Gaussian Likelihood Based Inference for Non-Invertible MA(1) Processes With SαS Noise

RICHARD A. DAVIS
Colorado State University

and

THOMAS MIKOSCH
University of Groningen

ABSTRACT

A limit theory was developed in the papers Davis and Dunsmuir (1996) and Davis, Chen, and Dunsmuir (1995) for the maximum likelihood estimator, based on a Gaussian likelihood, of the moving average parameter $\theta$ in an MA(1) model when $\theta$ is equal to or close to 1. Using the local parameterization, $\beta = T(1 - \theta)$, where $T$ is the sample size, it was shown that the likelihood, as a function of $\beta$, converged to a stochastic process, from which the limit distribution of $T(\hat{\theta} - 1)$ ($\hat{\theta}$ is the MLE) was established. As a byproduct of the likelihood convergence, the limit distribution of the likelihood ratio test for testing $H_0: \theta = 1$ vs. $\theta < 1$ was also determined. In this paper, we again consider the limit behavior of the maximum Gaussian likelihood estimator of $\theta$ and the corresponding likelihood ratio statistics when the non-invertible MA(1) process is generated by symmetric $\alpha$-stable noise with $\alpha \in (0, 2)$. Estimates of a similar nature have been studied for causal-invertible ARMA processes generated by infinite variance stable noise. In those situations, the scale normalization improves from the traditional $T^{1/2}$ rate obtained in the finite variance case to $(T/\ln T)^{1/\alpha}$. In the non-invertible setting of this paper, the rate is the same as in the finite variance case. That is, $T(\hat{\theta} - 1)$ converges in distribution and the pile-up effect, i.e., $\lim_{T \to \infty} P(\hat{\theta} = 1)$, is slightly less than in the finite variance case. It is also of interest to note that the limit distributions of $T(\hat{\theta} - 1)$ for different values of $\alpha \in (0, 2]$ are remarkably similar.

*This research supported in part by NSF DMS Grant No. DMS-9504596.
AMS 1980 Subject Classification: Primary 62M10; Secondary 62M09, 62M07
Key Words and Phrases. Moving average process, unit roots, non-invertible moving averages, maximum likelihood estimation, stable distribution, stable integral.
1 Introduction

The objective of this paper is to study the asymptotic properties for a class of estimators of the MA(1) parameter in the non-invertible or near non-invertible case when the noise distribution is heavy-tailed. Specifically, the model under consideration is the MA(1) process

\[ Y_t = \epsilon_t - \theta_0 \epsilon_{t-1}, \]

where \( \{\epsilon_t\} \sim \text{iid} \) with a symmetric \( \alpha \)-stable (S\( \alpha \)S) distribution, \( \alpha \in (0, 2) \), and \( \theta_0 \) is either equal to 1 or quite close to 1. With this restriction on \( \alpha \), the variance of \( \epsilon_t \), and hence \( Y_t \), is infinite. In the finite variance case, the 'near non-invertible' MA(1) arises quite frequently in a variety of time series modelling contexts. For example, a test that a time series has been over-differenced to achieve stationarity is equivalent to testing for the presence of a unit root in the moving average polynomial. Further applications and situations where the unit roots in moving averages occur can be found in the papers by Anderson and Takemura (1986), Stock (1994), Davis and Dunsmuir (1996), and in the books by Fuller (1996) and Tanaka (1996).

In Davis and Dunsmuir (1996), two estimators of \( \theta_0 \) were considered. The first, which was referred to as the local maximum estimate and denoted by \( \hat{\theta}_{LM} \), is defined as the local maximizer of the Gaussian likelihood closest to \( \theta = 1 \). The second, \( \hat{\theta}_{MLE} \), is the maximum (Gaussian) likelihood estimator (i.e., it is the value of \( \theta \) which maximizes the likelihood (based on a Gaussian distribution) over the interval \( \theta \in [-1, 1] \)). Unlike the standard statistical setting where any two consistent solutions to the likelihood equations are asymptotically equivalent, it was shown in Davis and Dunsmuir (1996) that \( \hat{\theta}_{LM} \) and \( \hat{\theta}_{MLE} \) have distinct limit distributions for a sequence of local alternatives converging to \( \theta_0 = 1 \). This same phenomenon remains valid in the non-normal stable noise setting.

