The Annals of Statistics
2011, Vol. 39, No. 6, 3062–3091
DOI: 10.1214/11-AOS935

UNIT ROOTS IN MOVING AVERAGES BEYOND FIRST ORDER[1]

BY RICHARD A. DAVIS AND LI SONG

Columbia University and Barclays Capital

The asymptotic theory of various estimators based on Gaussian likelihood has been developed for the unit root and near unit root cases of a first-order moving average model. Previous studies of the MA(1) unit root problem rely on the special autocovariance structure of the MA(1) process, in which case, the eigenvalues and eigenvectors of the covariance matrix of the data vector have known analytical forms. In this paper, we take a different approach to first consider the joint likelihood by including an augmented initial value as a parameter and then recover the exact likelihood by integrating out the initial value. This approach by-passes the difficulty of computing an explicit decomposition of the covariance matrix and can be used to study unit root behavior in moving averages beyond first order. The asymptotics of the generalized likelihood ratio (GLR) statistic for testing unit roots are also studied. The GLR test has operating characteristics that are competitive with the locally best invariant unbiased (LBIU) test of Tanaka for some local alternatives and dominates for all other alternatives.

1. Introduction. In this paper we consider inference for moving average models that possess one or more unit roots in the moving average polynomial. To introduce the problem, let's first consider the MA(1) model given by

$$X_t = Z_t - \theta_0 Z_{t-1}, \tag{1.1}$$

where $\theta_0 \in \mathbb{R}$, $\{Z_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbf{E}Z_t = 0, \mathbf{E}Z_t^2 = \sigma_0^2$ and density function f_Z. The MA(1) model is invertible if and only if $|\theta_0| < 1$, since in this case Z_t can be represented explicitly in terms of past values of X_t, that is,

$$Z_t = \sum_{j=0}^{\infty} \theta_0^j X_{t-j}.$$

Under this invertibility constraint, standard estimation procedures that produce asymptotically normal estimates are readily available. For example, if $\hat{\theta}$ represents the maximum likelihood estimator, found by maximizing the Gaussian likelihood based on the data $X_1, \dots, X_n$, then it is well known (see Brockwell and Davis [6])

Received June 2010; revised July 2011.
[1]Supported in part by NSF Grants DMS-07-43459 and DMS-11-07031.
MSC2010 subject classifications. 62M10.
Key words and phrases. Unit roots, moving average.

that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1 - \theta_0^2). \tag{1.2}$$

From the form of the limiting variance in (1.2), the asymptotic behavior of $\hat{\theta}$, let alone the scaling, is not immediately clear in the unit root case corresponding to $\theta_0 = 1$.

In the case f_Z is Gaussian, the parameters θ_0 and σ^2 are not identifiable without the constraint $|\theta_0| \le 1$. In particular, the profile Gaussian log-likelihood, obtained by concentrating out the variance parameter, satisfies

$$L_n(\theta) = L_n(1/\theta). \tag{1.3}$$

It follows that $\theta = 1$ is a critical value of the profile likelihood, and hence there is a positive probability that $\theta = 1$ is indeed the maximum likelihood estimator. If $\theta_0 = 1$, then it turns out that this probability does not vanish asymptotically (see, e.g., Anderson and Takemura [1], Tanaka [21] and Davis and Dunsmuir [10]). This phenomenon is referred to as the pile-up effect. For the case that $\theta_0 = 1$ or is near one in the sense that $\theta_0 = 1 + \gamma/n$, it was shown in Davis and Dunsmuir [10] that

$$n(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \xi_\gamma,$$

where ξ_γ is a random variable with a discrete component at 0, corresponding to the asymptotic pile-up effect, and a continuous component. Most of the early work on this problem was based on explicit knowledge of the eigenvectors and eigenvalues of the covariance matrix for observations from an MA(1) process; see Anderson and Takemura [1]. Recently, Breidt et al. [4] and Davis and Song [13] looked at model (1.1) under the Laplace likelihood and the Gaussian likelihood without resorting to knowledge of the precise form of eigenvectors and eigenvalues of the covariance matrix. Instead they introduced an auxiliary variable, which acts like an initial value and can be integrated out to form the likelihood.

With a couple exceptions, most of previous work dealt exclusively with the zero-mean case. Sargan and Bhargava [17] and Shephard [18] showed that for the nonzero mean case, the so-called pile-up effect is more severe than the zero mean case. Chen, Davis and Song [8] extended the results from Davis and Dunsmuir [10] to regression models with errors from a noninvertible MA(1) process. It is shown that, with a mean term present in the model, the pile-up probability goes up to more than 0.95.

The MA unit root problem can arise in many modeling contexts, especially if a time series exhibits trend and seasonality. For example, in personal communication, Richard Smith has mentioned the presence of a unit in modeling some environmental time series related to climate change [19]. After detrending and fitting an ARMA model to the time series, Smith noticed that the MA component appeared to have a unit root. One explanation for this phenomenon is that detrending often involves the application of a high-pass filter to the time series. In particular, the filter diminishes or obliterates any power in the time series at low frequencies (including the 0 frequency). Consequently, the detrended data will have a spectrum with 0 power at frequency 0, which can only be fitted with ARMA process that has a unit root in the MA component. While we only consider unit roots in higher order moving averages in this paper, we believe the techniques developed here will be applicable in a more general framework of an ARMA model. This will be the subject of future investigation.

In this paper, we will use the stochastic approaches described in [4] and [13] to first study the case when there is a regression component in the time series and errors are generated from noninvertible MA(1). A vital issue in extending these results to higher order MA models is the scaling required for the *auxiliary variable*. The scaling used for the regression problem in the MA(1) case provides insight into the way in which the auxiliary variable should be scaled in the higher order case. Quite surprisingly, when there is only one unit root in the MA(2) process, that is,

$$X_t = Z_t + c_1 Z_{t-1} + c_2 Z_{t-2}, \tag{1.4}$$

where $-c_1 - c_2 = 1$ and $\{Z_t\} \sim$ i.i.d. $(0, \sigma^2)$, the asymptotic distribution of the maximum likelihood estimator $(\hat{c}_1, \hat{c}_2)'$ is exactly the same as in invertible MA(2) case; see [6]. That is,

$$\sqrt{n}\begin{pmatrix} \hat{c}_1 - c_1 \\ \hat{c}_2 - c_2 \end{pmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \begin{bmatrix} 1 - c_2^2 & c_1(1 - c_2) \\ c_1(1 - c_2) & 1 - c_2^2 \end{bmatrix}\right). \tag{1.5}$$

One difference, however, is that $\hat{c}_1$ and $\hat{c}_2$ are now totally dependent asymptotically $[c_1(1 - c_2) = (1 - c_2)^2]$.

As seen from (1.3), the first derivative of the profile likelihood function is always 0 when $\theta = 1$. Therefore, the development of typical score tests or Wald tests is intractable in this case. Davis, Chen and Dunsmuir [9] used the asymptotic result from [10] to develop a test of $H_0 : \theta = 1$ based on $\hat{\theta}_{\mathrm{MLE}}$ and the generalized likelihood ratio. Interestingly, we will see that the estimator of the unit root in the MA(2) case has the same limit distribution as the corresponding estimator in the MA(1) case. Thus, we can extend the methods used in the MA(1) case to test for unit roots in the MA(2) case.

The paper is organized as follows. In Section 2, we demonstrate our method of proof applied to the MA(1) model with regression. This case plays a key role in the extension to higher order MAs. Section 3 contains the results for the unit root problem in the MA(2) case. In Section 4, we compare likelihood based tests with Tanaka's locally best invariant and unbiased (LBIU) test [20] for testing the presence of a unit root. It is shown that the likelihood ratio test performs quite well in comparison to the LBIU test. In Section 5, numerical simulation results are presented to illustrate the theory of Section 3. In Section 6, there is a brief discussion that connects the auxiliary variables in higher order MAs with terms in a regression model with MA(1) errors. Finally, in Section 7, the procedure for handling the MA(q) case with $q \geq 3$ is outlined. It is shown that the tools used in the MA(1) and MA(2) cases are still applicable and are, in fact, sufficient in dealing with higher order cases.

2. MA(1) with nonzero mean. In this section, we will extend the methods of Breidt et al. [4] and Davis and Song [13] to a regression model with MA(1) errors. These results turn out to have connections with the asymptotics in the higher order unit root cases (see Section 6). First, consider the model

$$X_t = \sum_{k=0}^{p} b_{k0} f_k(t/n) + Z_t - \theta_0 Z_{t-1}, \tag{2.1}$$

where $\{Z_t\}$ is defined as in (1.1), $\theta_0 = 1$, $b_{k0}, k = 0, \ldots, p$, are regression coefficients and $f_k(t/n), k = 0, \ldots, p$, are covariates at time t. Notice that the covariates $f_k(t/n)$ are also assumed to be functions on [0, 1]. Note that the detrended series $Y_t = X_t - \sum_{k=0}^{p} b_k f_k(t/n)$ has exactly the same likelihood as the one for the zero-mean case. As shown in [13], by concentrating out the scale parameter σ, maximizing the joint Gaussian likelihood is equivalent to minimizing the following objective function:

$$l_n(\vec{b}, \theta, z_{\text{init}}) = \sum_{t=0}^{n} z_t^2 \qquad \text{for } |\theta| \le 1, \tag{2.2}$$

where $\vec{b} = (b_0, \ldots, b_p)'$, $Z_{\text{init}} = Z_0$, and z_i is given by

$$\begin{aligned}
z_i &= Y_i + \theta Y_{i-1} + \cdots + \theta^{i-1} Y_1 + \theta^i z_{\text{init}} \\
&= \left(X_i - \sum_{k=0}^{p} b_k f_k(i/n)\right) + \theta\left(X_{i-1} - \sum_{k=0}^{p} b_k f_k\big((i-1)/n\big)\right) + \cdots \\
&\quad + \theta^{i-1}\left(X_1 - \sum_{k=0}^{p} b_k f_k(1/n)\right) + \theta^i z_{\text{init}} \\
&= \left(Z_i - Z_{i-1} + \sum_{k=0}^{p} b_{k0} f_k(i/n) - \sum_{k=0}^{p} b_k f_k(i/n)\right) + \cdots \\
&\quad + \theta^{i-1}\left(Z_1 - Z_0 + \sum_{k=0}^{p} b_{k0} f_k(1/n) - \sum_{k=0}^{p} b_k f_k(1/n)\right) + \theta^i z_{\text{init}} \\
&= Z_i - (1-\theta) \sum_{j=0}^{i-1} \theta^{i-1-j} - \theta^i (Z_0 - z_{\text{init}}) \\
&\quad + \sum_{k=0}^{p} (b_{k0} - b_k)\left(\sum_{j=1}^{i} \theta^{i-j} f_k(j/n)\right)
\end{aligned}$$

$$:= Z_i - y_i + \sum_{k=0}^{p}(b_{k0} - b_k)\left(\sum_{j=1}^{i}\theta^{i-j} f_k(j/n)\right)$$

$$:= Z_i - w_i.$$

As in [13], we adopt the parametrization for θ and z_{init} given by

$$\theta = 1 + \frac{\beta}{n} \quad \text{and} \quad z_{\text{init}} = Z_0 + \frac{\alpha\sigma_0}{\sqrt{n}}.$$

Further set

$$b_k = b_{k0} + \frac{\eta_k \sigma_0}{n^{3/2}}. \tag{2.3}$$

Note that (2.3) essentially characterizes the convergence rate of the estimated b_k to its true value b_{k0}. At first glance, this parameterization may look odd since it depends on the known parameter values, which are unavailable. This form of reparameterization is used only for deriving the asymptotic theory of the maximum likelihood estimators and not for estimation purposes. One notes that $\beta = n(\theta - 1)$, $\eta_k = n^{3/2}(b_k - b_{k0})$, so that the asymptotics of the MLE $\hat{\theta}$ and $\hat{b}_k$ of the associated parameters are found by the limiting behavior of $\hat{\beta} = n(\hat{\theta} - 1)$, $\hat{\eta}_k = n^{3/2}(\hat{b}_k - b_{k0})$. Hence, it is not necessary to know the true values in this analysis. The scaling $n^{3/2}$ for the regression coefficients is an artifact of the assumption that the regressors take the form $f_k(t/n)$ that is imposed on the problem. This also results in a clean expression for the limit.

