# Inference for Regression Models With Errors From a Non-invertible MA(1) Process

Meiching Chen*
Chaoyang University of Technology

Richard A. Davis*
Colorado State University

Abstract

This paper considers maximum likelihood estimation in a regression model when the errors follow a first order moving average model which is non-invertible or nearly non-invertible. The latter corresponds to a moving average parameter $\theta$ that is equal to or close to 1. The joint limiting distribution of the maximum likelihood estimators $\hat{\mathbf{b}}$ and $\hat{\theta}$ of the regression parameter vector $\mathbf{b}$ and the moving average parameter $\theta$ is described. Unlike the case with standard time series models, the limiting distribution of $\hat{\mathbf{b}}$ depends on whether or not $\theta$ is being estimated. Specifically, the limit distribution of $\hat{\mathbf{b}}$ is non-normal if $\theta$ is also being estimated and is normal if it is assumed that $\theta$ is fixed at either 1. The asymptotic behavior of the generalized likelihood ratio statistic for testing $\theta = 1$ vs. $\theta < 1$ is also studied and shown to perform well compared to the locally best invariant unbiased test of Tanaka (1990).

*This research supported in part by NSF DMS Grant No. 9504596 and by the National Science Council No. NSC87-2119-M-324-001.
*AMS 1980 Subject Classification:* 62M10; 62M09; 62M07
*Key Words and Phrases.* Regression model with moving average errors, unit roots, non-invertible moving averages, maximum likelihood estimator

# 1 Introduction

Let $Y_1, \ldots, Y_T$ be observations from the regression model given by

$$Y_t = \mathbf{x}_t' \mathbf{b}_0 + U_t \tag{1.1}$$

where $\mathbf{x}_t'$ is a $k$-dimensional row vector and $\mathbf{b}_0$ is a $k$-dimensional regression parameter. The error sequence $\{U_t\}$ is assumed to be generated from the MA(1) process

$$U_t = \epsilon_t - \theta_0 \epsilon_{t-1}, \tag{1.2}$$

where $\{\epsilon_t\} \sim \text{IID}(0, \sigma^2)$, $E\epsilon_t^4 < \infty$ and $|\theta_0| \leq 1$. The objective of this paper is to consider the asymptotic distribution of the maximum likelihood estimators of $\mathbf{b}$ and $\theta_0$ when the true value of $\theta_0$ is at or close to 1. Here, likelihood refers to the Gaussian likelihood based on the observations $Y_1, \ldots, Y_T$. Maximizing the Gaussian likelihood is a commonly used method of estimation in time series and the resulting estimates have a number of desirable properties even if the data are non-Gaussian.

One important application of model (1.1) is to the case of an MA(1) model with non-zero mean, i.e.,

$$Y_t = b + \epsilon_t - \theta_0 \epsilon_{t-1}. \tag{1.3}$$

A complete asymptotic theory of the maximum likelihood estimator, $\hat{\theta}_{MLE}$, and the local maximizer of the likelihood closest to 1, $\hat{\theta}_{LM}$, of $\theta_0$ was derived in Davis and Dunsmuir (1996) and Davis, Chen, and Dunsmuir (1995) (referred to as DD and DCD in the sequel) for the case that $\theta_0$ was close to or equal to 1 and $b = 0$. Most of the early work on this problem concentrated on the so-called pile-up effect for the mean zero case. The pile-up effect corresponds to the situation when the likelihood function has a local maximum at $\theta = 1$. If $\theta_0 = 1$, then $P[\hat{\theta}_{LM} = 1] \to .6575$ (see Anderson and Takemura (1986) and Tanaka and Satchell (1989)). As will be shown below (see also Sargan and Bhargava (1983) and Shephard (1993)), the pile-up effect is more severe when the model contains a mean term. As a byproduct of our general result in Section 2, it follows that the maximum likelihood estimator $\hat{b}$ of $b$ for model (1.3) is asymptotically normal for $\theta_0 = 1$, where $\hat{b} = (\mathbf{e}_T' \Omega^{-1}(1) \mathbf{e}_T)^{-1} \mathbf{e}_T' \Omega^{-1}(1) \mathbf{Y}$, $\mathbf{e}_T$ is a $T \times 1$ vector of ones, and $\sigma^2 \Omega(\theta)$ is the covariance matrix of $\mathbf{Y} = (Y_1, \ldots, Y_T)$. As is easy to see, the sample mean for this model is not asymptotically normal unless the $\epsilon_t$'s are normally distributed. On the other hand, if $\theta_0 = 1$, but must be estimated, then $\hat{b}$ is no longer asymptotically normal nor asymptotically independent of $\hat{\theta}_{MLE}$. This phenomenon is quite different than what occurs for standard time series models.

While the bulk of research on the unit root problem for moving averages has been devoted to the mean 0 case, Sargan and Bhargava (1983) and Pesaran (1983) showed that for the non-zero mean case, $\hat{\theta}_{LM} = 1 + O_P(T^{-1})$. They also calculated the probability that a local maximum of the likelihood function is equal to one. Shephard (1993) discusses the use of profile and marginal likelihoods for inference purposes in a regression model with stochastic trend components. He argues that for inferences about $q$, the signal-to-noise ratio which is related to the moving average parameter $\theta$, it is preferable to use the marginal likelihood. In our situation, inferences about both the regression and moving average parameters are of interest.

Our approach to studying the limit behavior of the maximum likelihood estimates is similar to the techniques employed in DD and DCD. We first build the location and scaling constants into the parameterization of the model. Specifically, define

$$\boldsymbol{\delta} \;=\; D_T^{-1}(\mathbf{b} - \mathbf{b}_0)$$

and

$$\beta \;=\; T(1-\theta),$$

where $D_T$ is a diagonal matrix to be specified. We now view the likelihood function, $L_T$, as a function of $\boldsymbol{\delta}$ and $\beta$ so that the *local maximum* estimators of $\mathbf{b}$ and $\theta$ satisfy

$$\hat{\boldsymbol{\delta}}_{LM} = D_T^{-1}(\hat{\mathbf{b}}_{LM} - \mathbf{b}_0) \;\text{ and }\; \hat{\beta}_{LM} = T(1-\hat{\theta}_{LM}),$$

where $\hat{\theta}_{LM}$ is the local maximizer of the profile likelihood that is closest to 1 and $\hat{\mathbf{b}}_{LM}$ is the resulting generalized least squares estimator of $\mathbf{b}$. The main step is show that the profile likelihood, suitably normalized, converges as a stochastic process indexed by $\beta$. From this, it can be shown that $\hat{\boldsymbol{\delta}}_{LM}$ and $\hat{\beta}_{LM}$ converge in distribution, with the latter converging to the corresponding maximizer of the limit process. The statements and proofs of these results are contained in Section 2.

In Section 3, we consider examples of regression functions that satisfy the growth conditions required for the asymptotic theory of Section 2 to hold. These include constant functions and polynomials in time. The theory is then applied to the overshorts data contained in Sections 3.2.4 and 6.3.2 of Brockwell and Davis (1996). This data come from a problem involving estimation of the leak rate of an underground storage tank. In this instance both the regression function consisting of a constant mean and the MA(1) parameter have physical interpretations.

In Section 4, we consider the asymptotic behavior of the likelihood ratio (LR) test for testing the null hypothesis that $\theta_0 = 1$ versus $\theta_0 < 1$ for the MA(1) model with non-zero mean. Simulation results show that the LR test is quite competitive with the locally best invariant unbiased (LBIU) test of Tanaka (1990). In particular, the power of the test is nearly the same for local alternatives close to $\theta_0 = 1$ and is much larger for alternatives further away from $\theta_0 = 1$. We also apply the LR test to the overshort data.

## 2 Asymptotic Theory

The regression model (1.1) can be written in matrix notation as

$$\mathbf{Y} = X\mathbf{b}_0 + \mathbf{U} \tag{2.1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_T)'$, X is the design matrix consisting of the rows $\mathbf{x}_i', i = 1, ..., T$, and $\mathbf{U} = (U_1, \ldots, U_T)'$ is the vector of errors which are assumed to follow the MA(1) model,

$$U_t = \epsilon_t - \theta_0 \epsilon_{t-1},$$

with $\{\epsilon_t\} \sim \text{IID}(0, \sigma^2)$, $E\epsilon_t^4 < \infty$ and $|\theta_0| \leq 1$. The logarithm of the Gaussian likelihood function, based on the observation vector $\mathbf{Y}$ is

$$-\frac{T}{2}\log(2\pi\gamma(0)) - \frac{1}{2}\log|G| - \frac{1}{2\gamma(0)}(\mathbf{Y} - X\mathbf{b})'G^{-1}(\mathbf{Y} - X\mathbf{b}),$$

where $\gamma(0) = \text{Var}(U_1)$ and $G = E(\mathbf{U}\mathbf{U}')/\gamma(0)$.

After concentrating $\gamma(0)$ out of the likelihood and deleting constant terms, the reduced 2log likelihood becomes

$$M(\mathbf{b}, \theta) = -\log|G| - T\log(\mathbf{Y} - X\mathbf{b})'G^{-1}(\mathbf{Y} - X\mathbf{b}). \tag{2.2}$$

We now reparameterize the model using $\boldsymbol{\delta} = D_T^{-1}(\mathbf{b} - \mathbf{b}_0)$ and $\beta = T(1 - \theta)$, where $D_T$ is a diagonal matrix of size $k$ (the dimension of the regression vector $\mathbf{x}_i$) to be specified shortly. Using this new parameterization, define $L_T(\boldsymbol{\delta}, \beta) = M(D_T\boldsymbol{\delta} + \mathbf{b}_0, 1 - \beta/T)$, so that the logarithm of the concentrated likelihood function is given by

$$\begin{aligned} L_T(\boldsymbol{\delta}, \beta) &= M(D_T\boldsymbol{\delta} + \mathbf{b}_0, 1 - \beta/T) \\ &= -\log|G| - T\log(\mathbf{Y} - X\mathbf{b}_0 - XD_T\boldsymbol{\delta})'G^{-1}(\mathbf{Y} - X\mathbf{b}_0 - XD_T\boldsymbol{\delta}). \end{aligned} \tag{2.3}$$

For $\beta$ fixed, the maximum of $L_T$ with respect to $\boldsymbol{\delta}$ occurs when $\hat{\boldsymbol{\delta}}(\beta) = (D_T X' G^{-1} X D_T)^{-1} D_T X' G^{-1} U$. We then show below that the profile likelihood suitably centered converges in distribution as a stochastic process indexed by $\beta \geq 0$.