In Section 2, we establish the limiting behavior of \( \hat{\theta}_{LM} \) for the model in (1.1) under a sequence of local values of \( \theta_0 \) which converge to 1 at rate \( 1/T \), where \( T \) is the number of observations in the time series. In particular, if \( \theta_0 = 1 - \gamma/T \) with \( \gamma \geq 0 \) a fixed constant, then \( T(1 - \hat{\theta}_{LM}) \xrightarrow{d} \beta_\gamma \), where \( \beta_\gamma \) is the minimizer of some stochastic process. The limit distribution has both a discrete component at the value 0, called the pile-up effect, and a continuous component.

There is now a number of papers in the literature concerned with the asymptotic behavior of maximum Gaussian-likelihood and other second-order based estimates for causal-invertible ARMA processes generated by non-normal stable noise; see for example Brockwell and Davis (1991), Section 13.3, Davis (1996), Davis et al. (1992), Davis and Resnick (1986), Kokoszka and Taqqu (1996), Mikosch et al. (1995). In these situations, the scale normalization typically improves from the traditional \( T^{1/2} \) rate obtained in the finite variance case to \( (T/\ln T)^{1/\alpha} \). However, for the non-invertible MA(1) model (1.1), the rate is the same as in the finite variance case. That is, \( T(\hat{\theta} - 1) \) converges in distribution in the finite variance case and for all \( \alpha \in (0, 2) \).
Section 3 contains the proof of the convergence in distribution of \( \hat{\theta}_{LM} \). In Section 4, we compare the limit distributions of \( T(\hat{\theta} - 1) \) for the four values of \( \alpha = 0.75, 1.0, 1.5, \) and \( 2.0 \). Perhaps surprisingly, the limit distributions of \( T(\hat{\theta} - 1) \) for different values of \( \alpha \) are remarkably similar. There are some important differences, however, that are worth pointing out. First, the pile-up effect, i.e., \( \lim_{T \to \infty} P[\hat{\theta} = 1] \), increases slightly with \( \alpha \). On the other hand, the tails of the limit distribution are heavier for increasing \( \alpha \). For example, the .05-quantile of the limit distribution for \( \alpha = 0.75 \) is larger than that for \( \alpha = 1.5 \). As in the finite variance case, the limit approximations are reasonably good for samples as small as 25 and for values of \( \theta_0 \) relatively far from 1.

Section 5 contains a number of technical results required in the proof of the main result.

2 Asymptotic Theory

Let \( \{Y_t\} \) be the MA(1) process defined by

\[
Y_t = \epsilon_t - \theta_0 \epsilon_{t-1},
\]

where \( |\theta_0| \leq 1 \) and \( \{\epsilon_t\} \) is an iid sequence of symmetric \( \alpha \)-stable random variables (SSS) with characteristic function \( E \exp\{it\epsilon\} = \exp\{-|t|^\alpha\} \) and \( \alpha \in (0, 2) \). Unlike the case with Gaussian noise, the parameter \( \theta_0 \) is now identifiable for all real values. However, since we are concerned with estimation of \( \theta_0 \) based solely on the Gaussian likelihood, we restrict the parameter space to \( |\theta_0| \leq 1 \) in order to achieve a parameter identification (see Brockwell and Davis (1991), p. 272).

Since we are interested in inference about \( \theta_0 \) when \( \theta_0 \) is at or near 1, we adopt the parameterization \( \theta = \theta_T = 1 - \beta / T \), where \( \beta \geq 0 \) and \( T \) is the sample size. Inference about \( \beta \) and hence \( \theta_0 \) will be based on the observations \( Y_1, \ldots, Y_T \) which are assumed to come from model (2.1) with true parameter \( \theta_0 = 1 - \gamma / T \), where \( \gamma \geq 0 \).

Even though this process has infinite variance, we still refer to \( \rho := \rho(\theta) = -\theta / (1 + \theta^2) \) as the lag 1 correlation of the model. After concentrating out the variance parameter of the 'Gaussian likelihood', the resulting concentrated likelihood, as a function of \( \rho \), is given by (see equation (10) of Anderson and Takemura (1986)),

\[
M(\rho) = -\log |G| - T \log Y'G^{-1}Y,
\]

where \( Y' = (Y_1, \ldots, Y_T) \) is the data vector and \( G \) is the correlation matrix of a finite variance MA(1) process with lag 1 correlation equal to \( \rho \). The concentrated likelihood as a function of \( \beta \) is then given by

\[
L_T(\beta) = M(\rho(1 - \beta / T))
\]

with

\[
L_T'(\beta) = \frac{dM}{d\rho} \frac{d\rho}{d\theta} \frac{d\theta}{d\beta}
\]
\[ \frac{dM}{dp} = \frac{2\beta/T - \beta^2/T^2}{2(1 - \beta/T) + \beta^2/T^2} \frac{1}{T} \]
\[ = \frac{\beta}{2T^2} \left( 1 - \frac{\beta}{2T} \right) a^2(\beta, T) \frac{dM}{dp}, \]

where

\[ a(\beta, T) = \left( 1 - \frac{\beta}{T} + \beta^2/(2T^2) \right)^{-1}. \]