Under the $(\vec{\eta}, \beta, \alpha)$ parameterization, it is easily seen [13], minimizing $l_n(\vec{b}, \theta, z_{\text{init}})$ with respect to $\vec{b}, \theta, z_{\text{init}}$ is equivalent to minimizing the function

$$U_n(\vec{\eta}, \beta, \alpha) \equiv \frac{1}{\sigma_0^2}[l_n(\vec{b}, \theta, z_{\text{init}}) - l_n(\vec{b}_0, 1, Z_0)] \tag{2.4}$$

with respect to $\vec{\eta}$, β and α. Then using the weak convergence results in Davis and Song [13],

$$\begin{aligned}
&U_n(\vec{\eta}, \beta, \alpha)\\
&\quad = \frac{1}{\sigma_0^2}\sum_{i=0}^{n} z_i^2 - Z_i^2 = -2\sum_{i=0}^{n}\frac{w_i Z_i}{\sigma_0^2} + \sum_{i=0}^{n}\frac{w_i^2}{\sigma_0^2}\\
&\quad \xrightarrow{\text{d}} 2\beta\int_0^1\int_0^s e^{\beta(s-t)}\,dW(t)\,dW(s) + 2\alpha\int_0^1 e^{\beta s}\,dW(s)\\
&\qquad - 2\sum_{k=0}^{p}\eta_k\int_0^1\left(\int_0^s e^{\beta(s-t)} f_k(t)\,dt\right)dW(s)\\
&\qquad + \int_0^1\left(\beta\int_0^s e^{\beta(s-t)}\,dW(t) + \alpha e^{\beta s} - \sum_{k=0}^{p}\eta_k\int_0^s e^{\beta(s-t)} f_k(t)\,dt\right)^2 ds\\
&\quad := U(\vec{\eta}, \beta, \alpha),
\end{aligned}$$

where "$\xrightarrow{\text{d}}$" indicates weak convergence on $C(\mathbb{R}^{p+1} \times (-\infty, 0] \times \mathbb{R})$. Throughout this paper, when referring to convergence of stochastic processes on $C(\mathbb{R}^k)$, the notation "$\xrightarrow{\text{d}}$" ("$\xrightarrow{\text{p}}$") means convergence in distribution (probability) on $C(\mathbb{K})$ where $\mathbb{K}$ is any compact set in $\mathbb{R}^k$.

As a special case of a polynomial, set $f_k(t) = t^k$. In this case, the limiting process $U(\vec{\eta}, \beta, \alpha)$ is

$$\begin{aligned}
&U(\vec{\eta}, \beta, \alpha) \\
&\quad = 2\beta \int_0^1 \int_0^s e^{\beta(s-t)}\, dW(t)\, dW(s) \\
&\qquad + 2\alpha \int_0^1 e^{\beta s}\, dW(s) - 2\sum_{k=0}^{p} \eta_k \int_0^1 \left(\int_0^s e^{\beta(s-t)} t^k dt\right) dW(s) \\
&\qquad + \int_0^1 \left(\beta \int_0^s e^{\beta(s-t)}\, dW(t) + \alpha e^{\beta s} - \sum_{k=0}^{p} \eta_k \int_0^s e^{\beta(s-t)} t^k\, dt\right)^2 ds.
\end{aligned}$$

From now on we consider the simple case of just a nonzero mean, that is, $p = 0$ and $f_0(t) = 1$. The formula further simplifies to

$$\begin{aligned}
&U(\eta_0, \beta, \alpha) \\
&\quad = 2\beta \int_0^1 \int_0^s e^{\beta(s-t)}\, dW(t)\, dW(s) \\
&\qquad + 2\alpha \int_0^1 e^{\beta s}\, dW(s) - 2\eta_0 \int_0^1 \frac{1 - e^{\beta s}}{\beta}\, dW(s) \\
&\qquad + \int_0^1 \left(\beta \int_0^s e^{\beta(s-t)}\, dW(t) + \alpha e^{\beta s} - \eta_0 \frac{1 - e^{\beta s}}{\beta}\right)^2 ds.
\end{aligned} \tag{2.5}$$

As shown in [13], one can recover the exact likelihood by integrating out the initial parameter effects. More specifically,

$$\begin{aligned}
f(\mathbf{x}_n, z_{\text{init}}) &= \prod_{t=0}^{n} f(z_t) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n+1} \exp\left\{-\frac{\sum_{t=0}^{n} z_t^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n+1} \exp\left\{-\frac{l_n(b_0, \theta, z_{\text{init}}) - l_n(b_{00}, 1, Z_0) + \sum_{t=0}^{n} Z_t^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n+1} \exp\left\{-\frac{\sum_{t=0}^{n} Z_t^2}{2\sigma^2}\right\} \exp\left\{-\frac{U_n(\eta_0, \beta, \alpha)\sigma_0^2}{2\sigma^2}\right\},
\end{aligned}$$

integrating out the augmented variable z_{init} yields

$$
\begin{aligned}
f(\mathbf{x}_n) &= \int_{-\infty}^{+\infty} f(\mathbf{x}_n, z_{\mathrm{init}})\,dz_{\mathrm{init}} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n+1} \exp\left\{-\frac{\sum_{t=0}^{n} Z_t^2}{2\sigma^2}\right\} \\
&\quad \times \frac{\sigma_0}{\sqrt{n}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{U_n(\eta_0,\beta,\alpha)\sigma_0^2}{2\sigma^2}\right\} d\alpha.
\end{aligned}
\tag{2.6}
$$

A similar argument as in [13] then shows that by profiling out the variance parameter σ^2 the exact profile log-likelihood $L_n(\eta_0, \beta)$ has the following property:

$$
\begin{aligned}
&L_n(\eta_0,\beta) - L_n(\eta_0, 0) \\
&\quad \xrightarrow{\mathrm{d}} L^*(\eta_0,\beta) \\
&\quad = \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\eta_0,\beta,\alpha)}{2}\right\} d\alpha \\
&\qquad - \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\eta_0,0,\alpha)}{2}\right\} d\alpha.
\end{aligned}
\tag{2.7}
$$

The weak convergence results on $C(\mathbb{R}^2)$ in (2.7) can be used to show convergence in distribution of a sequence of local maximizers of the objective functions L_n to the maximizer of the limit process L provided the latter is unique almost surely. This is the content of Remark 1 (see also Lemma 2.2) of Davis, Knight and Liu [12], which for ease of reference, we state a version here.

REMARK 2.1. Suppose $\{L_n(\cdot)\}$ is a sequence of stochastic processes which converge in distribution to $L(\cdot)$ on $C(\mathbb{R}^k)$. If L has a unique maximizer $\tilde{\beta}$ a.s., then there exists a sequence of local maximizers $\{\hat{\beta}_n\}$ of $\{L_n\}$ that converge in distribution to $\tilde{\beta}$. Note that this is consistent with many of the statements made in the classical theory for maximum likelihood (see, e.g., Theorem 7.1.1 of Lehmann [15]) and for inference in nonstandard time series models; see Theorems 8.2.1 and 8.6.1 in Rosenblatt [16], Breidt et al. [5], Andrews et al. [3] and Andrews et al. [2]. In some cases, for example, if the $\{L_n\}$ have concave sample paths, this can be strengthened to convergence of the global maximizers of L_n. See also Davis, Chen and Dunsmuir [9], Davis and Dunsmuir [11], Breidt et a.l [5] for examples of other cases when $\{L_n\}$ are not concave.

Returning to our example, under the case when $\theta_0 = 1$, that is, $\beta = 0$, the limit of the exact likelihood is $L(\eta_0, \beta = 0)$. This corresponds to the situation of inference about the mean term when it is known that the driving noise is an MA(1) process with a unit root. Since the Gaussian likelihood is a quadratic function of regression coefficients, $L(\eta_0, \beta = 0)$ is a quadratic function in η_0. Applying Remark 2.1, we obtain that the MLE $\hat{\eta}_0$ converges in distribution to $\tilde{\eta}_0$, the global maximizer of $L(\eta_0, \beta = 0)$. In particular, $\tilde{\eta}_0$ is the value that makes $\frac{\partial}{\partial \eta_0} L(\eta_0, \beta = 0) = 0$. Since

$$\begin{aligned}\frac{\partial}{\partial \eta_0} L(\eta_0, \beta = 0) \\ &= \frac{\int_{-\infty}^{+\infty} \exp\{-U(\eta_0, \beta = 0, \alpha)/2\}(-(1/2)(\partial U(\eta_0, \beta = 0, \alpha)/\partial \eta_0))\, d\alpha}{\int_{-\infty}^{+\infty} \exp\{-U(\eta_0, \beta = 0, \alpha)/2\}\, d\alpha},\end{aligned}$$

where

$$U(\eta_0, \beta = 0, \alpha) = 2\alpha W(1) - 2\eta_0 \int_0^1 s\, dW(s) + \int_0^1 (\alpha - s\eta_0)^2\, ds$$

and

$$\frac{\partial}{\partial \eta_0} U(\eta_0, \beta = 0, \alpha) = 2\eta_0 \int_0^1 s^2\, ds - 2\int_0^1 s\, dW(s) - 2\alpha \int_0^1 s\, ds.$$

Solving $\frac{\partial}{\partial \eta_0} L(\eta_0, \beta = 0) = 0$, we find that

$$\tilde{\eta}_0 = 12 \int_0^1 s\, dW(s) - 6W(1) \sim \mathrm{N}(0, 12) \tag{2.8}$$

and hence

$$n^{3/2}(\hat{b}_{0,n} - b_0) = \sigma_0 \hat{\eta}_{0,n} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 12\sigma_0^2). \tag{2.9}$$

This counter-intuitive result was also obtained earlier by Chen et al. [8]. It says the MLE of the mean term in the process would behave like a normal distribution asymptotically, but with convergence rate $n^{3/2}$. Notice that, even if one does not know the true value of θ, the MLE of the mean term would still behave very much like (2.9) due to the large pile-up effect in this case. However, the MLE is not asymptotically normal, if both b_0 and θ are estimated.

3. MA(2) with unit roots. The above approach, which also works in the invertible case, does not rely on detailed knowledge of the form of the eigenvectors and eigenvalues of the covariance matrix. Hence it has the potential to work in higher order models where the eigenvector and eigenvalue structure is not known explicitly. We will concentrate on the MA(2) process in this section and further illustrate our methods.

In the following section, we consider the model given in (1.4), where parameters $c_1, c_2 \in \bigtriangledown$, the triangular shaped region depicted in Figure 1. The interior of this region corresponds to the invertibility region of the parameter space. Note that the triangular region is separated into complex roots and real roots of the MA polynomial $1 + c_1 z + c_2 z^2$ by a quadratic curve $c_1^2 - 4c_2 = 0$.