As in standard regression problems, growth conditions on the regression variables are imposed to obtain limiting results. To set up the requisite conditions in this setting, let $P = [P_{st}]_{s,t=1}^T$ be the orthogonal matrix with entries $P_{st} = \sqrt{\frac{2}{T+1}} \sin \frac{\pi st}{T+1}$(see equation (15) in Anderson and Takemura (1986)). Assume that $x_{ti} = f_i(t)$ for some functions $f_i$ and put $a_{it,T} = \sum_{s=1}^T f_i(s) P_{st}$ , $i = 1, 2, ..., k$ and $t = 1, 2, ..., T$. If $h_T$ is a sequence of integers satisfying $h_T \to \infty$, $h_T/T \to 0$ and $h_T^2/T \to \infty$, then for each $i = 1, \ldots, k$, assume that there exist a negative real $m_i$ and constants $a_{it}^*$ such that

$$\sum_{t=1}^{\infty} \frac{|a_{it}^*|}{t} < \infty, \tag{$A0$}$$

$$(A1) \qquad \sup_{1\le t\le h_T} |T^{m_i} a_{it,T} - a^*_{it}| \to 0 \quad \text{as} \quad T \to \infty,$$

and

$$(A2) \qquad \sup_{1\le t\le T} |T^{m_i} a_{it,T}| \quad \text{is bounded in T.}$$

Set

$$\begin{aligned} D_T &= \operatorname{diag}(T^{m_i-1}, i = 1, \dots, k), \\ \boldsymbol{\delta} &= D_T^{-1}(\mathbf{b} - \mathbf{b}_0), \end{aligned}$$

and

$$\beta = T(1-\theta).$$

The following theorem describes the limiting behavior of $\hat{\boldsymbol{\delta}}(\beta)$ and $L_T(\hat{\boldsymbol{\delta}}(\beta), \beta) - L_T(\hat{\boldsymbol{\delta}}(0), 0)$ as random functions (on the space $C_k[0,\infty)$ and $C[0,\infty)$, respectively) of $\beta$. Here $C_k[0,\infty)$ denotes the space of $R^k$-valued continuous functions on $[0,\infty)$ and convergence is defined as uniform convergence on compact sets.

**Theorem 2.1** *Suppose $Y_1, \dots, Y_T$ are observations from model (2.1) with true parameters $\mathbf{b}_0$ and $\theta_0 = 1 - \frac{\gamma}{T}$ for some $\gamma \ge 0$. If the regression variables satisfy conditions $(A0) - (A2)$, then as $T \to \infty$*
*(i) $\hat{\boldsymbol{\delta}}(\beta) \xrightarrow{d} \sigma B^{-1}(\beta)\mathbf{C}_\gamma(\beta)$ on $C_k[0,\infty)$,*
*and*
*(ii) $L_T(\hat{\boldsymbol{\delta}}(\beta), \beta) - L_T(\hat{\boldsymbol{\delta}}(0), 0) \xrightarrow{d} W_\gamma(\beta)$ on $C[0,\infty)$,*
*where*

$$\begin{aligned} W_\gamma(\beta) &= Z_\gamma(\beta) + \mathbf{C}'_\gamma(\beta)B^{-1}(\beta)\mathbf{C}_\gamma(\beta) - \mathbf{C}'_\gamma(0)B^{-1}(0)\mathbf{C}_\gamma(0), \\ \mathbf{a}^*_t &= (a^*_{1t}, ...., a^*_{kt})', \\ B(\beta) &= \sum_{t=1}^{\infty} \frac{\mathbf{a}^*_t \mathbf{a}^{*'}_t}{\pi^2 t^2 + \beta^2}, \\ \mathbf{C}_\gamma(\beta) &= \sum_{t=1}^{\infty} \frac{\mathbf{a}^*_t \sqrt{\pi^2 t^2 + \gamma^2} X_t}{\pi^2 t^2 + \beta^2}, \\ Z_\gamma(\beta) &= \sum_{t=1}^{\infty} \log \frac{\pi^2 t^2}{(\pi^2 t^2 + \beta^2)} + \sum_{t=1}^{\infty} \frac{\beta^2(\pi^2 t^2 + \gamma^2) X_t^2}{(\pi^2 t^2 + \beta^2)\pi^2 t^2}, \end{aligned}$$

*and $\{X_t\}$ is a sequence of iid $\mathrm{N}(0,1)$ random variables.*

**Remark 2.1.** Using the argument given for the Corollary on p.8 of DD, it can be shown that $\hat{\beta}_{LM} \xrightarrow{d} \tilde{\beta}_{LM}$, where $\tilde{\beta}_{LM}$ is the local maximizer of the $W_\gamma(\beta)$ that is closest to 0. It follows that

$$\begin{bmatrix} T(\hat{\theta}_{LM} - 1) \\ D_T^{-1}(\mathbf{b}_{LM} - \mathbf{b}_0) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} -\tilde{\beta}_{LM} \\ \tilde{\boldsymbol{\delta}}(\tilde{\beta}_{LM}) \end{bmatrix},$$

where $\tilde{\boldsymbol{\delta}}(\beta) = B^{-1}(\beta)C_\gamma(\beta)$. In particular, $P(\hat{\beta}_{LM} = 0) = P(\hat{\theta}_{LM} = 1) \xrightarrow{d} P(\tilde{\beta}_{LM} = 0)$.

**Remark 2.2.** As mentioned in DD, convergence on $C[0,\infty)$ does not necessarily imply convergence of the corresponding global maximizers. Additional arguments were required in the zero-mean case (see DCD) to show that the maximum likelihood estimator converged in distribution to the global maximizer of the limit process. We suspect that the same holds here for $\hat{\beta}_{MLE}$ and simulation results, some of which are contained in Section 4, bear this out.

**Proof of Theorem 2.1**: We start from the profile or concentrated likelihood given by (2.3),

$$L_T(\hat{\boldsymbol{\delta}}(\beta),\beta) = -\log|G(\beta)| - T\log\left(\mathbf{U}'G^{-1}(\beta)\mathbf{U} - \mathbf{U}'G^{-1}(\beta)XD_T\hat{\boldsymbol{\delta}}(\beta)\right), \tag{2.4}$$

where $\hat{\boldsymbol{\delta}}(\beta)$ is the generalized least squares estimate,

$$\hat{\boldsymbol{\delta}}(\beta) = (D_TX'G^{-1}XD_T)^{-1}D_TX'G^{-1}U. \tag{2.5}$$

Now, using equation (15) in Anderson and Takemura (1986), we have

$$G = P\Lambda P'$$

where $\Lambda = \text{diag}(1+2p_Td_1, 1+2p_Td_2, \ldots, 1+2p_Td_T)$, $P$ is the orthogonal matrix described earlier, $p_T = cor(U_t, U_{t+1}) = -(1-\beta/T)/(1+(1-\beta/T)^2)$, and $d_t = \cos\frac{\pi t}{T+1}$.

It follows that

$$\begin{aligned} X'G^{-1}X &= X'P\Lambda^{-1}P'X \\ &= \sum_{t=1}^{T}\frac{\mathbf{a}_{t,T}\mathbf{a}'_{t,T}}{1+2p_Td_t}, \end{aligned}$$

where

$$[\mathbf{a}_{1,T}, \ldots., \mathbf{a}_{T,T}] = X'P.$$

Moreover,

$$\begin{aligned} X'G^{-1}\mathbf{U} &= X'P\Lambda^{-1}P'U \\ &= \sigma_T\sum_{t=1}^{T}\frac{\mathbf{a}_{t,T}\sqrt{1+2q_Td_t}U_{t,T}}{1+2p_Td_t}, \end{aligned}$$

where

$$\begin{aligned} U_{t,T} &= \sqrt{\frac{2}{T+1}}\sigma_T^{-1}(1+2q_Td_t)^{-\frac{1}{2}}\sum_{s=1}^{T}U_s\sin\frac{\pi st}{T+1}, \\ \sigma_T^2 &= (1+(1-\frac{\gamma}{T})^2)\sigma^2, \end{aligned}$$

and

$$q_T = -(1-\frac{\gamma}{T})\sigma^2/\sigma_T^2.$$

We show for $M > 0$ fixed and $\theta_0 = 1 - \gamma/T$, that

$$\text{(2.6)} \qquad \begin{aligned} D_T X' G^{-1} X D_T &\to 2\sum_{t=1}^{\infty} \frac{\mathbf{a}_t^* \mathbf{a}_t^{*'}}{\pi^2 t^2 + \beta^2} \quad \text{uniformly in } \beta \in [0, M] \\ &=: 2B(\beta) \end{aligned}$$

and

$$\begin{aligned} D_T X' G^{-1} U &\stackrel{d}{\to} \sum_{t=1}^{\infty} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2 + \gamma^2} X_t}{\pi^2 t^2 + \beta^2} \\ &=: 2\sigma \mathbf{C}_\gamma(\beta) \quad \text{on } C_k[0, M]. \end{aligned}$$

