885x^2y^3 - 1.6245x^4y^3 - 4.9495y^4 + 3.78599x^2y^4 + 5.41049y^5 + 3.4295x^2y^5 + 1.54327y^6 - 1.62901y^7 = 0 \tag{4.6}$$

The real solutions of two equations above are shown in Figure.4.2

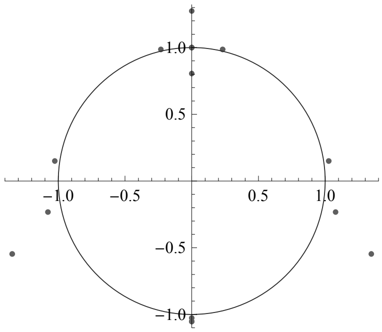


Figure. 4.2. Real solutions

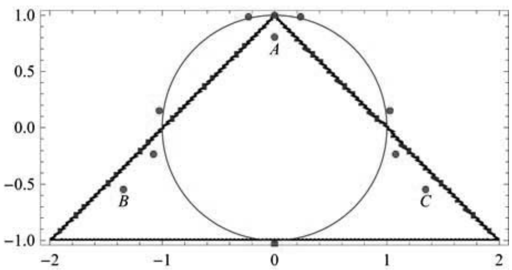


Figure. 4.3. Three locally minimal points

We can see in Figure.4.3 that there are three locally minimal points in the domain Ω_{2A} such that

A: {0.0, 0.805225}, B: { -1.3453, -0.546645}, C: { 1.3453, -0.546645}.

Corresponding to these points, we have three MA(2) models which have the points for their parameter. We show three spectral density functions of these models and that of the true model.

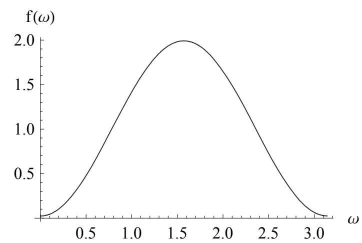


Figure.4.4. Spectral density function for A.

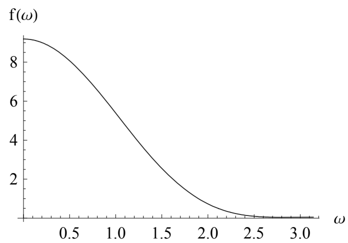


Figure.4.5. Spectral density function for B.

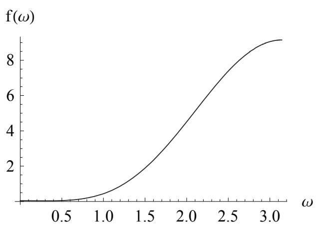


Figure.4.6. Spectral density function for C.

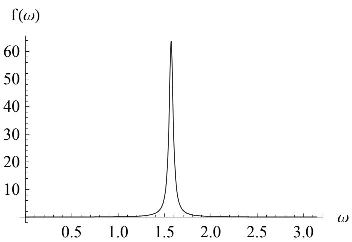


Figure.4.7. Spectral density of the true model.

Furthermore, in the case of $a_1 = 0.0$ and $a_2 \geq 0.94$, we can also determine that there are three MA(2) models which are fitted to the AR(2) process.

5. Conclusion

In Section 2, we have considered the misspecified ARMA(1,1) model fitting to MA(2) processes following to the previous paper[11] in 2012. The conditions for MA(2) parameters on which ARMA(1,1) quasi-likelihood function has more than one local maximum points in the stationary and invertible parameter space were given as the domain D_2 for MA(2) parameters (b_1, b_2) shown in Figure.5. It related to critical point theory and the behavior of degenerate critical points of the function of two variables in Catastrophe theory, considering the ARMA(1,1) quasi-likelihood function as a potential function with two external parameters b_1 and b_2.

In Section 4, we have also considered on the misspecified MA(2) model fitting to AR(2) processes. It was already given the domain for AR(2) parameters on which the MA(1) quasi-likelihood function has more than one local maximum point. Our new result presented here is that the MA(2) quasi-likelihood function has three local maximum points in the invertible parameter space Ω_2. Furthermore we have shown that more general ARMA model has more than three local maximum points in the stationary and invertible parameter space Ω_2. However, I have not performed yet determining the domain where three models exist in parameter space Ω_2. We will wait for future research findings about this problem. Moreover, is the number of a misspecified model estimated to at most three? We have discovered an example to which six models are estimated by the initial value in a simulation for an ARMA(3,3) model fitting to ARMA(3,6) processes. However, though regrettable, theoretical proof is not made to this result. It is also a future subject about this problem.

Considering these researches, we shall also conjecture that an ARMA(p,q) model has more than one locally maximum points in the stationary and invertible parameter space, if it fitted to a series belongs to an ARMA(p, q+r) process for any positive integer p, q and some r≥1.

The purpose of our research at the last is to investigate what kind of phenomenon happens, when the misspecified model is applied to a certain time series, but probably, it may be insufficient. It will be necessary to utilize well two or more models obtained there, and to make it useful for the estimation of a true model, as we discussed in Section 3.

References

[1] Åström, K.J. and Söderström, T., 1974, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model", IEEE Trans. Automat. Contr., 19, 769-773.

[2] Bernardo, J. M. and Smith, A.F.M., 1994, *Bayesian theory*, John Wiley & Sons, New York.
[3] Box, G.E.P. and Jenkins, G.M., 1970, *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
[4] Brockwell, P.J. and Davis, R.A., 1991, *Time Series : Theory and Methods,* Springer, New York.
[5] Castrigiano, D.P.L. and Hayes, S.A., 2004, *Catastrophe theory*, Westview Press.
[6] Huzii, M., 1988, "Some properties of conditional quasi-likelihood functions for time series model fitting", Journal of Time Series Analysis, 9, 345-352.
[7] He,Y., 1995, *Time Series Pack for Mathematica*, Wolfram Research.
[8] Kabaila, P., 1983, "Parameter values of ARMA models minimizing the one-step-ahead prediction error when the true system is not in the model set", J. Appl . Prob., 20, 405-408.
[9] Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D., 2013, *The BUGS Book*, CRC Press, Boca Raton, FL.
[10] Poston,T. and Stewart, I.N., 1978, *Catastrophe theory and its applications,* Pitman Publishing Limited.
[11] Tanaka, M., 2012, "On Some Properties of ARMA(1,1) Model Fitting to AR(2) Processes", Bulletin of the Institute of Information Science, Vol.20, 1 - 15.
[12] Tanaka, M. and Aoki, K., 1991, "On a moving average time series model fitting" (in Japanese), Bulletin of the Institute of Information Science, Vol.12, 42 - 54.
[13] Tanaka, M. and Huzii, M., 1992, "Some properties of moving - average model fittings", J. Japan Statist. Soc., Vol.22, No .1, 33 - 44.