If the parameters are on the boundary of the $\bigtriangledown$ region, it indicates presence of unit roots. Otherwise, the model is said to be invertible; see also Brockwell and

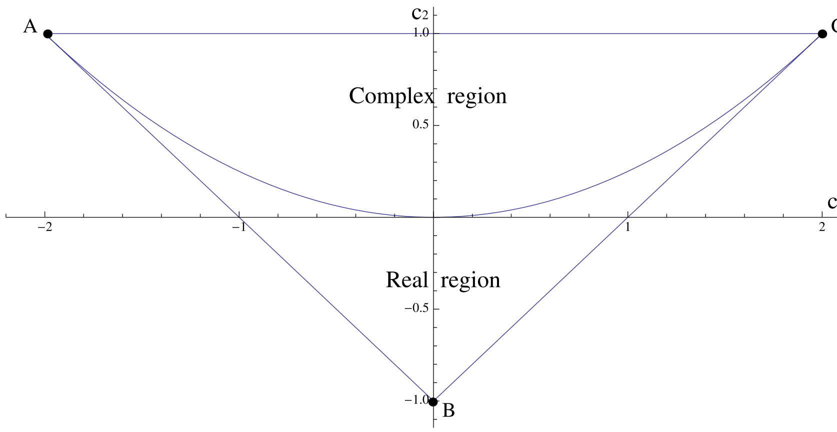

FIG. 1. ∇ *region defined by* $-c_1 - c_2 \le 1, c_1 - c_2 \le 1, |c_2| \le 1$.

Davis [6]. Model (1.4) can also be represented in terms of the roots of the MA polynomial by

$$\begin{aligned} X_t &= (1 + c_1\mathrm{B} + c_2\mathrm{B}^2)Z_t \\ &= (1 - \theta_0\mathrm{B})(1 - \alpha_0\mathrm{B})Z_t, \end{aligned}$$

where $c_1 = -\theta_0 - \alpha_0$ and $c_2 = \theta_0\alpha_0$.

3.1. *Case* 1: $|\alpha_0| < 1$ *and* $\theta_0 = 1$. This case corresponds to the situation of only one unit root in the MA polynomial, that is, the boundary AB in Figure 1. Let $L_n(\theta, \alpha)$ be the profile likelihood of an MA(2) process. Again, we adopt the parametrization

$$\theta = 1 + \frac{\beta}{n}, \qquad \beta \le 0,$$

and

$$\alpha = \alpha_0 + \frac{\gamma}{\sqrt{n}}, \qquad \gamma \in \mathbb{R}.$$

For convenience, define the intermediate process $Y_t = (1 - \alpha_0\mathrm{B})Z_t$ and observe that

$$X_t = (1 - \theta_0\mathrm{B})(1 - \alpha_0\mathrm{B})Z_t = (1 - \theta_0\mathrm{B})Y_t.$$

In the MA(2) case, two augmented initial variables Z_{init} and Y_{init} are needed. These initial variables and the joint likelihood have a simple form, that is,

$$Z_{\text{init}} = Z_{-1} \quad \text{and} \quad Y_{\text{init}} = Z_0 - \alpha_0 Z_{\text{init}}, \tag{3.1}$$

$$
\begin{aligned}
f_{\mathbf{X},Y_{\text{init}},Z_{\text{init}}}(\mathbf{x}_n, y_{\text{init}}, z_{\text{init}}) &= f_{\mathbf{Y},Y_{\text{init}},Z_{\text{init}}}(\mathbf{y}_n, y_{\text{init}}, z_{\text{init}}) \\
&= f_{\mathbf{Z},Z_{\text{init}}}(\mathbf{z}_n, z_{\text{init}}) \\
&= \prod_{j=-1}^{n} f_Z(z_j).
\end{aligned}
$$

As what has been shown in the MA(1) case, the key of our method is to calculate the formula for the residual $r_i := Z_i - z_i$, which can be obtained from

$$
\begin{aligned}
z_i &= y_i + \alpha y_{i-1} + \cdots + \alpha^{i-1} y_1 + \alpha^i y_{\text{init}} + \alpha^{i+1} z_{\text{init}} \\
&= \left(\sum_{j=1}^{i} \theta^{i-j} X_j + \theta^i y_{\text{init}}\right) + \alpha\left(\sum_{j=1}^{i-1} \theta^{i-1-j} X_j + \theta^{i-1} y_{\text{init}}\right) + \cdots \\
&\quad + \alpha^{i-1}(X_1 + \theta y_{\text{init}}) + \alpha^i y_{\text{init}} + \alpha^{i+1} z_{\text{init}} \\
&= \sum_{j=1}^{i} \frac{\theta^{i-j+1} - \alpha^{i-j+1}}{\theta - \alpha} X_j + \frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha} y_{\text{init}} + \alpha^{i+1} z_{\text{init}} \\
&= Z_i - \frac{(\theta_0 - \theta)(\theta - \alpha_0)}{\theta - \alpha} \sum_{j=-1}^{i-1} \theta^{i-j-1} Z_j \qquad (3.2) \\
&\quad - \frac{(\alpha_0 - \alpha)(\theta_0 - \alpha)}{\theta - \alpha} \sum_{j=-1}^{i-1} \alpha^{i-j-1} Z_j + \frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha}(y_{\text{init}} - Y_0) \\
&\quad + \alpha^{i+1}(z_{\text{init}} - Z_{-1}) + (\theta_0 - \theta)\frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha} Z_{-1} \\
&= Z_i - r_i, \qquad (3.3)
\end{aligned}
$$

where the fourth equation (3.2) comes from the fact that $X_j = Z_j - (\theta_0 + \alpha_0) \times Z_{j-1} + \theta_0\alpha_0 Z_{j-2}$ and $Y_0 = Z_0 - \alpha_0 Z_{-1}$. Therefore, the residuals r_i are given by

$$
\begin{aligned}
r_i &= \frac{(\theta_0 - \theta)(\theta - \alpha_0)}{\theta - \alpha} \sum_{j=-1}^{i-1} \theta^{i-j-1} Z_j \\
&\quad + \frac{(\alpha_0 - \alpha)(\theta_0 - \alpha)}{\theta - \alpha} \sum_{j=-1}^{i-1} \alpha^{i-j-1} Z_j - \frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha}(y_{\text{init}} - Y_0) \qquad (3.4) \\
&\quad - \alpha^{i+1}(z_{\text{init}} - Z_{-1}) - (\theta_0 - \theta)\frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha} Z_{-1}.
\end{aligned}
$$

Notice that the residuals r_i no longer have a neat form as in the MA(1) case. This is what makes the MA(2) case more interesting yet more complicated.

In the following calculations, let

$$y_{\text{init}} = Y_0 + \frac{\sigma_0 \eta_1}{\sqrt{n}} \quad \text{and} \quad z_{\text{init}} = Z_{-1} + \frac{\sigma_0 \eta_2}{\sqrt{n}}.$$

With a similar argument as in [13], we opt to minimize the objective function

$$U_n(\beta, \gamma, \eta_1, \eta_2) = -2 \sum_{i=-1}^{n} \frac{r_i Z_i}{\sigma_0^2} + \sum_{i=-1}^{n} \frac{r_i^2}{\sigma_0^2}. \tag{3.5}$$

First note that $r_i = A_i + B_i + C_i + D_i$, where

$$A_i := \frac{(\theta_0 - \theta)(\theta - \alpha_0)}{\theta - \alpha} \sum_{j=-1}^{i-1} \theta^{i-j-1} Z_j - \frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha}(y_{\text{init}} - Y_0),$$

$$B_i := \frac{(\alpha_0 - \alpha)(\theta_0 - \alpha)}{\theta - \alpha} \sum_{j=-1}^{i-1} \alpha^{i-j-1} Z_j,$$

$$C_i := -\alpha^{i+1}(z_{\text{init}} - Z_{-1}),$$

$$D_i := -(\theta_0 - \theta)\frac{\theta^{i+1} - \alpha^{i+1}}{\theta - \alpha} Z_{-1}.$$

To determine the weak limit of $-2\sum_{i=-1}^{n} \frac{r_i Z_i}{\sigma_0^2}$ in (3.5) in the continuous function space, note that

$$\begin{aligned}
-2 \sum_{i=-1}^{n} \frac{A_i Z_i}{\sigma_0^2} &= 2\frac{(\theta - \theta_0)(\theta - \alpha_0)}{\theta - \alpha} \sum_{i=-1}^{n} \sum_{j=-1}^{i-1} \theta^{i-j-1} \frac{Z_j}{\sigma_0} \frac{Z_i}{\sigma_0} \\
&\quad + \frac{2\eta_1}{\sqrt{n}(\theta - \alpha)} \sum_{i=-1}^{n} \theta^{i+1} \frac{Z_i}{\sigma_0} - \frac{2\eta_1}{\sqrt{n}(\theta - \alpha)} \sum_{i=-1}^{n} \alpha^{i+1} \frac{Z_i}{\sigma_0} \\
&= 2\frac{\beta(1 - \alpha_0 + \beta/n)}{1 - \alpha_0 + \beta/n - \gamma/\sqrt{n}} \sum_{i=-1}^{n} \sum_{j=-1}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j-1} \frac{Z_j}{\sigma_0} \frac{Z_i}{\sqrt{n}\sigma_0} \\
&\quad + \frac{2\eta_1}{(1 - \alpha_0 + \beta/n - \gamma/\sqrt{n})} \sum_{i=-1}^{n} \left(1 + \frac{\beta}{n}\right)^{i+1} \frac{Z_i}{\sqrt{n}\sigma_0} \\
&\quad - \frac{2\eta_1}{\sqrt{n}(1 - \alpha_0 + \beta/n - \gamma/\sqrt{n})} \sum_{i=-1}^{n} \left(\alpha_0 + \frac{\gamma}{\sqrt{n}}\right)^{i+1} \frac{Z_i}{\sqrt{n}\sigma_0} \\
&\xrightarrow{\mathrm{d}} 2\beta \int_0^1 \int_0^s e^{\beta(s-t)}\, dW(t)\, dW(s) + \frac{2\eta_1}{1 - \alpha_0} \int_0^1 e^{\beta s}\, dW(s),
\end{aligned} \tag{3.6}$$

where the last term disappears in the limit due to the fact that $|\alpha_0| < 1$. Similarly, we have

$$
\begin{aligned}
-2\sum_{i=-1}^{n}\frac{B_i Z_i}{\sigma_0^2} &= 2\frac{(\alpha-\alpha_0)(1-\alpha)}{\theta-\alpha}\sum_{i=-1}^{n}\sum_{j=-1}^{i-1}\alpha^{i-j-1}\frac{Z_j}{\sigma_0}\frac{Z_i}{\sigma_0} \\
&= 2\frac{\gamma(1-\alpha_0-\gamma/\sqrt{n})}{1-\alpha_0+\beta/n-\gamma/\sqrt{n}} \\
&\quad\times\sum_{i=-1}^{n}\sum_{j=-1}^{i-1}\left(\alpha_0+\frac{\gamma}{\sqrt{n}}\right)^{i-j-1}\frac{Z_j}{\sigma_0}\frac{Z_i}{\sqrt{n}\sigma_0} \\
&= 2\gamma\sum_{i=-1}^{n}\sum_{j=-1}^{i-1}\alpha_0^{i-j-1}\frac{Z_j}{\sigma_0}\frac{Z_i}{\sqrt{n}\sigma_0}+o_p(1) \qquad (3.7) \\
&\xrightarrow{\mathrm{d}} 2\gamma N, \qquad (3.8)
\end{aligned}
$$

where $N \sim \mathrm{N}(0, \frac{1}{1-\alpha_0^2})$. The third equality holds because $|\alpha_0|$ is strictly smaller than 1, and $o_p(1)$ is uniform in γ on any compact set of $\mathbb{R}$. The weak convergence from (3.7) to (3.8) follows from martingale central limit theorem; see Hall and Heyde [14]. It can also be shown that N and the $W(t)$ process from (3.6) are independent; see Theorem 2.2 in Chan and Wei [7].