By Assumption $(A1)$ , we have

$$\sup_{1 \le t \le h_T} |T D_T \mathbf{a}_{t,T} - \mathbf{a}_t^*| \to 0$$

and

$$\text{(2.8)} \qquad \sup_{1 \le t \le h_T} |T^2 D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T - \mathbf{a}_t^* \mathbf{a}_t^{*'}| \to 0,$$

where we define the absolute value of a matrix to be the corresponding matrix of absolute values of the components. Using assumption $(A2)$ and (2.17) in DD, there exists a positive constant C such that

$$\begin{aligned} \left|\sum_{t=h_T+1}^{T} \frac{D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T}{1 + 2p_T d_t}\right| &\le C \left|\frac{1}{(T+1)^2} \sum_{t=h_T+1}^{T} \frac{1}{1 + 2p_T d_t}\right| \\ &\to 0\,. \end{aligned}$$

Now from (2.9) and (2.11) in DD, we have

$$\begin{aligned} &\left|\sum_{t=1}^{h_T} \frac{D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T}{1 + 2p_T d_t} - 2\sum_{t=1}^{h_T} \frac{\mathbf{a}_t^* \mathbf{a}_t^{*'}}{\pi^2 t^2 + \beta^2}\right| \\ &\quad = \left|\sum_{t=1}^{h_T} \frac{2(T+1)^2 D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T (\pi^2 t^2 + \beta^2) - 4(T+1)^2 \mathbf{a}_t^* \mathbf{a}_t^{*'} (1 + 2p_T d_t)}{2(T+1)^2 (1 + 2p_T d_t)(\pi^2 t^2 + \beta^2)}\right| \\ &\quad \le \left|\sum_{t=1}^{h_T} \frac{[2(T+1)^2 D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T - 2\mathbf{a}_t^* \mathbf{a}_t^{*'}](\pi^2 t^2 + \beta^2) - 2\mathbf{a}_t^* \mathbf{a}_t^{*'} (\pi^2 t^2 + \beta^2) o(1)}{(\pi^2 t^2 + \beta^2)(1 + o(1)) \pi^2 t^2}\right| \\ &\quad \le \sup_{1 \le t \le h_T} |2(T+1)^2 D_T \mathbf{a}_{t,T} \mathbf{a}_{t,T}' D_T - 2\mathbf{a}_t^* \mathbf{a}_t^{*'}| \sum_{t=1}^{h_T} \frac{1}{\pi^2 t^2} \\ &\qquad + 2 \sup_{1 \le t \le h_T} |\mathbf{a}_t^* \mathbf{a}_t^{*'}| o(1) \sum_{t=1}^{h_T} \frac{1}{\pi^2 t^2} \\ &\quad \to 0, \end{aligned}$$

where the limit follows from (2.7) and $(A2)$. This convergence is in fact uniform on $\beta \in [0, M]$ for any $M > 0$ from which (2.6) follows.

Similarly, on $C_k[0, M]$

$$|\sum_{t=h_T+1}^{T} \frac{D_T \mathbf{a}_{t,T}\sqrt{1+2q_T d_t} U_{t,T}}{1+2p_T d_t}| \xrightarrow{P} 0$$

and

$$\begin{aligned} &|\sum_{t=1}^{h_T} \frac{D_T \mathbf{a}_{t,T}\sqrt{1+2q_T d_t} U_{t,T}}{1+2p_T d_t} - \sqrt{2}\sum_{t=1}^{h_T} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} U_{t,T}}{\pi^2 t^2+\beta^2}| \\ &\leq |\sum_{t=1}^{h_T} \frac{[\sqrt{2}(T+1)D_T \mathbf{a}_{t,T} - \sqrt{2}\mathbf{a}_t^*]\sqrt{\pi^2 t^2+\gamma^2}(\pi^2 t^2+\beta^2) U_{t,T}}{(\pi^2 t^2+\beta^2)(1+o(1))\pi^2 t^2}| \\ &\leq \sup_{1\leq t\leq h_T} |\sqrt{2}(T+1)D_T \mathbf{a}_{t,T} - \sqrt{2}\mathbf{a}_t^*| \sum_{t=1}^{h_T} \frac{1}{\pi^2 t^2} |U_{t,T}|. \end{aligned}$$

Because $E|U_{t,T}| \leq EU_{t,T}^2 = 1$, we have from (2.8) that

$$|\sum_{t=1}^{T} \frac{D_T \mathbf{a}_{t,T}\sqrt{1+2q_T d_t} U_{t,T}}{1+2p_T d_t} - \sqrt{2}\sum_{t=1}^{h_T} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} U_{t,T}}{\pi^2 t^2+\beta^2}| \xrightarrow{P} 0.$$

By the weak convergence part of Proposition $A2$ in DD and the continuous mapping theorem,

$$\sqrt{2}\sum_{t=1}^{m} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} U_{t,T}}{\pi^2 t^2+\beta^2} \xrightarrow{d} \sqrt{2}\sum_{t=1}^{m} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} X_t}{\pi^2 t^2+\beta^2}$$

on $C_k[0, M]$. By Assumption $(A0)$, there exists a positive constant $C_1$ such that

$$\begin{aligned} |\sqrt{2}\sum_{t=m+1}^{\infty} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} X_t}{\pi^2 t^2+\beta^2}| &\leq C_1 \sum_{t=m+1}^{\infty} \frac{|\mathbf{a}_t^*||X_t|}{t} \\ &\xrightarrow{d} 0. \end{aligned}$$

as $m \to \infty$ Therefore, we have

$$\sqrt{2}\sum_{t=1}^{m} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} X_t}{\pi^2 t^2+\beta^2} \xrightarrow{d} \sqrt{2}\sum_{t=1}^{\infty} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} X_t}{\pi^2 t^2+\beta^2}$$

as $m \to \infty$. Similarly, there exists a positive constant $C_2$ such that

$$\limsup_{T\to\infty} E[\sup_{0\leq\beta\leq M} |\sqrt{2}\sum_{t=m+1}^{h_T} \frac{\mathbf{a}_t^* \sqrt{\pi^2 t^2+\gamma^2} U_{t,T}}{\pi^2 t^2+\beta^2}|] \leq C_2 \sum_{t=m+1}^{\infty} \frac{|\mathbf{a}_t^*|}{t} \to 0$$

as $m \to 0$. Applying Theorem 4.2 in Billingsley (1968), (2.7) now follows.

Combining (2.5), (2.6) and (2.7), we conclude that

$$\hat{\boldsymbol{\delta}}(\beta) \xrightarrow{d} \sigma B^{-1}(\beta)\mathbf{C}_\gamma(\beta) \tag{2.9}$$

which proves (i).

Now let

$$A_1(\beta) = \mathbf{U}'G^{-1}(\beta)\mathbf{U} = \sigma_T^2 \sum_{t=1}^{T} \frac{(1+2q_T d_t)U_{t,T}^2}{1+2p_T d_t}$$

and

$$A_2(\beta) = \mathbf{U}'G^{-1}(\beta)XD_T\hat{\boldsymbol{\delta}}(\beta).$$

Then equation (2.4) becomes

$$\begin{aligned} L_T(\hat{\boldsymbol{\delta}}(\beta),\beta) &= -\sum_{t=1}^{T}\log(1+2p_T d_t) - T\log[A_1(\beta) - A_2(\beta)] \\ &= -\sum_{t=1}^{T}(1+2p_T d_t) - T\log A_1(\beta) - T\log[1 - \frac{A_2(\beta)}{A_1(\beta)}]. \end{aligned}$$

Therefore,

$$\begin{aligned} &L_T(\hat{\boldsymbol{\delta}}(\beta),\beta) - L_T(\hat{\boldsymbol{\delta}}(0),0) \\ &\quad = \sum_{t=1}^{T}\log\frac{1-d_t}{1+2p_T d_t} + T\log\frac{A_1(0)}{A_1(\beta)} - T\log(1-\frac{A_2(\beta)}{A_1(\beta)}) + T\log(1-\frac{A_2(0)}{A_1(0)}). \end{aligned}$$

Observe that

$$\begin{aligned} &-T\log(1-\frac{A_2(\beta)}{A_1(\beta)}) + T\log(1-\frac{A_2(0)}{A_1(0)}) \\ &\quad \sim T(\frac{A_2(\beta)}{A_1(\beta)} - \frac{A_2(0)}{A_1(0)}) \\ &\quad = T(\frac{1}{A_1(\beta)} - \frac{1}{A_1(0)})A_2(\beta) + \frac{T}{A_1(0)}((A_2(\beta) - A_2(0)) \end{aligned}$$

Using (2.13) and (2.14) in DD, we have

$$\frac{T}{A_1(\beta)} - \frac{T}{A_1(0)} \xrightarrow{P} 0,$$

and

(2.10)
$$\frac{T}{A_1(0)} \xrightarrow{P} \frac{1}{2\sigma^2},$$

Also, from (2.6) - (2.9)

(2.11)
$$A_2(\beta) \xrightarrow{d} 2\sigma^2\mathbf{C}'_\gamma(\beta)B^{-1}(\beta)\mathbf{C}_\gamma(\beta)$$

and hence

$$T(\frac{1}{A_1(\beta)} - \frac{1}{A_1(0)})A_2(\beta) \xrightarrow{P} 0.$$

Now applying Theorem 2.1 in DCD, we have

(2.12)
$$\sum_{t=1}^{T}\log\frac{1-d_t}{1+2p_T d_t} + T\log\frac{A_1(0)}{A_1(\beta)} \xrightarrow{d} Z_\gamma(\beta),$$

which combined with (2.10) - (2.12) implies

$$L_T(\hat{\boldsymbol{\delta}}(\beta),\beta) - L_T(\hat{\boldsymbol{\delta}}(0),0) \xrightarrow{d} Z_\gamma(\beta) + \mathbf{C}'_\gamma(\beta)B^{-1}(\beta)\mathbf{C}_\gamma(\beta) - \mathbf{C}'_\gamma(0)B^{-1}(0)\mathbf{C}_\gamma(0),$$

which proves (ii). □

# 3 Examples

In this section, Theorem 2.1 is applied to two important cases: an MA(1) model with non-zero mean and an MA(1) model with a polynomial trend. In the former, the pile-up effect turns out to be extreme. We also illustrate the theory on the overshort data contained in Brockwell and Davis (1996).

MA(1) model with non-zero mean: Let $Y_t$ be the MA(1) process with mean $b$ given by the equations,

$$Y_t = b + \epsilon_t - \theta\epsilon_{t-1}, \quad \epsilon_t \sim \text{IID}(0,\sigma^2).$$

In this case, $x_{t1} = 1$ so that $x_{t1} = f_1(t) = 1$ and

$$\begin{aligned} a_{1t,T} = \sum_{s=1}^{T} f_1(s)P_{st} = \sqrt{\frac{2}{T+1}}\sum_{s=1}^{T}\sin\left(\frac{\pi st}{T+1}\right) \quad &\sim \quad \frac{\sqrt{2(T+1)}}{\pi t}\int_0^{\pi t}\sin x dx \\ &= \quad \frac{\sqrt{2(T+1)}}{\pi t}(1-\cos\pi t). \end{aligned}$$

Hence,

$$\begin{aligned} T^{-\frac{1}{2}}a_{1t,T} \quad &\rightarrow \quad \begin{cases} 0, & \text{if t is even,} \\ \frac{2\sqrt{2}}{\pi t}, & \text{if t is odd,} \end{cases} \\ &=: \quad a^*_{1t}\,. \end{aligned}$$

Let $\delta = T^{\frac{3}{2}}(b - b_0)$, where $b_0$ is the true mean. Then from Theorem 2.1, we have

$$\begin{aligned} \hat{\delta}(\beta) \quad &\xrightarrow{d} \quad \frac{\sqrt{2}\sigma}{4}\left(\sum_{t=1}^{\infty}\frac{1}{(\pi^2(2t-1)^2+\beta^2)\pi^2(2t-1)^2}\right)^{-1} \\ &\qquad \times\left(\sum_{t=1}^{\infty}\frac{\sqrt{\pi^2(2t-1)^2+\gamma^2}X_{2t-1}}{(\pi^2(2t-1)^2+\beta^2)\pi(2t-1)}\right) \\ &=: \quad \tilde{\delta}(\beta) \end{aligned}$$

and

$$\begin{aligned} L(\hat{\delta},\beta) \quad &- \quad L(\hat{\delta},0) \\ &\rightarrow \sum_{t=1}^{\infty}\ln\frac{\pi^2t^2}{\pi^2t^2+\beta^2} + \sum_{t=1}^{\infty}\frac{\beta^2(\pi^2t^2+\gamma^2)X_t^2}{(\pi^2t^2+\beta^2)\pi^2t^2} \\ &\quad + \left(\sum_{t=1}^{\infty}\frac{\sqrt{\pi^2(2t-1)^2+\gamma^2}X_{2t-1}}{(\pi^2(2t-1)^2+\beta^2)\pi(2t-1)}\right)^2\left(\sum_{t=1}^{\infty}\frac{1}{(\pi^2(2t-1)^2+\beta^2)\pi^2(2t-1)^2}\right)^{-1} \\ &\quad - \left(\sum_{t=1}^{\infty}\frac{\sqrt{\pi^2(2t-1)^2+\gamma^2}X_{2t-1}}{\pi^3(2t-1)^3}\right)^2\left(\sum_{t=1}^{\infty}\frac{1}{\pi^4(2t-1)^4}\right)^{-1}. \end{aligned}$$

In the special case that $\theta_0 = 1$ ($\gamma = 0$) is assumed known and unestimated, then

$$\begin{aligned}
\hat{\delta}(0) &= T^{\frac{3}{2}}(\hat{b} - b_0) \\
&\xrightarrow{d} \frac{\sqrt{2}\sigma}{4}\left(\sum_{t=1}^{\infty}\frac{X_{2t-1}}{\pi^2(2t-1)^2}\right)\left(\sum_{t=1}^{\infty}\frac{1}{\pi^4(2t-1)^4}\right)^{-1} \\
&\sim N(0, \sigma^2\left(4\sum_{t=1}^{\infty}\frac{1}{\pi^4(2t-1)^4}\right)^{-1}) \\
&= N(0, 12\sigma^2),
\end{aligned}$$

whereas if $\theta$ is estimated but $\theta_0 = 1$,

$$\begin{aligned}
\hat{\delta}(\hat{\beta}_{LM}) \xrightarrow{d} & \frac{\sqrt{2}\sigma}{4}\left(\sum_{t=1}^{\infty}\frac{1}{(\pi^2(2t-1)^2+\tilde{\beta}_{LM}^2)\pi^2(2t-1)^2}\right)^{-1} \\
& \cdot\sum_{t=1}^{\infty}\frac{X_{2t-1}}{(\pi^2(2t-1)^2+\tilde{\beta}_{LM}^2)}.
\end{aligned}$$

In order to get a feel for the limiting behavior of $\hat{\delta}(\beta)$, 10,000 replicates of the limiting random variable $\tilde{\delta}(\beta)$ were generated for the case $\gamma = 0$. Figure 1 shows that the variance of $\tilde{\delta}(\beta)$ increases with $\beta$. We break out the distribution of $\tilde{\delta}(\tilde{\beta}_{LM})$ into two pieces corresponding to the cases $\tilde{\beta}_{LM} = 0$ and $\tilde{\beta}_{LM} > 0$. In the former, $\tilde{\delta}(0)$ has a $\text{N}(0, 12\sigma^2)$ distribution. The sample variances of $\tilde{\delta}(\tilde{\beta}_{LM})$ for the two cases are 11.993 and 17.077, respectively. Estimated densities for these two cases are displayed in Figure 2. (The estimate for $\tilde{\delta}(\tilde{\beta}_{LM})$ when $\tilde{\beta}_{LM} > 0$ is based on far fewer points than in the $\tilde{\beta}_{LM} = 0$ case which results in a more ragged graph.) As seen from Figure 2, the density functions of $\tilde{\delta}(\tilde{\beta}_{LM})$ in the two cases can be reasonably approximated by the densities of a $N(0, 12\sigma^2)$ and a $N(0, 17.077\sigma^2)$ distribution, respectively. Since $\tilde{\beta}_{LM}$ is equal to 0 with probability 0.9692 (see Table 3.1 below), the limiting density function of $\hat{\delta}(\hat{\beta}_{LM})$ tends to resemble a normal density with variance that is a mixture of the two variances, i.e., $0.9692 \times 11.993 + (1 - 0.9692) \times 17.077 = 12.145$. This variance is close to the sample variance (12.151) based on the 10,000 replicates of $\tilde{\delta}(\tilde{\beta}_{LM})$.

| $\gamma$ | $P(\hat{\beta}_{LM} = 0)$ | $P(\hat{\beta}_{MLE} = 0)$ | $P(\tilde{\beta}_{LM,\gamma} = 0)$ | $P(\tilde{\beta}_{MLE,\gamma} = 0)$ |
|---|---|---|---|---|
| 0.00 | .9715 | .9541 | .9692 | .9547 |
| 0.50 | .9715 | .9510 | .9683 | .9540 |
| 1.00 | .9700 | .9503 | .9664 | .9507 |
| 2.50 | .9568 | .9318 | .9524 | .9335 |
| 5.00 | .9053 | .8657 | .8968 | .8684 |
| 10.00 | .6893 | .6294 | .6852 | .6312 |
| 15.00 | .4603 | .3913 | .4646 | .3875 |

**Table 3.1.** *Probabilities of $\hat{\beta}_{LM}$, $\hat{\beta}_{MLE}$, $\tilde{\beta}_{LM}$, and $\tilde{\beta}_{MLE}$ being 0. For $\hat{\beta}_{LM}$ and $\hat{\beta}_{MLE}$, the sample size $T = 50$.*

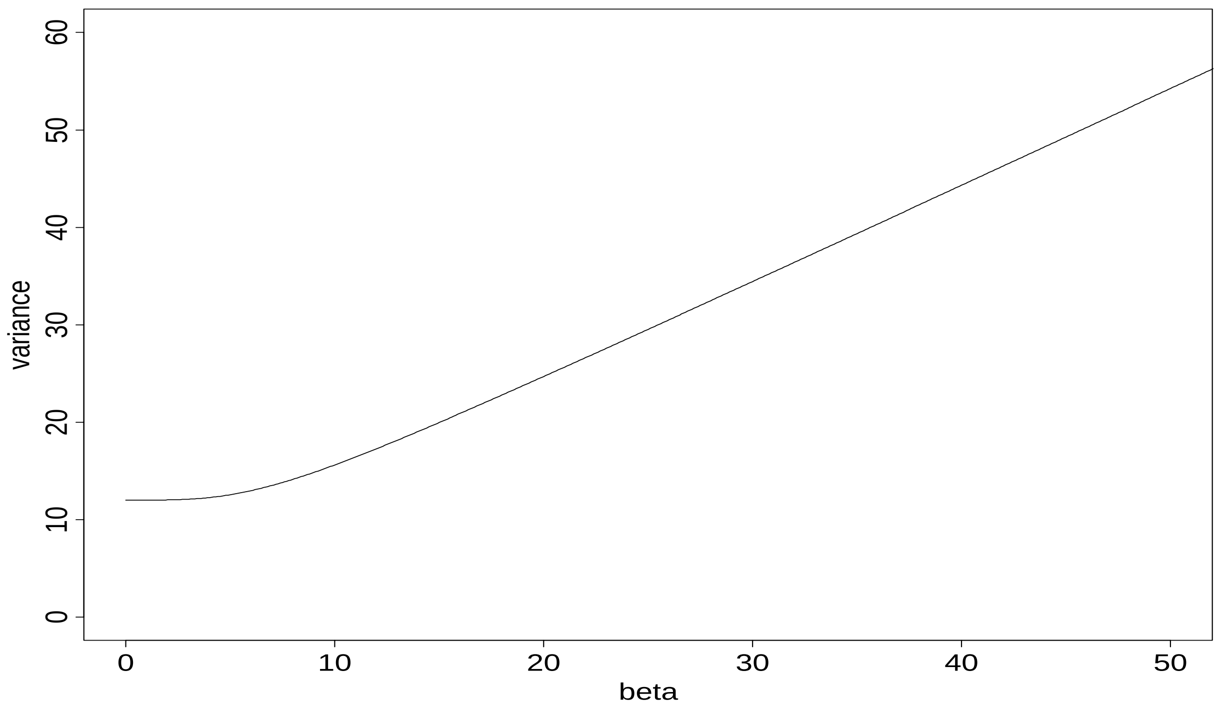


Figure 1: *The variance of* $\boldsymbol{\delta}(\beta)$ *for* $\gamma = 0$ *and* $\sigma^2 = 1$.