Following similar arguments, it is easy to show that

$$
-2\sum_{i=-1}^{n}\frac{C_i Z_i}{\sigma_0^2}\xrightarrow{\mathrm{p}} 0 \quad \text{and} \quad -2\sum_{i=-1}^{n}\frac{D_i Z_i}{\sigma_0^2}\xrightarrow{\mathrm{p}} 0.
$$

For the second term in (3.5), writing

$$
\begin{aligned}
\sum_{i=-1}^{n}\frac{r_i^2}{\sigma_0^2} &= \sum_{i=-1}^{n}\frac{A_i^2+B_i^2+C_i^2+D_i^2}{\sigma_0^2} \\
&\quad+\sum_{i=-1}^{n}\frac{2A_iB_i+2A_iC_i+2A_iD_i+2B_iC_i+2B_iD_i+2C_iD_i}{\sigma_0^2},
\end{aligned}
$$

and using Corollary 2.10 in [13], we have

$$
\sum_{i=-1}^{n}\frac{A_i^2}{\sigma_0^2}\xrightarrow{\mathrm{d}}\int_0^1\left(\beta\int_0^s e^{\beta(s-t)}\,dW(t)+\frac{\eta_1}{1-\alpha_0}e^{\beta s}\right)^2 ds, \qquad (3.9)
$$

$$
\sum_{i=-1}^{n}\frac{B_i^2}{\sigma_0^2}\xrightarrow{\mathrm{p}}\gamma^2\operatorname{var}(N). \qquad (3.10)
$$

Moreover, it is relatively easy to show that

$$\sum_{i=-1}^{n} \frac{C_i^2}{\sigma_0^2} \xrightarrow{\text{p}} 0 \quad \text{and} \quad \sum_{i=-1}^{n} \frac{D_i^2}{\sigma_0^2} \xrightarrow{\text{p}} 0. \tag{3.11}$$

Next we show that all the cross product terms also vanish in the limit, namely,

$$\sum_{i=-1}^{n} \frac{2A_iB_i + 2A_iC_i + 2A_iD_i + 2B_iC_i + 2B_iD_i + 2C_iD_i}{\sigma_0^2} \xrightarrow{\text{p}} 0. \tag{3.12}$$

Here we only give the details for showing $\sum_{i=-1}^{n} \frac{A_iB_i}{\sigma_0^2} \xrightarrow{\text{p}} 0$; the other cases can be proved in an analogous manner. Notice that for any fixed $M > 0$ and any $\beta \in [-M, 0]$,

$$\begin{aligned}
\sum_{i=-1}^{n} \frac{A_iB_i}{\sigma_0^2} &= \frac{\beta/n(1-\alpha_0+\beta/n)\gamma/\sqrt{n}(1-\alpha_0-\gamma/\sqrt{n})}{(1-\alpha_0+\beta/n-\gamma/\sqrt{n})^2} \\
&\quad \times \sum_{i=-1}^{n}\left(\sum_{j=-1}^{i-1}\alpha^{i-j-1}\frac{Z_j}{\sigma_0}\right)\left(\sum_{j=-1}^{i-1}\theta^{i-j-1}\frac{Z_j}{\sigma_0}\right) \\
&\quad + \frac{(\gamma/\sqrt{n})(1-\alpha_0-\gamma/\sqrt{n})(\eta_1/\sqrt{n})}{(1-\alpha_0+\beta/n-\gamma/\sqrt{n})^2} \\
&\quad \times \sum_{i=-1}^{n}\left[(\theta^{i+1}-\alpha^{i+1})\sum_{j=-1}^{i-1}\alpha^{i-j-1}\frac{Z_j}{\sigma_0}\right] \\
&= \frac{\beta\gamma}{n}\sum_{i=-1}^{n}\left(\sum_{j=-1}^{i-1}\alpha_0^{i-j-1}\frac{Z_j}{\sigma_0}\right)\left(\sum_{j=-1}^{i-1}\left(1+\frac{\beta}{n}\right)^{i-j-1}\frac{Z_j}{\sqrt{n}\sigma_0}\right) \\
&\quad + \frac{\gamma\eta_1}{n}\sum_{i=1}^{n}\left[\left(1+\frac{\beta}{n}\right)^{i+1}\sum_{j=-1}^{i-1}\alpha_0^{i-j-1}\frac{Z_j}{\sigma_0}\right] \\
&\quad \times \frac{\gamma\eta_1}{n}\sum_{i=-1}^{n}\sum_{j=-1}^{i-1}\alpha_0^{2i-j}\frac{Z_j}{\sigma_0} + o_p(1),
\end{aligned} \tag{3.13}$$

where $o_p(1)$ is uniform in β and γ on any compact set in $\mathbb{R}^- \times \mathbb{R}$. Setting $R_i = \sum_{j=-1}^{i} \alpha_0^{i-j} Z_j/\sigma_0$, it follows that R_i is a stationary AR(1) process satisfying

$$R_i = \alpha_0 R_{i-1} + Z_i/\sigma_0.$$

Since $|\alpha_0| < 1$, we can apply Theorem 3.7. in Tanaka [21] to obtain

$$S_n(t) := \frac{1}{\sqrt{n}} \sum_{i=0}^{[nt]} R_i \xrightarrow{\text{d}} \tilde{\alpha} S(t),$$

where $\tilde{\alpha} = \sum_{l=0}^{\infty} \alpha_0^l = \frac{1}{1-\alpha_0}$ and $S(t)$ is a standard Brownian motion. Also, since R_i is adapted to the σ-fields $\mathcal{F}_i$ generated by $Z_0, \ldots, Z_i$. By Theorem 2.1 in [13], we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=-1}^{n} \left(1 + \frac{\beta}{n}\right)^{i+1} R_{i-1} \xrightarrow{\mathrm{d}} \tilde{\alpha} \int_0^1 e^{\beta s} \, dS(s) \qquad \text{on } C[-M, 0].$$

Therefore,

$$\frac{\gamma \eta_1}{n} \sum_{i=-1}^{n} \left[\left(1 + \frac{\beta}{n}\right)^{i+1} \sum_{j=-1}^{i-1} \alpha_0^{i-j-1} \frac{Z_j}{\sigma_0}\right] \xrightarrow{\mathrm{p}} 0 \qquad \text{on } C[-M, 0]. \tag{3.14}$$

It is also easy to see that

$$\frac{\gamma \eta_1}{n} \sum_{i=-1}^{n} \sum_{j=-1}^{i-1} \alpha_0^{2i-j} \frac{Z_j}{\sigma_0} = \frac{\gamma \eta_1}{n} \sum_{i=1}^{n} \alpha_0^{i+1} R_{i-1} \xrightarrow{\mathrm{p}} 0. \tag{3.15}$$

Since

$$\sum_{i=-1}^{n} \left(\sum_{j=-1}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j-1} \frac{Z_j}{\sqrt{n}\sigma_0} \right) \frac{R_{i-1}}{\sqrt{n}} \tag{3.16}$$

is in the form of the double sum in Theorem 2.8 in [13], except that $\{R_i\}$ is no longer a martingale difference sequence. However, we can still follow the proof of Theorem 2.8 in [13] and show that (3.16) has a nondegenerate weak limit in $C[-M, 0]$. It follows that

$$\begin{aligned} &\frac{\beta\gamma}{n} \sum_{i=-1}^{n} \left(\sum_{j=-1}^{i-1} \alpha_0^{i-j-1} \frac{Z_j}{\sigma_0} \right) \left(\sum_{j=-1}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j-1} \frac{Z_j}{\sqrt{n}\sigma_0} \right) \\ &\quad = \frac{\beta\gamma}{\sqrt{n}} \sum_{i=-1}^{n} \left(\sum_{j=-1}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j-1} \frac{Z_j}{\sqrt{n}\sigma_0} \right) \frac{R_{i-1}}{\sqrt{n}} \xrightarrow{\mathrm{p}} 0. \end{aligned} \tag{3.17}$$

Thus, combining (3.14), (3.15) and (3.17), we conclude that the terms in (3.13) go to 0 in probability on $C[-M, 0]$. The convergence in probability of the other terms in (3.12) can also be proved in a similar way. To sum up, we have shown the key stochastic process convergence result, that is,

$$\begin{aligned} &U_n(\beta, \gamma, \eta_1, \eta_2) \\ &\quad \xrightarrow{\mathrm{d}} U(\beta, \gamma, \eta_1) \\ &\quad = 2\beta \int_0^1 \int_0^s e^{\beta(s-t)} \, dW(t) \, dW(s) + 2\gamma N + \frac{2\eta_1}{1-\alpha_0} \int_0^1 e^{\beta s} \, dW(s) \\ &\qquad + \int_0^1 \left(\beta \int_0^s e^{\beta(s-t)} \, dW(t) + \frac{\eta_1}{1-\alpha_0} e^{\beta s} \right)^2 ds + \gamma^2 \operatorname{var}(N). \end{aligned} \tag{3.18}$$

Using (3.18), one can easily derive the asymptotics for the exact profile log-likelihood denoted by $L_n(\beta, \gamma)$. In particular,

$$
\begin{aligned}
&L_n(\beta,\gamma) - L_n(0,0) \\
&\quad \xrightarrow{\mathrm{d}} \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\beta,\gamma,\eta_1)}{2}\right\} d\eta_1 \qquad (3.19)\\
&\qquad - \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(0,0,\eta_1)}{2}\right\} d\eta_1 \\
&\quad := L^*(\beta,\gamma) \\
&\quad = -\gamma N - \frac{\gamma^2}{2}\operatorname{var}(N) + \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\beta,\eta^*)}{2}\right\} d\eta^* \qquad (3.20)\\
&\qquad - \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(0,\eta^*)}{2}\right\} d\eta^*,
\end{aligned}
$$

where $\eta^* = \frac{\eta_1}{1-\alpha_0}$ and $U(\beta, \eta^*)$ is given by

$$
\begin{aligned}
U(\beta,\eta^*) &= 2\int_0^1 \left[\beta \int_0^s e^{\beta(s-t)}\, dW(t) + \eta^* e^{\beta s}\right] dW(s) \\
&\quad + \int_0^1 \left[\beta \int_0^s e^{\beta(s-t)}\, dW(t) + \eta^* e^{\beta s}\right]^2 ds,
\end{aligned} \tag{3.21}
$$

which is the limiting process of the joint likelihood obtained in the unit root MA(1) case, see also Davis and Song [13]. We state the key result of this paper in the following theorem.