Table 3.1 provides a summary of the pile-up probabilities of $\hat{\beta}_{LM}$ and $\hat{\beta}_{MLE}$ for a variety of $\gamma$ values. As expected the pile-up probabilities of the local maximum estimate are larger than the analogous probabilities for the global maximum. This disparity is much larger than in the mean zero case considered in DCD (see Table 3.1), and the pile-up probabilities of both are quite high even for large values of $\gamma$. These results demonstrate that the severity of the pile-up effect for the non-zero mean model.

**Example (overshort data):** We now apply the foregoing results to the the overshort data discussed in Examples 3.2.8 and 6.3.2 of Brockwell and Davis (1996). A graph of the data and its sample ACF is displayed in Figure 3. The value of the series $Y_t$ represents the net change in the contents of an underground fuel tank after adjusting for the amount of fuel sold and delivered during day $t$. The ACF of the data strongly suggests modelling the data as an MA(1) model with a non-constant mean and possibly a unit root. The appropriateness of an MA(1) model for the data can also be argued on a structural model formulation as described in Example 3.2.8 of Brockwell and Davis (1996). Both the mean and moving average parameter have physical interpretations and inferences about both parameters are desired. The mean $b$ corresponds to the daily "leak rate" of

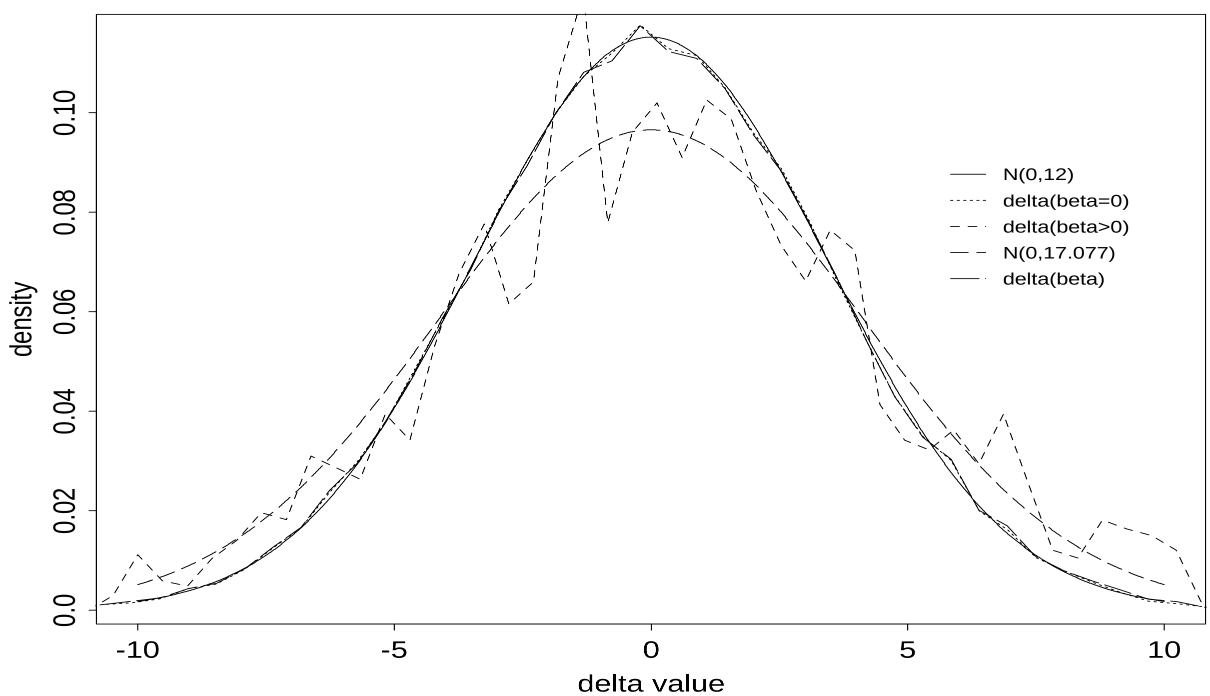


Figure 2: *The estimated density function of* $\tilde{\boldsymbol{\delta}}(\beta)$ *for* $\gamma = 0$ *and* $\sigma^2 = 1$*, based on 10,000 replications*

the tank while the moving average parameter $\theta$ is related to the signal to noise ratio. Here the noise refers to measurement error in recording the actual number of gallons of fuel in the tank. A value of $\theta$ near 1 indicates no measurement error, while a value of $\theta$ far from 1 suggests substantial measurement error. The maximum likelihood estimator, $\hat{b}_{GLS}$, of $b$ for $\theta_0 = 1$ is -5.025 with an asymptotic standard error of $\sqrt{12\sigma^2/(57)^3} = .3531$. This is in close agreement with the actual standard error of $\hat{b}_{GLS}$ given by .3440. Now the maximum likelihood estimates of $b, \theta$ and $\sigma^2$ are -4.781, .8478, and 2019.81, respectively. Using the $\mathrm{N}(0, 17.077\sigma^2/T^3)$ distribution as an approximation to $\hat{b}_{MLE}$ when $\theta_0 = 1$, the resulting asymptotic standard error is .4316, which is nearly 30% larger than the case when $\theta_0$ is known to be 1. On the other hand, if one applied standard asymptotic theory to the case of an invertible MA(1) model, then the estimated standard error of $\hat{b}_{MLE}$ is $\sqrt{(1 - .8478)^2 2019.81/57} = .9060$, nearly double the standard error when $\theta_0 = 1$. These calculations illustrate the enormous impact that assumptions about $\theta_0$ have on the magnitude of the standard error of $\hat{b}_{MLE}$.

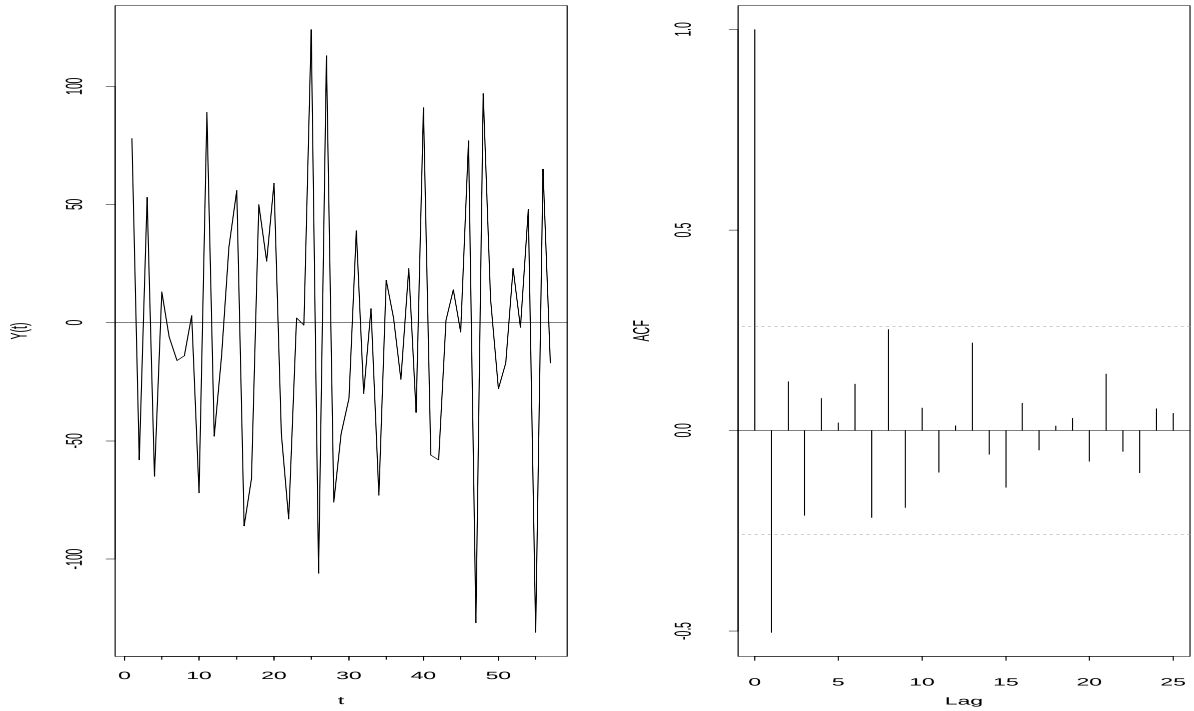


Figure 3: *Time series and ACF plots of overshort data.*

Polynomial trend: Let $Y_1, \ldots, Y_T$ observations from the MA(1) model with a linear trend given by

$$Y_t = b_0 + b_1 t + \cdots + b_{k-1} t^{k-1} + U_t,$$

where

$$U_t = \epsilon_t - \theta_0 \epsilon_{t-1}.$$

In this case,

$$a_{1t,T} = \sqrt{\frac{2}{T+1}} \sum_{s=1}^{T} \sin\left(\frac{\pi s t}{T+1}\right)$$

which, as found in the constant mean case, satisfies

$$T^{-\frac{1}{2}} a_{1t,T} \quad \to \quad a^*_{1t} = \begin{cases} 0, & \text{if t is even,} \\ \frac{2\sqrt{2}}{\pi t}, & \text{if t is odd.} \end{cases}$$

Similarly, for $i = 2, 3, \ldots, k$

$$\begin{aligned} a_{it,T} &= \sqrt{\frac{2}{T+1}} \sum_{s=1}^{T} s^{i-1} \sin\left(\frac{\pi s t}{T+1}\right) \\ &\sim \frac{\sqrt{2} T^{i-1/2}}{(\pi t)^i} \int_0^{\pi t} x^{i-1} \sin x dx, \end{aligned}$$

so that

$$\begin{aligned} T^{1/2-i} a_{it,T} &\to \frac{\sqrt{2}}{(\pi t)^i} \int_0^{\pi t} x^{i-1} \sin x dx \\ &=: a_{it}^*. \end{aligned}$$

For a linear trend, i.e., $k = 2$, we have

$$D_T = \mathrm{diag}(T^{-\frac{3}{2}}, T^{-\frac{5}{2}})$$

and

$$\mathbf{a}_t^{*'} = \begin{cases} [0, -\sqrt{2}(\pi t)^{-1}], & \text{if t is even}, \\ [2\sqrt{2}(\pi t)^{-1}, \sqrt{2}(\pi t)^{-1}], & \text{if t is odd}, \end{cases}$$

$$\mathbf{a}_t^* \mathbf{a}_t^{*'} = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 2(\pi t)^{-2} \end{bmatrix}, & \text{if t is even}, \\ \begin{bmatrix} 8(\pi t)^{-2} & 4(\pi t)^{-2} \\ 4(\pi t)^{-2} & 2(\pi t)^{-2} \end{bmatrix}, & \text{if t is odd}. \end{cases}$$

If $\mathbf{b} = \mathbf{b}_0 + D_T \boldsymbol{\delta}$, where $\mathbf{b}_0$ is the true parameter vector, then

$$\hat{\boldsymbol{\delta}}(\beta) \to \sigma B^{-1}(\beta) \mathbf{C}_\gamma(\beta),$$

where

$$\begin{aligned} B_{1,1}(\beta) &= 8 \sum_{t=1}^{\infty} \frac{1}{(\pi^2(2t-1)^2 + \beta^2)\pi^2(2t-1)^2}, \\ B_{1,2}(\beta) &= B_{2,1}(\beta) \\ &= 4 \sum_{t=1}^{\infty} \frac{1}{(\pi^2(2t-1)^2 + \beta^2)\pi^2(2t-1)^2}, \\ B_{2,2}(\beta) &= 2 \sum_{t=1}^{\infty} \frac{1}{(\pi^2 t^2 + \beta^2)\pi^2 t^2}, \\ \mathbf{C}_{\gamma,1}(\beta) &= 2\sqrt{2} \sum_{t=1}^{\infty} \frac{\sqrt{\pi^2(2t-1)^2 + \gamma^2} X_{2t-1}}{(\pi^2(2t-1)^2 + \beta^2)\pi(2t-1)}, \end{aligned}$$

and

$$\mathbf{C}_{\gamma,2}(\beta) = \sqrt{2} \sum_{t=1}^{\infty} \left( \frac{\sqrt{\pi^2(2t-1)^2 + \gamma^2} X_{2t-1}}{(\pi^2(2t-1)^2 + \beta^2)\pi(2t-1)} - \frac{\sqrt{4\pi^2 t^2 + \gamma^2} X_{2t}}{2(4\pi^2 t^2 + \beta^2)\pi t} \right)$$

In the special case when $\theta_0 = 1$ ($\gamma = 0$) is known, i.e., $\beta = 0$, then

$$B(0) = \begin{bmatrix} \frac{1}{12} & \frac{1}{24} \\ \frac{1}{24} & \frac{1}{45} \end{bmatrix},$$

$$B^{-1}(0) = \begin{bmatrix} 192 & -360 \\ -360 & 720 \end{bmatrix},$$

and

$$\mathbf{C}_0(0) = \begin{bmatrix} 2\sqrt{2}\sum_{t=1}^{\infty} \frac{X_{2t-1}}{\pi^2(2t-1)^2} \\ \sqrt{2}\sum_{t=1}^{\infty} \left(\frac{X_{2t-1}}{\pi^2(2t-1)^2} - \frac{X_{2t}}{4\pi^2 t^2}\right) \end{bmatrix}.$$

These imply

$$\begin{aligned} \hat{\boldsymbol{\delta}}(0) &\to \sigma B^{-1}(0)\mathbf{C}_0(0) \\ &= N(\mathbf{0}, \sigma^2 B^{-1}(0)). \end{aligned}$$

For the case of a quadratic trend, we have

$$D_T = \text{diag}(T^{-\frac{3}{2}}, T^{-\frac{5}{2}}, T^{-\frac{7}{2}})$$

$$\mathbf{a}_t^{*'} = \begin{cases} [0, -\sqrt{2}(\pi t)^{-1}, -\sqrt{2}(\pi t)^{-1}], & \text{if t is even,} \\ [2\sqrt{2}(\pi t)^{-1}, \sqrt{2}(\pi t)^{-1}, \sqrt{2}(\pi t)^{-3}(-4+\pi^2 t^2)], & \text{if t is even.} \end{cases}$$

$$\mathbf{a}_t^{*}\mathbf{a}_t^{*'} = \begin{cases} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2(\pi t)^{-2} & 2(\pi t)^{-2} \\ 0 & 2(\pi t)^{-2} & 2(\pi t)^{-2} \end{bmatrix}, & \text{if t is even,} \\ \begin{bmatrix} 8(\pi t)^{-2} & 4(\pi t)^{-2} & 4(\pi t)^{-4}(-4+\pi^2 t^2) \\ 4(\pi t)^{-2} & 2(\pi t)^{-2} & 2(\pi t)^{-4}(-4+\pi^2 t^2) \\ 4(\pi t)^{-4}(-4+\pi^2 t^2) & 2(\pi t)^{-4}(-4+\pi^2 t^2) & 2(\pi t)^{-6}(-4+\pi^2 t^2)^2 \end{bmatrix}, & \text{if t is odd.} \end{cases}$$

If $\mathbf{b} = \mathbf{b}_0 + D_T\boldsymbol{\delta}$, then

$$\hat{\boldsymbol{\delta}}(\beta) \to \sigma B^{-1}(\beta)\mathbf{C}_\gamma(\beta),$$

where

$$\begin{aligned} B_{1,3}(\beta) &= B_{3,1}(\beta) \\ &= 4\sum_{t=1}^{\infty} \frac{-4+\pi^2(2t-1)^2}{(\pi^2(2t-1)^2+\beta^2)\pi^4(2t-1)^4}, \\ B_{2,3}(\beta) &= B_{3,2}(\beta) \\ &= \sum_{t=1}^{\infty}\left(\frac{1}{2(4\pi^2t^2+\beta^2)\pi^2t^2} + \frac{2(-4+\pi^2(2t-1)^2)}{(\pi^2(2t-1)^2+\beta^2)\pi^4(2t-1)^4}\right), \\ B_{3,3}(\beta) &= \sum_{t=1}^{\infty}\left(\frac{1}{2(4\pi^2t^2+\beta^2)\pi^2t^2} + \frac{2(-4+\pi^2(2t-1)^2)^2}{(\pi^2(2t-1)^2+\beta^2)\pi^6(2t-1)^6}\right), \end{aligned}$$

and

$$\mathbf{C}_{\gamma,3}(\beta) = \sqrt{2}\sum_{t=1}^{\infty}\left(\frac{(\pi^2(2t-1)^2-4)\sqrt{\pi^2(2t-1)^2+\gamma^2}X_{2t-1}}{(\pi^2(2t-1)^2+\beta^2)\pi(2t-1)} - \frac{\sqrt{\pi^2t^2+\gamma^2}X_{2t}}{(4\pi^2t^2+\beta^2)2\pi t}\right)$$

and the other elements are the same as in the $k = 2$ case considered above.

Again in the special case when $\theta_0 = 1$ ($\gamma = 0$) and $\theta$ is not estimated, i.e., $\beta = 0$, then

$$B(0) = \begin{bmatrix} \frac{1}{12} & \frac{1}{24} & \frac{1}{40} \\ \frac{1}{24} & \frac{1}{45} & \frac{1}{72} \\ \frac{1}{40} & \frac{1}{72} & \frac{1}{112} \end{bmatrix},$$

$$B^{-1}(0) = \begin{bmatrix} 1200 & -5400 & 5040 \\ -5400 & 25920 & -25200 \\ 5040 & -25200 & 25200 \end{bmatrix},$$

and

$$\mathbf{C}_0(0) = \begin{bmatrix} 2\sqrt{2}\sum_{t=1}^{\infty}\frac{X_{2t-1}}{\pi^2(2t-1)^2} \\ \sqrt{2}\sum_{t=1}^{\infty}\left(\frac{X_{2t-1}}{\pi^2(2t-1)^2} - \frac{X_{2t}}{8\pi^2t^2}\right) \\ \sqrt{2}\sum_{t=1}^{\infty}\left(\frac{(\pi^2(2t-1)^2-4)X_{2t-1}}{\pi^2(2t-1)^2} - \frac{X_{2t}}{4\pi^2t^2}\right) \end{bmatrix}.$$

by using

$$\sum_{t=1}^{\infty}(2t-1)^{-4} = \frac{\pi^4}{96},$$
$$\sum_{t=1}^{\infty}(2t-1)^{-6} = \frac{\pi^6}{960},$$
$$\sum_{t=1}^{\infty}(2t-1)^{-8} = \frac{17\pi^8}{161280},$$

and

$$\sum_{t=1}^{\infty}t^{-4} = \frac{\pi^4}{90}$$

These imply

$$\begin{aligned}\hat{\boldsymbol{\delta}}(0) &\to \sigma B^{-1}(0)\mathbf{C}_0(0) \\ &= N(\mathbf{0}, \sigma^2 B^{-1}(0)).\end{aligned}$$

# 4 Testing of $H_0 : \theta = 1$ in MA(1) Model with Non-zero Mean

We now turn to the problem of testing $H_0 : \theta_0 = 1$ versus $H_A : \theta_0 < 1$ in the MA(1) model with non-zero mean. Let $Y_1, \cdots, Y_T$ be observations from

$$Y_t = b_0 + \epsilon_t - \theta_0\epsilon_{t-1}.$$

The asymptotic theory of Section 2 allows us to construct tests of $H_0$ using $\hat{\beta}_{LM}$ and the likelihood ratio statistic.