THEOREM 3.1. *Consider the model given in* (1.4) *with two roots* θ *and* α *which are parameterized by*

$$\theta = 1 + \frac{\beta}{n} \quad \textit{and} \quad \alpha = \alpha_0 + \frac{\gamma}{\sqrt{n}}.$$

Denote the profile log-likelihood based on a Gaussian likelihood as $L_n(\beta, \gamma)$. *Then* $L_n(\beta, \gamma)$ *satisfies*

$$L_n(\beta,\gamma) - L_n(0,0) \xrightarrow{\mathrm{d}} L^*(\beta,\gamma) \qquad \text{on } C([-\infty,0]\times\mathbb{R}),$$

where

$$
\begin{aligned}
L^*(\beta,\gamma) &= -\gamma N - \frac{\gamma^2}{2}\operatorname{var}(N) + U^*(\beta) \\
&\overset{\mathrm{d}}{=} -\gamma N - \frac{\gamma^2}{2}\operatorname{var}(N) + \frac{1}{2} Z_0(\beta).
\end{aligned} \tag{3.22}
$$

The processes $U^*(\beta)$ and $Z_0(\beta)$ are defined by

$$U^*(\beta) = \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\beta,\alpha)}{2}\right\} d\alpha - \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(0,\alpha)}{2}\right\} d\alpha \tag{3.23}$$

and

$$Z_0(\beta) = \sum_{k=1}^{\infty} \frac{\beta^2\pi^2k^2X_k^2}{(\pi^2k^2+\beta^2)\pi^2k^2} + \sum_{k=1}^{\infty} \log\left(\frac{\pi^2k^2}{\pi^2k^2+\beta^2}\right). \tag{3.24}$$

Furthermore, there exists a sequence of local maxima $\hat{\beta}_n, \hat{\gamma}_n$ of $L_n(\beta,\gamma)$ converging in distribution to $\tilde{\beta}_{\mathrm{MLE}}, \tilde{\gamma}_{\mathrm{MLE}}$, the global maximum of the limiting process $U^*(\beta,\gamma)$. If model (1.4) has, at most, one unit root, then for the estimators $\hat{c}_1$ and $\hat{c}_2$, we have

$$\sqrt{n}\begin{pmatrix} \hat{c}_1 - c_1 \\ \hat{c}_2 - c_2 \end{pmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \begin{bmatrix} 1-c_2^2 & c_1(1-c_2) \\ c_1(1-c_2) & 1-c_2^2 \end{bmatrix}\right). \tag{3.25}$$

REMARK 3.2. The equivalence in distribution of the processes $U^*(\beta)$ and $\frac{1}{2}Z_0(\beta)$ is given in Theorem 4.3 in Davis and Song [13]. As mentioned in Davis and Dunsmiur [10], convergence on C(-∞,0] does not necessarily imply convergence of the corresponding global maximizers. Additional arguments were required 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 $\hat{\beta}_{\mathrm{MLE},n}$ and $\hat{\gamma}_{\mathrm{MLE},n}$ and simulation results, some of which are contained in Sections 4 and 5, bear this out.

REMARK 3.3. To establish the convergence in (3.25), if there is exactly one unit root, then

$$\begin{aligned}
\sqrt{n}(\hat{c}_1 - c_1) &= -\frac{\hat{\beta}_{\mathrm{MLE}}}{\sqrt{n}} - \hat{\gamma}_{\mathrm{MLE}} \xrightarrow{\mathrm{d}} -\tilde{\gamma}_{\mathrm{MLE}} = \frac{N}{\mathrm{var}(N)} \\
&\stackrel{\mathrm{d}}{=} \mathrm{N}(0, 1-\alpha_0^2) = \mathrm{N}(0, 1-c_2^2), \\
\sqrt{n}(\hat{c}_2 - c_2) &= \hat{\gamma}_{\mathrm{MLE}} + \frac{\alpha_0\hat{\beta}_{\mathrm{MLE}}}{\sqrt{n}} + \frac{\hat{\gamma}_{\mathrm{MLE}}\hat{\beta}_{\mathrm{MLE}}}{n} \xrightarrow{\mathrm{d}} \tilde{\gamma}_{\mathrm{MLE}} \\
&= -\frac{N}{\mathrm{var}(N)} \stackrel{\mathrm{d}}{=} \mathrm{N}(0, 1-\alpha_0^2) = \mathrm{N}(0, 1-c_2^2).
\end{aligned}$$

Here, we use the fact that $\tilde{\beta}_{\mathrm{MLE}} < \infty$ a.s. as stated in (Theorem 4.3 in [13]). One can also calculate the limiting asymptotic covariance of $\hat{c}_1$ and $\hat{c}_2$ as

$$\begin{aligned}
-\mathrm{var}(\tilde{\gamma}_{\mathrm{MLE}}) &= -(1-\alpha_0^2) = -(1+\alpha_0)(1-\alpha_0) \\
&= c_1(1-c_2).
\end{aligned}$$

REMARK 3.4. The above theorem says that when $|\alpha_0| < 1$ and $\theta_0 = 1$, we have a similar asymptotic result for c_1 and c_2 as in the invertible case. If we only consider the original parameters c_1 and c_2, the effect of the unit root disappears in the limit. But $\sqrt{n}(\hat{c}_1 - c_1)$ and $\sqrt{n}(\hat{c}_2 - c_2)$ are perfectly dependent in the limit, since $c_1(1 - c_2) = 1 - c_2^2$.

REMARK 3.5. The estimated roots $\hat{\theta}$ and $\hat{\alpha}$ calculated from $\hat{c}_1$ and $\hat{c}_2$ are asymptotically independent. Interestingly, $\tilde{\beta}_{\mathrm{MLE}}$ corresponding to the unit root in MA(2) has exactly the same distribution as the $\tilde{\beta}_{\mathrm{MLE}}$ in the MA(1) case. So the pile-up and other properties of $\tilde{\beta}_{\mathrm{MLE}}$ follow exactly from those in the MA(1) case. It may seem surprising that the unit root in the MA(2) model (when there is only one unit root) behaves asymptotically just like the unit root in MA(1) case. To see this, consider the situation where we are given the parameter α and $\alpha = \alpha_0$. In this case, $\gamma = 0$ and

$$L_n(\beta, 0) - L_n(0, 0) \xrightarrow{\mathrm{d}} \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(\beta, \eta^*)}{2}\right\} d\eta^* - \log \int_{-\infty}^{+\infty} \exp\left\{-\frac{U(0, \eta^*)}{2}\right\} d\eta^*,$$

which is the limiting process of the exact profile log-likelihood in the MA(1) case. On the other hand when α is given, θ becomes the only parameter that needs to be estimated

$$X_t = (1 - \alpha_0 \mathrm{B})(1 - \theta \mathrm{B}) Z_t. \tag{3.26}$$

Because of the invertibility of the operator $1 - \alpha_0\mathrm{B}$, we can get an intermediate process Y_t by inverting the operator. Namely,

$$Y_t := \frac{1}{(1 - \alpha_0 \mathrm{B})} X_t = \sum_{k=0}^{\infty} \alpha_0^k X_{t-k} = (1 - \theta \mathrm{B}) Z_t. \tag{3.27}$$

Since we are dealing with asymptotics, inverting the operator $1 - \alpha_0\mathrm{B}$ is feasible. Therefore, the transformed process Y_t is indeed an MA(1) process with the true parameter $\theta_0 = 1$. Then it follows naturally that the properties of the estimator of θ in this situation should be equivalent to those of θ in a unit root MA(1) process.

3.2. *MA(2) with two unit roots.* In moving from the unit root problem for the MA(1) model to the MA(2) model, several new and challenging problems arise. In this subsection, we discuss some issues when there are two unit roots in the MA polynomial.

3.2.1. *Case* 2: $c_2 = 1$ *and* $c_1 \neq \pm 2$. This corresponds to the case that the true parameters are on the boundary $c_2 = 1$, that is, the boundary AC in Figure 1, which means the two roots live on the unit circle and are not real valued. Denote the two generic complex valued roots of the MA polynomial by $\phi = re^{\vec{i}\theta}$ and $\bar{\phi} = re^{-\vec{i}\theta}$. To avoid confusion in notation, we use $\vec{i}$ to represent $\sqrt{-1}$. A rather different representation of the residuals r_i is used in this case, that is,

$$
\begin{aligned}
r_i = {} & \frac{(\phi_0 - \phi)(\phi - \bar{\phi}_0)}{\phi - \bar{\phi}} \sum_{j=-1}^{i-1} \phi^{i-j-1} Z_j \\
& + \frac{(\bar{\phi}_0 - \bar{\phi})(\phi_0 - \bar{\phi})}{\phi - \bar{\phi}} \sum_{j=-1}^{i-1} \bar{\phi}^{i-j-1} Z_j \\
& - \frac{\phi^{i+1} - \bar{\phi}^{i+1}}{\phi - \bar{\phi}} (z_{\text{init},0} - Z_0) \\
& + \frac{\phi\bar{\phi}(\phi^i - \bar{\phi}^i)}{\phi - \bar{\phi}} (z_{\text{init},-1} - Z_{-1}) \\
& + \frac{(\phi + \bar{\phi} - \phi_0 - \bar{\phi}_0)(\phi^{i+1} - \bar{\phi}^{i+1})}{\phi - \bar{\phi}} Z_{-1}.
\end{aligned}
\tag{3.28}
$$

We also adopt the parameterization for r, θ and two initial variables given by

$$
r = 1 + \frac{\beta}{n} \quad \text{and} \quad \theta = \theta_0 + \frac{\gamma}{n},
$$

$$
z_{\text{init},0} = Z_0 + \frac{\sigma_0 \eta_1}{\sqrt{n}} \quad \text{and} \quad z_{\text{init},-1} = Z_{-1} + \frac{\sigma_0 \eta_2}{\sqrt{n}}.
$$

Again, we study the limiting process of $-2\sum_{i=-1}^{n} \frac{r_i Z_i}{\sigma_0^2} + \sum_{i=-1}^{n} \frac{r_i^2}{\sigma_0^2}$. Here we only present the first term of $\sum_{i=-1}^{n} \frac{r_i Z_i}{\sigma_0^2}$ for illustration; the limit of the other terms can be derived in a similar fashion. By Theorem 2.8 in [13], we obtain

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=0}^{n} \sum_{j=-1}^{i-1} \phi^{i-j} \frac{Z_j Z_i}{\sigma_0^2} \\
& \quad = \sum_{i=0}^{n} \left(\sum_{j=-1}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j} \exp\left\{ \vec{i}\gamma \frac{i-j}{n} \right\} \frac{e^{-\vec{i}\theta_0 j} Z_j}{\sqrt{n}\sigma_0} \right) \frac{e^{\vec{i}\theta_0 i} Z_i}{\sqrt{n}\sigma_0} \\
& \quad \xrightarrow{\text{d}} \int_0^1 \int_0^s e^{\beta(s-t) + \vec{i}\gamma(s-t)} \, d\overline{\mathbb{W}}(t) \, d\mathbb{W}(s),
\end{aligned}
$$

where $\mathbb{W}(t)$ is a two-dimensional Brownian motion, $\mathbb{W}(t) = W_1(t) + \vec{i} W_2(t)$, $\overline{\mathbb{W}}(t) = W_1(t) - \vec{i} W_2(t)$ and $W_1(t)$ and $W_2(t)$ are the corresponding weak limits of the sum

$$W_{1,n}(t) = \sum_{k=0}^{[nt]} \cos(k\theta_0) \frac{Z_k}{\sqrt{n}\sigma_0} \quad \text{and} \quad W_{2,n}(t) = \sum_{k=0}^{[nt]} \sin(k\theta_0) \frac{Z_k}{\sqrt{n}\sigma_0}.$$

The weak convergence of $W_{1,n}(t)$ and $W_{2,n}(t)$ to two independent Brownian motions is guaranteed by Theorem 2.2 in Chan and Wei [7].