Type I error probabilities and the power functions for testing the null hypothesis that $H_0 : \theta_0 = 1$ versus $H_A : \theta_0 < 1$ for this model can easily be computed via simulation. Tests based on $\hat{\beta}_{LM}$, the likelihood ratio statistic and Tanaka's score statistic are considered here. The asymptotic theory of Section 2 also allows us to approximate the nominal power of tests based on $\hat{\theta}_{LM}$ and LR test against local alternatives of the form:

$$H_A : \theta = \theta_A,$$

where $\theta_A = 1 - \gamma/T$. The tests considered here all have asymptotic power equal to 1 against any fixed local alternative.

For the test based on the LM point estimate, let $b_{LM}(\alpha)$ be the $(1-\alpha)$th quantile defined by

$$P(\tilde{\beta}_{LM,0} > b_{LM}(\alpha)) = \alpha.$$

The generalized likelihood ratio test is based on the likelihood ratio given by

$$\hat{L}_T = L_T(\hat{\boldsymbol{\delta}}(\hat{\beta}_{MLE}), \hat{\beta}_{MLE}) - L_T(\hat{\boldsymbol{\delta}}(0), 0)$$

Let $\tilde{W}_\gamma = W_\gamma(\tilde{\beta}_{MLE,\gamma})$ denote the limit random variable of $\hat{L}_T$ when $\gamma$ is the true value. The $(1-\alpha)$th asymptotic quantile $b_{LR}(\alpha)$ is defined as:

$$P(\tilde{W}_0 > b_{LR}(\alpha)) = \alpha.$$

In the results to follow the tests are defined using these asymptotic quantiles to define the critical regions. For example the LR rejects $H_0$ at level $\alpha$ if $\hat{\beta}_{LM} = T(1 - \hat{\theta}_{LM}) > b_{LM}(\alpha)$ or, equivalently, if $\hat{\theta}_{LM} < 1 - b_{LM}(\alpha)/T$.

The values of $b_{LR}(\alpha)$ and $b_{LM}(\alpha)$ are found using the asymptotic results of Section 2. The infinite sums required in the limiting process $W_\gamma(\beta)$ and $W'_\gamma(\beta)$ are approximated by truncating them at $k = 1000$. For all results reported below, 10,000 replications were used. For each replicate the two statistics were evaluated thereby reducing the between replicate variability as a component in the comparison of the three methods. For finite sample results further details on the methods used to compute the likelihood based estimates are given in DD. Probabilities reported below are accurate to $\pm 0.01$ with 95% confidence. The standard errors displayed in Table 4.1 were based on estimates of the asymptotic variance of the sample $p$th quantile given by $\frac{p(1-p)}{N\hat{f}^2(\hat{\xi}_p)}$, where $\hat{\xi}_p$ is the sample $p$th quantile, $\hat{f}(\cdot)$ is the estimated pdf computed using the function DENSITY in SPLUS, and $N$ is the number of replications (10,000 in our case).

| | $\alpha$ | |
|---|---|---|
| | .01 | .025 |
| $b_{LM}(\alpha)$ | 9.01 (.180) | 5.80 (.301) |
| $b_{LR}(\alpha)$ | 1.67 (.138) | 0.63 (.069) |

**Table 4.1.** $(1-\alpha)$ quantiles for the distribution of $\tilde{\beta}_{LM,0}$ and $\tilde{W}_0$.

Table 4.2 compares the power of the tests based on LM and the LR derived above for the finite sample. These power values are plotted in Figure 4. The LR test procedure exhibits slightly greater power than the test based on the local maximum estimate. They have almost identical power curves for values of $\gamma$ near 0 but the LR test dominates for $\gamma$'s in the range of 7–15. Table 4.3 compares the limiting powers of the LM and LR tests. The LR test has slightly higher power than the LM test (also see Figure 5), however, the differences are small. Table 4.4 compares the finite sample powers with those using the asymptotic approximations for the LR test. In both cases the actual finite sample values almost gives slightly higher values than the asymptotic values for $\gamma < 5$ (i.e., $\theta > 0.9$, here $\theta = 1 - \gamma/50$).

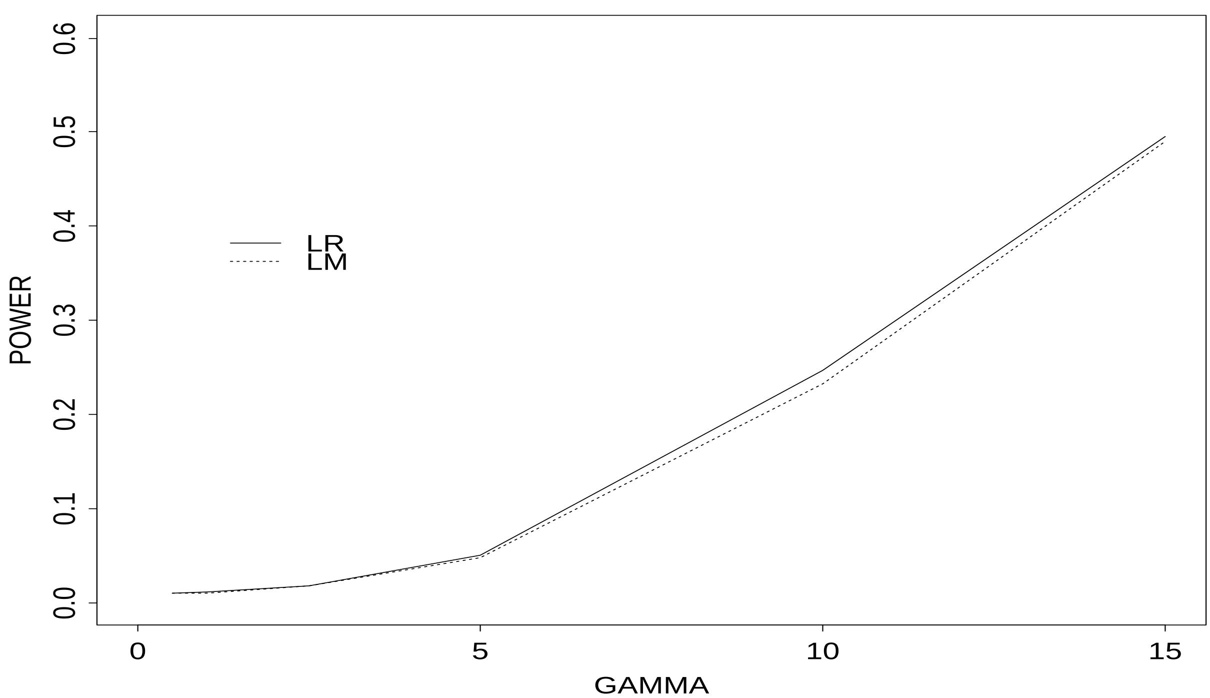


Figure 4: *Power curves of tests based on LR and LM for $T = 50$.*

| $\gamma$ | $P(\hat{\beta}_{LM} > b_{LM}(\alpha)\|\gamma)$ | $P(\hat{L}_T > b_{LR}(\alpha)\|\gamma)$ |
|---|---|---|
| 0.50 | 0.0100 | 0.0099 |
| 1.00 | 0.0102 | 0.0113 |
| 2.50 | 0.0181 | 0.0184 |
| 5.00 | 0.0481 | 0.0505 |
| 10.0 | 0.2321 | 0.2468 |
| 15.0 | 0.4896 | 0.4954 |

**Table 4.2** Comparison of the finite sample ($T = 50$) power of the LM test with the LR test for $\alpha = 0.01$.

| $\gamma$ | $P(\tilde{\beta}_{LM,\gamma} > b_{LM}(\alpha))$ | $P(\tilde{W}_\gamma > b_{LR}(\alpha))$ |
|---|---|---|
| 0.50 | 0.0106 | 0.0101 |
| 1.00 | 0.0116 | 0.0111 |
| 2.50 | 0.0173 | 0.0181 |
| 5.00 | 0.0468 | 0.0507 |
| 10.0 | 0.2260 | 0.2359 |
| 15.0 | 0.4853 | 0.4853 |

**Table 4.3** Comparison of the limiting power of LM and LR tests for $\alpha = 0.01$.

| | | $\alpha$ | |
|---|---|---|---|
| $\theta_A$ | | .01 | .025 |
| .7 | Exact | .4954 | .5721 |
| | Limit | .4853 | .5572 |
| .8 | Exact | .2468 | .3237 |
| | Limit | .2359 | .3088 |
| .9 | Exact | .0505 | .0892 |
| | Limit | .0507 | .0895 |
| .95 | Exact | .0184 | .0377 |
| | Limit | .0181 | .0398 |
| .99 | Exact | .0099 | .0251 |
| | Limit | .0101 | .0254 |

**Table 4.4** Comparison of the limiting power and exact power of the LR test using quantiles from Table 4.1 with sample of size 50. Exact values are computed via simulation.