By Theorem 2.1 in [13] we have

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n} \phi^i \frac{Z_i}{\sigma_0} \xrightarrow{\mathrm{d}} \int_0^1 e^{\beta s + \vec{i}\gamma s} \, d\mathbb{W}(s).$$

Therefore, (3.28) leads to

$$\begin{aligned}
-2 \sum_{i=-1}^{n} \frac{r_i Z_i}{\sigma_0^2} \xrightarrow{\mathrm{d}} {} & 4\Re\biggl\{(\gamma \cos\theta_0 - \gamma \sin\theta_0 + \beta\cos\theta_0 + \vec{i}\beta\sin\theta_0) e^{-\vec{i}\theta_0} \\
& \qquad \times \int_0^1 \int_0^s e^{\beta(s-t) + \vec{i}\gamma(s-t)} \, d\overline{\mathbb{W}}(t) \, d\mathbb{W}(s)\biggr\} \\
& + 4\eta_1 \Re\biggl\{\frac{e^{\vec{i}\theta_0}}{2\vec{i}\sin\theta_0} \int_0^1 e^{\beta s + \vec{i}\gamma s} \, d\mathbb{W}(s)\biggr\} \\
& - 4\eta_2 \Re\biggl\{\frac{1}{2\vec{i}\sin\theta_0} \int_0^1 e^{\beta s + \vec{i}\gamma s} \, d\mathbb{W}(s)\biggr\},
\end{aligned} \tag{3.29}$$

where $\Re\{\cdot\}$ means the real part of a complex function. The weak limit of $\sum_{i=-1}^{n} r_i^2/\sigma_0^2$ can also be computed in an analogous manner using Corollary 2.10 in [13]. However, the weak limit of $\sum_{i=-1}^{n} r_i^2/\sigma_0^2$ has an even more complicated form than (3.29).

By integrating out the auxiliary variables, the exact likelihood can be recovered as well. However, the form of the joint likelihood function is much more complicated than the one computed in the one unit root case. The asymptotic properties and pile-up probabilities in this case remain unknown.

3.2.2. *Case* 3: $c_2 = 1$ *and* $c_1 = -2$. This corresponds to the vertex A in the $\bigtriangledown$-region in Figure 1. It is convenient to first consider a special case of local asymptotics when the approach to the corner is through the boundary $-c_1 - c_2 = 1$. With this constraint, the dimension of the parameters has been reduced from two to one. We parameterize the MA(2) in this case by

$$X_t = Z_t - (\theta + 1) Z_{t-1} + \theta Z_{t-2} \tag{3.30}$$

and define a Z_{init} and a Y_{init} as in (3.1), but with different normalization, that is,

$$\theta = 1 + \frac{\beta}{n}, \qquad Y_{\text{init}} = Y_0 + \frac{\sigma_0 \eta_1}{n^{3/2}} \quad \text{and} \quad Z_{\text{init}} = Z_{-1} + \frac{\sigma_0 \eta_2}{\sqrt{n}}. \tag{3.31}$$

Then, with the help of the theorems in Davis and Song [13], it follows that

$$
\begin{aligned}
U_n(\beta,\eta_1,\eta_2) &= -2\sum_{i=-1}^{n}\frac{r_i Z_i}{\sigma_0^2}+\sum_{i=-1}^{n}\frac{r_i^2}{\sigma_0^2}\\
&\xrightarrow{\mathrm{d}} 2\beta\int_0^1\int_0^s e^{\beta(s-t)}\,dW(t)\,dW(s)\\
&\quad + 2\eta_2\int_0^1 e^{\beta s}\,dW(s)-\frac{2\eta_1}{\beta}\int_0^1(1-e^{\beta s})\,dW(s)\\
&\quad + \int_0^1\left(\beta\int_0^s e^{\beta(s-t)}\,dW(t)+\eta_2 e^{\beta s}-2\eta_1\frac{1-e^{\beta s}}{\beta}\right)^2 ds.
\end{aligned}
\tag{3.32}
$$

There is a connection between this limiting process and the one in (2.5) derived for the limiting process for an MA(1) model with a nonzero mean. Notice that in (2.5), $U(\eta_0,\beta,\alpha)$ is exactly the process we just derived with η_1 and η_2 replaced by α and η_0. This leads us to an interesting connection of the mean term in the lower order MA model and the initial value in the higher order MA model, which we will discuss further in the Section 6.

Alternatively, if we do not impose the constraint $-c_1 - c_2 = 1$, there are two possible ways to parameterize the roots. First, the vertex can be approached through the real region, where $c_1 = -\theta - \alpha$, $c_2 = \theta\alpha$ and the roots are parameterized further as

$$\theta = 1+\frac{\beta}{n} \quad \text{and} \quad \alpha = 1+\frac{\gamma}{n},$$

which makes

$$c_1 = -1-\left(1+\frac{\beta+\gamma}{n}\right) \quad \text{and} \quad c_2 = 1+\frac{\beta+\gamma}{n}+o\left(\frac{1}{n}\right).$$

The second parameterization is through the complex region, in which the roots are $re^{\vec{i}\theta}$ and $re^{-\vec{i}\theta}$ with $c_1 = -2r\cos(\theta)$, $c_2 = r^2$. The radius and the angular parts are further parameterized as

$$r = 1+\frac{\beta}{n} \quad \text{and} \quad \theta = \frac{\gamma}{n},$$

which implies

$$c_1 = -1-\left(1+\frac{2\beta}{n}\right)+o\left(\frac{1}{n}\right) \quad \text{and} \quad c_2 = 1+\frac{2\beta}{n}+o\left(\frac{1}{n}\right).$$

Therefore, in either case, if we ignore the higher order terms, c_1 and c_2 can be approximated as

$$c_1 = -1-\left(1+\frac{\zeta}{n}\right) \quad \text{and} \quad c_2 = 1+\frac{\zeta}{n}.$$

This parameterization, however, is exactly the one we have seen in the conditional case, which suggests that one of the unit roots has pile-up with probability one asymptotically while the other unit root behaves like the unit root in the conditional case; see (3.30) and (3.32). This claim is also supported by the simulation results; see Table 4 in Section 5.

4. Testing for a unit root in an MA(2) model. A direct application of the results in the previous section is testing for the presence of a unit root in the MA(2) model. For the testing problem, we extend the idea of a generalized likelihood ratio test proposed in Davis, Chen and Dunsmuir [9] to the MA(2) case. Tests based on $\hat{\beta}_{\text{MLE}}$ are also considered in this section. We will compare these tests with the score-type test of Tanaka [20].

To specify our hypothesis testing problem in the MA(2) case, the null hypothesis is H_0: there is exactly one unit root in the MA polynomial, and the alternative is H_A: there are no unit roots. The asymptotic theory of the previous section allows us to approximate the nominal power against local alternatives. To set up the problem, for the model

$$X_t = Z_t - \left(\alpha + 1 + \frac{\beta}{n}\right) Z_{t-1} + \alpha\left(1 + \frac{\beta}{n}\right) Z_{t-2}$$

with $|\alpha| < 1$. We want to test $H_0 : \beta = 0$ versus $H_A : \beta < 0$.

To describe the test based on the generalized likelihood ratio, let $\mathrm{GLR}_n = 2(L_n(\hat{\beta}_{\text{MLE}}, \hat{\gamma}_{\text{MLE}}) - L_n(0, \hat{\gamma}_{\text{MLE},0}))$, where $\hat{\gamma}_{\text{MLE},0}$ is the MLE of γ when $\beta = 0$. An application of Theorem 3.1 gives $\mathrm{GLR}_n \xrightarrow{\mathrm{d}} L^*(\tilde{\beta}_{\text{MLE}}, \tilde{\gamma}_{\text{MLE}}) - L^*(0, \tilde{\gamma}_{\text{MLE}}) = U^*(\tilde{\beta}_{\text{MLE}})$, where $L^*(\beta, \gamma)$ and $U^*(\beta)$ are given in (3.22) and (3.23) and $\tilde{\gamma}_{\text{MLE}} = -N/\operatorname{var}(N)$. Notice that the limit distribution of GLR_n only depends on $\tilde{\beta}_{\text{MLE}}$, and γ serves as a nuisance parameter, which does not play a role in the limit. Define the $(1-\alpha)$th asymptotic quantile $b_{\text{GLR}}(\alpha)$ and $b_{\text{MLE}}(\alpha)$ as

$$\mathbf{P}\big(U^*(\tilde{\beta}_{\text{MLE}}) > b_{\text{GLR}}(\alpha)\big) = \alpha \quad \text{and} \quad \mathbf{P}\big(\tilde{\beta}_{\text{MLE}} > b_{\text{MLE}}(\alpha)\big) = \alpha.$$

Since the limiting random variables $U^*(\tilde{\beta}_{\text{MLE}})$ and $\tilde{\beta}_{\text{MLE}}$ are the same as in the MA(1) unit root case, the critical values of $b_{\text{GLR}}(\alpha)$ and $b_{\text{MLE}}(\alpha)$ are the same as those provided in Table 3.2 of Davis, Chen and Dunsmuir [9].

There has been limited research on the testing for a unit root in the MA(2) case. One approach, proposed by Tanaka, was based on a score type of statistic, which is locally best invariant and unbiased (LBIU). However, implementation of this test requires choosing a sequence $l_n \to \infty$ at a suitable rate. One choice is $l_n = o(n^{1/4})$, yet this may not always work well, especially if $\alpha > 0$; see also [20]. Next we compare the power curves of the three tests for sample size $n = 50$.

Figure 2 below shows the power curves based on MLE, GLR and LBIU tests, when the invertible root α in the MA(2) model is -0.3 and -0.5, respectively. Since the score-type test of Tanaka is demonstrated to be locally best invariant unbiased, it has a very small edge on the GLR test up to the local alternative 4 or so. Thereafter, the GLR test increasingly outperforms the LBIU test by a wide

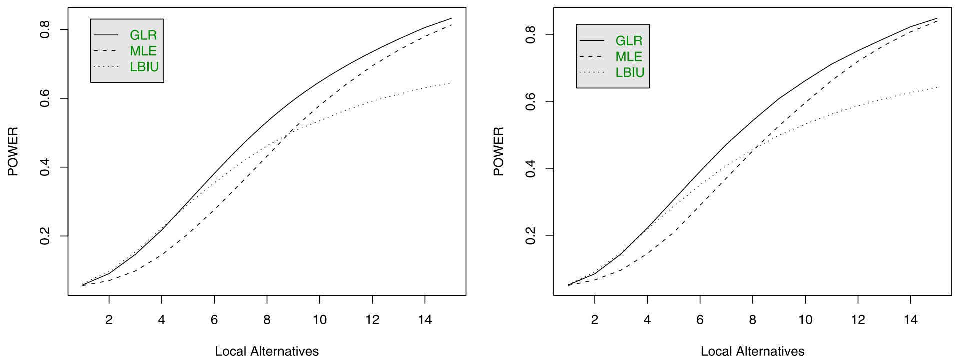

FIG. 2. *Power curve with respect to local alternatives when* $\alpha = -0.3$ (*upper*) *and when* $\alpha = -0.5$ (*lower*). *Sample size* $n = 50$. *The size of the test is set to be* 0.05.

margin. When the sample size is 50, the local alternative parameter corresponds to $\theta = 1 - 4/50 = 0.92$. Also, as seen in Figure 2, the power function based on the MLE dominates the power function of the LBIU test for local alternatives greater than 8 or 9.

In the case when $\alpha > 0$ especially for small sample sizes like 50, the behavior of the tests based on MLE and LBIU are very poor. This is because when $\alpha > 0$ and there is one unit root, the two parameters c_1 and c_2 lie on the boundary $-c_1 - c_2 = 1$ which is close to the complex region boundary $c_1^2 - 4c_2 = 0$. But our asymptotic results are derived in a way which assumes that the two roots are only approaching the limit through the real region. This holds asymptotically, but in finite sample cases, when we maximize the likelihood jointly over c_1 and c_2, it is likely that the two maximizers would fall into the complex region. As α gets closer to -1 this effect becomes more severe. Thus we do not recommend using the test based on the MLE when the invertible root is likely to be negative. Using the test based on MLE usually gives larger size of the test. The LBIU is not good in this case either as pointed out in Tanaka [20]. The upper tail probabilities are greatly underestimated when α gets closer to -1, and hence H_0 tends to be accepted much more often. Simulation results show that when the sample size is 50, and the true α is 0.3 and 0.5, the corresponding size of the LBIU test is 0.0119 and 0.0015 which are much smaller than the nominal size 0.05. GLR seems to be the best among the three choices. This is due to the fact that the GLR only considers the maximum value of the likelihood ratio instead of the MLE of c_1 and c_2. Therefore, even if $\hat{c}_1$ and $\hat{c}_2$ are in the complex region, the GLR test can still be carried out whereas the test based on $\hat{\beta}_{\mathrm{MLE}}$ is not even well defined in this case. Although the size of the GLR test is often slightly greater than the nominal size, GLR gives the best performance under this situation.

Finally, we compare these tests when $\alpha = 0$; that is, the model is in fact a unit root MA(1). The test developed for the MA(2) case is still applicable. The results are summarized in Figure 3. Clearly, the power functions of the tests designed for

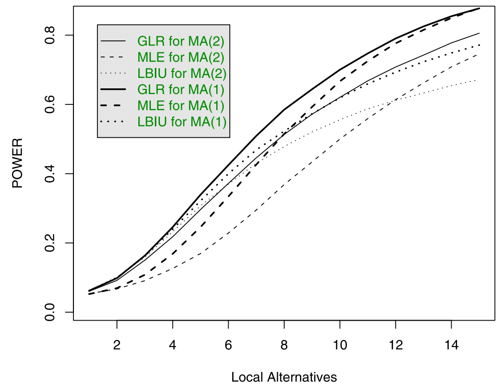

FIG. 3. *Power curve with respect to local alternatives when* $\alpha = 0$. *Sample size* $n = 50$. *The size of the test is set to be* 0.05.

the MA(1) dominate the power functions of their counterparts designed for the MA(2). However, it is surprising that for large local alternatives (greater than 9 or so), the GLR for the MA(2) model outperforms the LBIU for the MA(1) model.

5. Numerical simulations. In this section, we present simulation results that illustrate the theory from Section 3. Realizations were simulated from the MA(2) process given by

$$X_t = Z_t - (1+\alpha)Z_{t-1} + \alpha Z_{t-2}, \tag{5.1}$$

where α takes the values 0.3, 0 and -0.3, respectively. The MA(2) model was replicated 10,000 times for each choice of α, and then the MLEs for the MA(2) coefficients θ_1 and θ_2 were calculated for each replicate. The empirical pile-up probability, the empirical variance and MSE of the MLEs are reported in Tables 1 to 3. Notice that the numbers in the tables for the variance and the MSE are reported for the normalized estimates $\sqrt{n}(\hat{c}_i - c_i)$, $i = 1, 2$.

As seen in the tables, the correlation of $\hat{c}_1$ and $\hat{c}_2$ is increasing to 1 with the sample size. The variances and the MSEs are converging to the theoretical value

TABLE 1
Summary of the case: $\alpha = 0.3$

Sample size	**Pile-up probability**	**Variance of c_1**	**MSE of c_1**	**Variance of c_2**	**MSE of c_2**	**Correlation of c_1 and c_2**
25	0.5436	2.1701	2.1970	2.4455	2.6536	0.9347
50	0.6041	1.4063	1.4118	1.4967	1.5553	0.9644
100	0.6234	1.1108	1.1108	1.1490	1.1636	0.9815
400	0.6398	0.9788	0.9788	0.9854	0.9890	0.9953
1,000	0.6437	0.9290	0.9290	0.9327	0.9338	0.9981

TABLE 2
Summary of the case: $\alpha = 0$ [*MA*(1) *with a unit root*]

Sample size	**Pile-up probability**	**Variance of c_1**	**MSE of c_1**	**Variance of c_2**	**MSE of c_2**	**Correlation of c_1 and c_2**
25	0.5870	2.1624	2.1629	2.5037	2.6355	0.8792
50	0.6182	1.3661	1.3670	1.4690	1.5053	0.9378
100	0.6220	1.1661	1.1670	1.2082	1.2224	0.9662
400	0.6318	1.0440	1.0441	1.0544	1.0578	0.9918
1,000	0.6334	1.0329	1.0330	1.0351	1.0384	0.9966

$1 - c_2^2$. As pointed out in [10] and [9], the asymptotic results work remarkably well even for small sample sizes in the MA(1) case. Here, although the pile-up probability is still 0.6518, the rates vary depending on α. For $\alpha > 0$, rates are slow while for $\alpha < 0$ rates are much faster. From the derivation of the asymptotic results, there are error terms in the likelihood that vanish asymptotically and contribute to a more lethargic rate of convergence. Again the asymptotic results were derived assuming the roots are always in the real region, which only holds asymptotically. When the sample size is small and $\alpha > 0$, the MLEs of c_1 and c_2 are more likely to be in the complex region than those when $\alpha < 0$. Thus the limiting process would approximate the likelihood function poorly when $\alpha > 0$, which in turn results in less pile-up in smaller sample sizes.

Table 4 summarizes the pile-up effects for the model considered in Section 3.2.2, where the two roots of the MA polynomial are both 1. In one realization, the estimators are said to exhibit a pile-up if the MLEs of c_1 and c_2 are on the boundary $-c_1 - c_2 = 1$.

As seen in the table, the pile-up probability is increasing to 1 with sample size. However, the claimed 100% probability of pile-up is not a good approximation for small sample sizes. Even when $n = 500$, the pile-up is only about 80%.

TABLE 3
Summary of the case: $\alpha = -0.3$

Sample size	**Pile-up probability**	**Variance of c_1**	**MSE of c_1**	**Variance of c_2**	**MSE of c_2**	**Correlation of c_1 and c_2**
25	0.6171	1.8370	1.8806	2.1654	2.2287	0.7950
50	0.6347	1.2820	1.3053	1.3647	1.3820	0.8938
100	0.6447	1.0748	1.0853	1.1215	1.1299	0.9397
400	0.6472	0.9245	0.9267	0.9316	0.9339	0.9822
1,000	0.6511	0.9232	0.9242	0.9256	0.9263	0.9933

TABLE 4
Pile-up probabilities for the case: $c_1 = -2$

Sample size	Pile-up probability
100	0.246
500	0.804
1,000	0.961
5,000	0.999

6. Unit roots and differencing.

As pointed out in Section 3.2.2, there is a link between the mean term in the lower order MA model and the initial value in the higher order MA model. To illustrate this, consider the simple case when

$$Y_t = \mu_0 + Z_t,$$

where $\{Z_t\} \sim$ i.i.d. $(0, \sigma_0^2)$. So Y_t is an i.i.d. sequence with a common mean. It is clear that

$$\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_0^2),$$

where $\hat{\mu}$ is the MLE of μ obtained by maximizing the objective Gaussian likelihood function. Now suppose we difference the time series to obtain

$$X_t = (1 - \mathrm{B})Y_t = Z_t - Z_{t-1},$$

which becomes an MA(1) process with a unit root. The initial value as defined before of this differenced process is

$$Z_{\text{init}} = Z_0 = Y_0 - \mu_0.$$

From the results in Theorem 4.2 in [13], if it is known that an MA(1) time series has a unit root, that is, $\beta = 0$, we have

$$U(\beta = 0, \alpha) = 2\alpha W(1) + \alpha^2.$$

Clearly, $\tilde{\alpha} = -W(1)$ and with our parameterization of z_{init}, we have

$$\begin{aligned}\hat{\alpha} &= \frac{\sqrt{n}(z_{\text{init}} - Z_0)}{\sigma_0} = \frac{\sqrt{n}(Y_0 - \hat{\mu} - Y_0 + \mu_0)}{\sigma_0} \\ &= -\frac{\sqrt{n}(\hat{\mu} - \mu_0)}{\sigma_0} \xrightarrow{\mathrm{d}} \tilde{\alpha} = -W(1) \stackrel{\mathrm{d}}{=} \mathrm{N}(0, 1),\end{aligned}$$

which is consistent with the classical result. Therefore we can conclude that whenever we have an MA model with a unit root, the information stored in the initial value comes from the information of the mean term from the undifferenced series. So differencing the series will not get rid of the mean parameter; instead, differencing creates a new parameter Z_{init} which behaves like the mean in the undifferenced series and its effect persists even asymptotically. With this, we can now explain easily the result in (2.9). Turning to a little more complicated model consisting of i.i.d. noise and a linear trend, that is,

$$Y_t = \mu_0 + b_0 t + Z_t, \tag{6.1}$$

which, after differencing, delivers an MA(1) model with a unit root and a nonzero mean given by

$$X_t = (1 - \mathrm{B})Y_t = b_0 + Z_t - Z_{t-1}.$$

From (2.9), we know $n^{3/2}(\hat{b} - b_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, 12\sigma_0^2)$. But this can be obtained much more easily by analyzing the model (6.1). This is just a simple application of linear regression, and we can get exactly the same asymptotic result for $\hat{b}$.

Now consider the model from Section 2,

$$Y_t = b_0 + Z_t - \theta Z_{t-1},$$

where $\theta = 1 + \frac{\beta}{n}$ is near or on the unit circle. By differencing we obtain

$$X_t = (1 - \mathrm{B})Y_t = Z_t - (1 + \theta)Z_{t-1} + \theta Z_{t-2}.$$

If we define Z_{init} as before and

$$Y_{\text{init}} = Y_0 = b_0 + Z_0 - Z_{-1},$$

then $y_{\text{init}} - Y_{\text{init}}$ can be viewed as $\hat{b} - b_0$. Since $\hat{b}$ converges at the rate of $n^{3/2}$, so does y_{init}. This explains the parametrization given in (3.31) as well as the resemblance of (2.5) and (3.32).

7. Going beyond second order. The techniques proposed in this paper can be adapted to handle the unit root problem for MA(q) with $q \geq 3$. However, the complexity of the argument, mostly in terms of bookkeeping, also increases with the order q. In this section, we outline the procedure for the MA(3) case, from which extensions to larger orders are straightforward.