Figures 4 and 5 show that for the purposes of testing the null hypothesis that $\theta = 1$ it is better to use the LR test for all alternatives. For this reason we will now turn to a comparison of the LR test and the LBIU test of Tanaka when $\alpha = 0.01, 0.05$, and 0.1. Due to the large pile-up effect, the

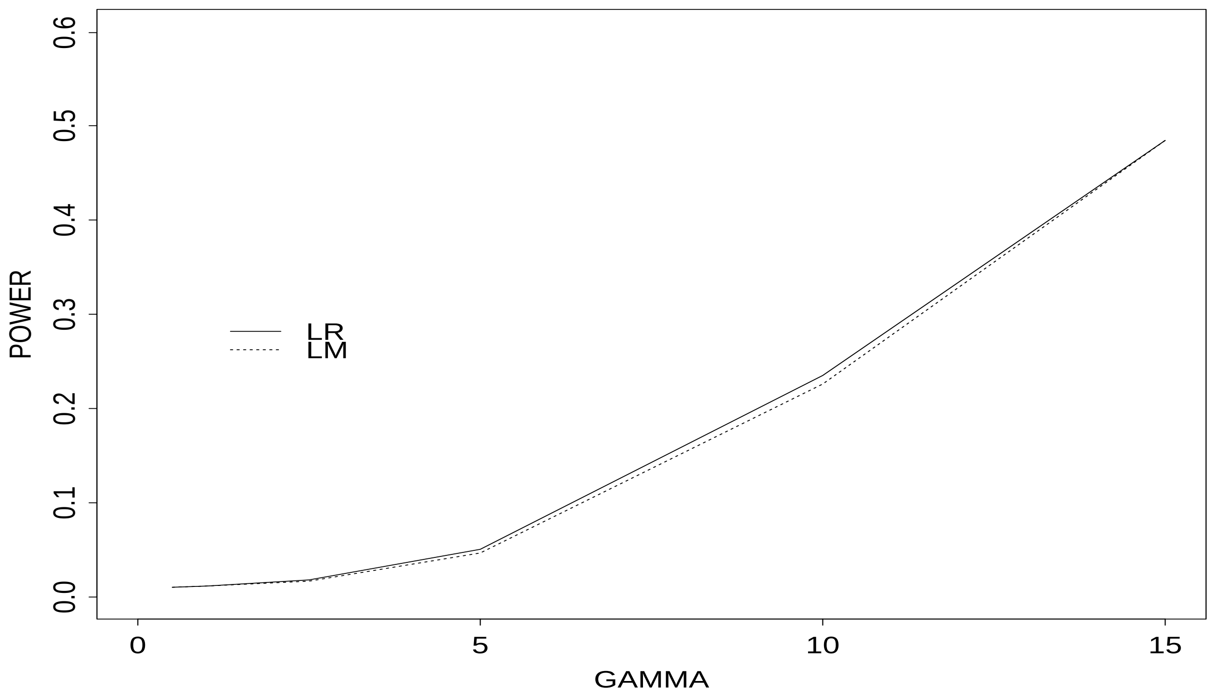


Figure 5: *Limiting power curves of tests based on LR and LM.*

0.95 and 0.9 quantiles of the $\tilde{W}_\gamma$ are zero. Therefore, we use a randomized test to achieve a test based on the LR test with type I errors = 0.05 and 0.1. The critical function $\phi$ is defined by

$$\phi(\tilde{W}_\gamma) = \begin{cases} 1, & \text{if } \tilde{W}_\gamma > 0, \\ C_\alpha, & \text{if } \tilde{W}_\gamma = 0, \end{cases}$$

where $C_\alpha$ is chosen such that

$$\begin{aligned} E(\phi(\tilde{W}_0)) &= P(\tilde{W}_0 > 0) + C_\alpha P(\tilde{W}_0 = 0) \\ &= 0.0453 + C_\alpha(0.9547) \\ &= \alpha. \end{aligned}$$

Solving for $C_\alpha$, we obtain $C_{0.05} = 0.0049$ and $C_{0.1} = 0.0573$.

Table 4.5 compares the limiting power of the LBIU test (see Table 10.3 in Tanaka (1996)) with the limiting power of the LR test for $\alpha$ =0.01, 0.05, and 0.1. For $\alpha = 0.01$, they have almost identical power values for $\gamma \leq 5$. Thereafter the LR test increasingly outperforms the LBIU test by a wide margin (see Figure 6). For $\alpha = 0.05$, the differences are small except when $\gamma = 20$. But for $\alpha = 0.1$, the LBIU dominates the power function for the LR test. This is due to the large randomized component of the LR test required to achieve a significance level of 0.1.

| | 0.01 | | 0.05 | | 0.1 | |
|---|---|---|---|---|---|---|
| $\gamma$ | LBIU | LR | LBIU | LR | LBIU | LR |
| 1.00 | 0.0110 | 0.0111 | 0.0531 | 0.0540 | 0.1048 | 0.1038 |
| 5.00 | 0.0476 | 0.0507 | 0.1369 | 0.1359 | 0.2169 | 0.1814 |
| 10.0 | 0.2089 | 0.2359 | 0.3671 | 0.3719 | 0.4695 | 0.4050 |
| 20.0 | 0.6027 | 0.6924 | 0.7477 | 0.7844 | 0.8162 | 0.7957 |
| 50.0 | 0.9660 | 0.9920 | 0.9873 | 0.9944 | 0.9933 | 0.9947 |
| 60.0 | 0.9848 | 0.9969 | 0.9951 | 0.9974 | 0.9976 | 0.9975 |

**Table 4.5** Comparison of the limiting power of the LBIU test with the limiting power of the LR test for $\alpha = 0.01, 0.05$, and 0.1.

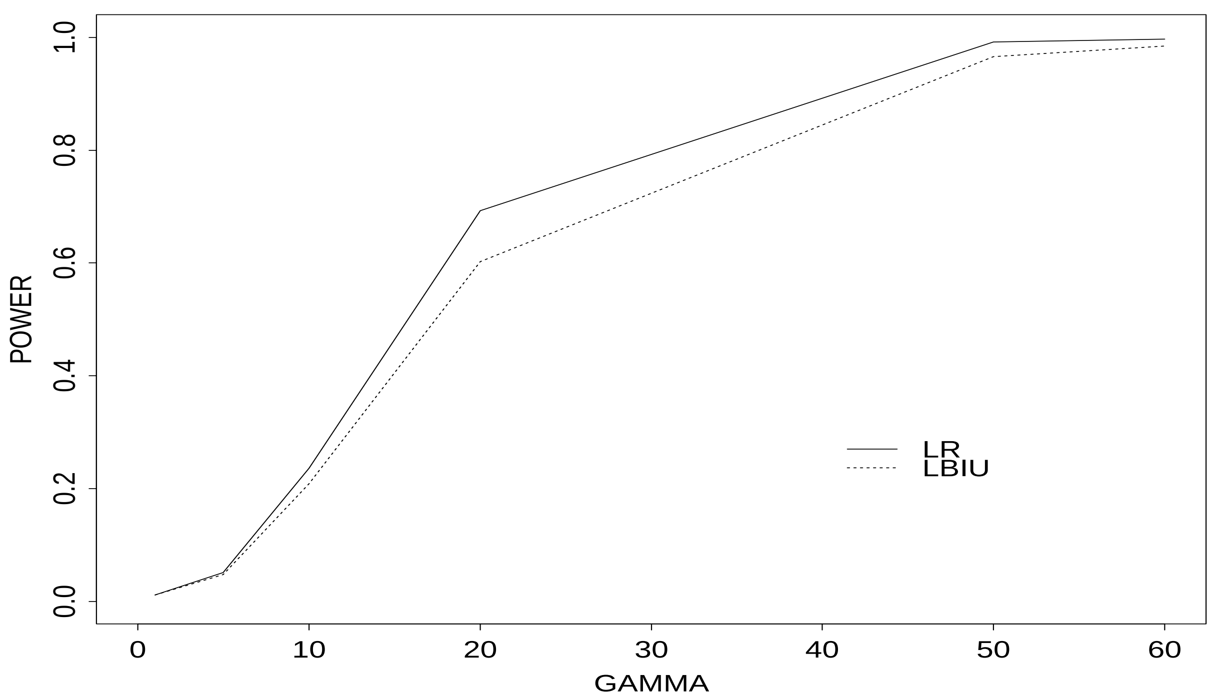


Figure 6: *The limiting power curves of tests based on LR and LM.*

<u>**Example (overshort data continued):**</u> For this data, $\tilde{\beta}_{LM} = 57(1 - .8478) = 8.675$, and the hypothesis $\theta_0 = 1$ is rejected at the .025 level but not at the .01 level. On the other hand, the likelihood ratio test statistic is .009 which corresponds to a $p$-value between .025 and .05.

## REFERENCES


Brockwell, P. J. and R. A. Davis (1996). *Introduction to Time Series and Forecasting* . Springer-Verlag, New York.

Davis, R. A., Chen, M. C. and W. T. M. Dunsmuir (1995). Inference for MA(1) processes with a root on or near the unit circle. *Probability and Mathematical Statistics* **15** 227–242.

Davis, R. A. and Dunsmuir, W.T.M. (1996). Maximum Likelihood Estimation for MA(1) Processes with a Root On or Near the Unit Circle. *Preprint.*

Pesaran, M. H. (1983). A note on the maximum likelihood estimation of regression models with first order moving average errors with roots on the unit circle. *Australian Journal of Statistics* **25**, 442-448.

Sargan, J. D. and A. Bhargava (1983). Maximum likelihood estimation of regression mode ls with first-order moving average errors when the root lies on the unit circle. *Econometrics* **51**, 799-820.

Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components. *J. American Statistical Association* **88** 590–595.

Tanaka, K. (1990). Testing for a Moving Average Unit Root. *Econometric Theory* **6** 433-444.

Tanaka, K. (1996). *Time Series Analysis : Nonstationary and Noninvertible Distribution Theory.* Wiley.