Suppose $\{X_t\}$ follows an MA(3) model, which is parameterized in terms of the reciprocals of the zeros of the MA polynomial, that is,

$$\begin{aligned} X_t &= Z_t - (\theta_0 + \phi_0 + \psi_0)Z_{t-1} \\ &\quad + (\theta_0\phi_0 + \theta_0\psi_0 + \phi_0\psi_0)Z_{t-2} - \theta_0\phi_0\psi_0 Z_{t-3} \\ &= (1 - \theta_0 B)(1 - \phi_0 B)(1 - \psi_0 B)Z_t \\ &= (1 - \theta_0 B)(1 - \phi_0 B)Y_t \\ &= (1 - \theta_0 B)W_t. \end{aligned} \tag{7.1}$$

For simplicity, assume $\theta_0 \neq \phi_0 \neq \psi_0$. Now we form two intermediate processes Y_t and W_t and consider three augmented initial variables defined by $Z_{\text{init}} = Z_{-2}$,

$Y_{\mathrm{init}} = Z_{-1} + \psi_0 Z_{\mathrm{init}}$ and $W_{\mathrm{init}} = Y_0 + \phi_0 Y_{\mathrm{init}}$. Similar arguments as in Section 3 show that the joint likelihood of ($\mathbf{X}$, W_{init}, Y_{init}, Z_{init}) has a simple form given by

$$f_{\mathbf{X}, W_{\mathrm{init}}, Y_{\mathrm{init}}, Z_{\mathrm{init}}}(\mathbf{x}_n, w_{\mathrm{init}}, y_{\mathrm{init}}, z_{\mathrm{init}}) = \prod_{j=-2}^{n} f_Z(z_j).$$

As in the MA(1) and MA(2) cases, maximizing this joint likelihood is essentially equivalent to minimizing the objective function

$$U_n = \frac{1}{\sigma_0^2} \sum_{i=-2}^{n} (z_i^2 - Z_i^2).$$

The key to this analysis is to write out the explicit expression for z_i which is basically an estimator for Z_i. The following equations are straightforward to derive:

$$w_k = \sum_{l=1}^{k} \theta^{k-l} X_l + \theta^k w_{\mathrm{init}}, \tag{7.2}$$

$$y_j = \sum_{k=1}^{j} \phi^{j-k} w_k + \phi^j w_{\mathrm{init}} + \phi^{j+1} y_{\mathrm{init}}, \tag{7.3}$$

$$z_i = \sum_{j=1}^{i} \psi^{i-j} y_j + \psi^i w_{\mathrm{init}} + \psi^i (\phi + \psi) y_{\mathrm{init}} + \psi^{i+2} z_{\mathrm{init}}. \tag{7.4}$$

Plugging (7.2) into (7.3), we obtain

$$y_j = \sum_{k=1}^{j} \frac{\theta^{j-k+1} - \phi^{j-k+1}}{\theta - \phi} X_k + \frac{\theta^{j+1} - \phi^{j+1}}{\theta - \phi} w_{\mathrm{init}} + \phi^{j+1} y_{\mathrm{init}}, \tag{7.5}$$

and plugging this into (7.4), we obtain

$$\begin{aligned} z_i = \sum_{j=1}^{i} \bigl(&\theta\phi(\theta^{i-j+1} - \phi^{i-j+1}) \\ &+ \theta\psi(\psi^{i-j+1} - \theta^{i-j+1}) + \phi\psi(\phi^{i-j+1} - \psi^{i-j+1})\bigr) \\ &\times \bigl((\theta - \psi)(\psi - \phi)(\phi - \theta)\bigr)^{-1} X_j \\ + &\left(\frac{\theta^2(\theta^i - \psi^i)}{(\theta - \phi)(\theta - \psi)} - \frac{\phi^2(\phi^i - \psi^i)}{(\theta - \phi)(\phi - \psi)} + \psi^i\right) w_{\mathrm{init}} \\ + &\frac{\phi^{i+2} - \psi^{i+2}}{\phi - \psi} y_{\mathrm{init}} + \psi^{i+2} z_{\mathrm{init}}. \end{aligned}$$

While this is a more complicated looking expression than the one encountered in the MA(2) case, the coefficient of X_j in the sum looks very similar to (3.2), only with more terms. Now replacing X_j with (7.1), z_i can be written as

$$
\begin{aligned}
z_i &= Z_i - \sum_{j=-2}^{i-1} C_{i,j}^z Z_j \\
&\quad - C_i^w (w_{\text{init}} - W_{\text{init}}) - C_i^y (y_{\text{init}} - Y_{\text{init}}) - C_i^z (z_{\text{init}} - Z_{\text{init}}) \\
&= Z_i - r_i,
\end{aligned}
\tag{7.6}
$$

where $C_{i,j}^z$ is the coefficient for Z_j in z_i and is a combination of θ^{i-j}, ϕ^{i-j} and ψ^{i-j}, and C_i^w, C_i^y and C_i^z are coefficients for $w_{\text{init}} - W_{\text{init}}$, $y_{\text{init}} - Y_{\text{init}}$ and $z_{\text{init}} - Z_{\text{init}}$. They are linear combinations of θ^i, ϕ^i and ψ^i. For illustration, assume the MA(3) model has only one unit root with $|\psi_0| < 1$, $|\phi_0| < 1$ and $\theta_0 = 1$. We can then reparameterize the parameters as

$$
\theta = 1 + \frac{\beta}{n}, \qquad \beta \le 0, \qquad \phi = \phi_0 + \frac{\alpha}{\sqrt{n}} \quad \text{and} \quad \psi = \psi_0 + \frac{\gamma}{\sqrt{n}},
$$

and the initial values as

$$
w_{\text{init}} = W_{\text{init}} + \frac{\sigma_0 \eta_w}{\sqrt{n}}, \qquad y_{\text{init}} = Y_{\text{init}} + \frac{\sigma_0 \eta_y}{\sqrt{n}} \quad \text{and} \quad z_{\text{init}} = Z_{\text{init}} + \frac{\sigma_0 \eta_z}{\sqrt{n}}.
$$

Then the objective function U_n becomes

$$
\begin{aligned}
&U_n(\beta, \alpha, \gamma, \eta_w, \eta_y, \eta_z) \\
&\quad = -2 \sum_{i=-2}^{n} \frac{r_i Z_i}{\sigma_0^2} + \sum_{i=-2}^{n} \frac{r_i^2}{\sigma_0^2} \\
&\quad = -2 \sum_{i=-2}^{n} \left(\sum_{j=-2}^{i-1} C_{i,j}^z \frac{Z_j}{\sigma_0} + \frac{C_i^w \eta_w}{\sqrt{n}} + \frac{C_i^y \eta_y}{\sqrt{n}} + \frac{C_i^z \eta_z}{\sqrt{n}} \right) \frac{Z_i}{\sigma_0} \\
&\qquad + \sum_{i=-2}^{n} \left(\sum_{j=-2}^{i-1} C_{i,j}^z \frac{Z_j}{\sigma_0} + \frac{C_i^w \eta_w}{\sqrt{n}} + \frac{C_i^y \eta_y}{\sqrt{n}} + \frac{C_i^z \eta_z}{\sqrt{n}} \right)^2.
\end{aligned}
\tag{7.7}
$$

Because of the special structure of $C_{i,j}^z$, C_i^w, C_i^y and C_i^z, the sum in (7.7) consists of terms that have a similar structure to quantities like

$$
\frac{1}{n} \sum_{i=-2}^{n} \sum_{j=-2}^{i-1} \left(1 + \frac{\beta}{n}\right)^{i-j} \frac{Z_j}{\sigma_0} \frac{Z_i}{\sigma_0} \quad \text{and} \quad \sum_{i=-2}^{n} \left(1 + \frac{\beta}{n}\right)^{i} \frac{Z_i}{\sqrt{n}\sigma_0},
$$

that were used in the MA(1) and MA(2) cases. By using a martingale central limit theorem and theorems proved in Davis and Song [13], one can establish the weak convergence of $U_n(\beta, \alpha, \gamma, \eta_w, \eta_y, \eta_z)$ to a random element $U(\beta, \alpha, \gamma, \eta_w, \eta_y, \eta_z)$ in $\mathbb{C}(\mathbb{R}^6)$. Now arguing as in Section 3, the initial variables can be integrated out, and the limiting process of the exact profile log-likelihood can be established.

For general $q > 3$, the residual $r_i = z_i - Z_i$ has the form

$$r_i = \sum_{j=-q+1}^{i-1} C_{i,j}^z Z_j + \sum_{k=1}^{q} C_i^k (\mathrm{init}_k - \mathrm{INIT}_k),$$

where $\{\mathrm{INIT}_1, \dots, \mathrm{INIT}_q\}$ are q augmented initial variables, defined either through the i.i.d. random variables Z_t or through the intermediate processes like Y_t in the above example. Furthermore, $C_{i,j}^z$ is only a linear combination of $(\theta_1^{i-j}, \dots, \theta_q^{i-j})$, where $(\theta_1, \dots, \theta_q)$ are reciprocals of the roots of the MA(q) polynomial. Coefficients C_i^k, $k = 1, \dots, q$, are only linear combinations of $(\theta_1^i, \dots, \theta_q^i)$. This special structure of r_i allows us to apply the weak convergence theorems proved in Davis and Song [13] to find the limiting process of $U_n = -2\sum_{i=-q+1}^{n} r_i Z_i/\sigma_0^2 + \sum_{i=-q+1}^{n} r_i^2/\sigma_0^2$, from which the limiting behavior of the maximum likelihood estimators of the θ_i's can be derived.

Acknowledgments. We would like to thank the referees and the Associate Editor for their insightful comments, which were incorporated into the final version of this paper.

REFERENCES

[1] ANDERSON, T. W. and TAKEMURA, A. (1986). Why do noninvertible estimated moving averages occur? *J. Time Series Anal.* **7** 235–254. MR0883008

[2] ANDREWS, B., CALDER, M. and DAVIS, R. A. (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Ann. Statist.* **37** 1946–1982. MR2533476

[3] ANDREWS, B., DAVIS, R. A. and BREIDT, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *J. Multivariate Anal.* **97** 1638–1659. MR2256234

[4] BREIDT, F. J., DAVIS, R. A., HSU, N.-J. and ROSENBLATT, M. (2006). Pile-up probabilities for the Laplace likelihood estimator of a non-invertible first order moving average. In *Time Series and Related Topics*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **52** 1–19. IMS, Beachwood, OH. MR2427836

[5] BREIDT, F. J., DAVIS, R. A. and TRINDADE, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Ann. Statist.* **29** 919–946. MR1869234

[6] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*. Springer, New York.

[7] CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877

[8] CHEN, M. C., DAVIS, R. A. and SONG, L. (2011). Inference for regression models with errors from a non-invertible MA(1) process. *J. Forecast.* **30** 6–30.

[9] DAVIS, R. A., CHEN, M. and DUNSMUIR, W. T. M. (1995). Inference for MA(1) processes with a root on or near the unit circle. *Probab. Math. Statist.* **15** 227–242. MR1369801

[10] 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. *Econometric Theory* **12** 1–29. MR1396378

[11] DAVIS, R. A. and DUNSMUIR, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. *J. Theoret. Probab.* **10** 481–497. MR1455154

[12] DAVIS, R. A., KNIGHT, K. and LIU, J. (1992). *M*-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40** 145–180. MR1145464

[13] DAVIS, R. A. and SONG, L. (2012). Functional convergence of stochastic integrals with application to statistical inference. *Stochastic Process. Appl.* **122** 725–757.

[14] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

[15] LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York. MR1663158

[16] ROSENBLATT, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer, New York. MR1742357

[17] SARGAN, J. D. and BHARGAVA, A. (1983). Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. *Econometrica* **51** 799–820. MR0712371

[18] SHEPHARD, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components. *J. Amer. Statist. Assoc.* **88** 590–595. MR1224385

[19] SMITH, R. L. (2008). Statistical trend analysis. In *Weather and Climate Extremes in a Changing Climate* (*Appendix A*) 127–132.

[20] TANAKA, K. (1990). Testing for a moving average unit root. *Econometric Theory* **6** 433–444. MR1094221

[21] TANAKA, K. (1996). *Time Series Analysis*: *Nonstationary and Noninvertible Distribution Theory*. Wiley, New York. MR1397269

DEPARTMENT OF STATISTICS
1255 AMSTERDAM AVE
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
USA

BARCLAYS CAPITAL
745 7TH AVE
NEW YORK, NEW YORK 10019
USA