Clearly, \( L_T \) is zero at \( \beta = 0, 2T \) and hence \( M'(\pm 1/2) = 0 \). Consequently, \( L_T \) vanishes at every local maximum. Using equations (14)-(20) in Anderson and Takemura (1986) (see also equations (2.2)-(2.6) of Davis and Dunsmuir (1996)), we have under the true \( \theta_0 = 1 - \gamma/T \), that

\[ \frac{dM}{dp} = -\sum_{i=1}^{T} \frac{2d_t}{1 + 2p_r d_t} + \left( \frac{T}{T} \sum_{t=1}^{T} \frac{1 + 2q_r d_t}{1 + 2p_r d_t} U_{t,t}^2 \right)^{-1} \sum_{t=1}^{T} \frac{2d_t(1 + 2q_r d_t)}{(1 + 2p_r d_t)^2} U_{t,t}^2, \]

where \( d_t = \cos \omega_t \),

\[ \omega_t = \frac{\pi t}{T + 1}, \]

\[ p_r = -\left( 1 - \frac{\beta}{T} \right) / \left( 1 + \left( 1 - \frac{\beta}{T} \right)^2 \right), \]

\[ = -\frac{1}{2} + \frac{\beta^2}{4T^2} a(\beta, T), \]

\[ q_r = -\frac{1}{2} + \frac{\gamma^2}{4T^2} a(\gamma, T), \]

\[ U_{t,t} = (2/(T + 1))^{1/2}(1 + 2q_r d_t)^{-1/2} \sum_{s=1}^{T} Y_s \sin \omega_{st}. \]

Similarly, we find that

\[ L''_T(\beta) = \frac{\beta^2}{4T^4} \left( 1 - \frac{\beta}{2T} \right)^2 a^2(\beta, T) \frac{d^2 M}{dp^2} + \frac{b(\beta, T)}{2T^2} \frac{dM}{dp}, \]

where \( b(\beta, T) \) is the derivative of \( \beta(1 - \beta/(2T))a(\beta, T) \) with respect to \( \beta \) which converges to 1 uniformly on compact sets.

The following theorem describes the joint limiting behavior of \( L'_T \) and \( L''_T \) as random elements with values in \( C[0, \infty) \). Here \( C[0, \infty) \) denotes the set of continuous functions on \([0, \infty)\) endowed with the uniform topology on compact sets. Thus weak convergence on \( C[0, \infty) \) is equivalent to weak convergence on \( C[0, N] \) for every \( N > 0 \); see Pollard (1984).

**Theorem 2.1** Suppose \( Y_1, \ldots, Y_T \) are observations from model (2.1) with \( \theta_0 = 1 - \gamma/T \) for some \( \gamma \geq 0 \). Let \( \{X_t\} \) be a sequence of random variables with representation

\[ X_0 = \left( 2 \sum_{0 \leq x \leq 1} (\Delta M(x))^2 \right)^{1/2} \]
and
\[ X_t = 2 \int_0^1 \frac{-\pi t \cos(\pi tx) + \gamma \sin(\pi tx)}{\left(\pi^2 t^2 + \gamma^2\right)^{1/2}} dM(x), \ t = 1, 2, \ldots, \]

where \( M \) is an \( \alpha S \) Lévy motion on \([0, 1]\) with \( M(1) \overset{d}{=} \epsilon_1 \). (See Remark 2.2 below for the interpretation of these integrals and Lemma 3.5 for the joint distribution of the \( X_t \).) Further, let \( S \) denote the state space \( S = \{ f = (f_1, f_2) : \text{where } f_1, f_2 \in C[0, \infty) \text{ and } f_1(0) = 0 \} \). Then with \( L_T \) defined by (2.2),

(i) \( (L_T', L_T'') \overset{d}{\to} (Z_\gamma', Z_\gamma'') \) as \( T \to \infty \),

where \( \overset{d}{\to} \) denotes weak convergence on \( S \),

\[ Z_\gamma(\beta) = \sum_{k=1}^{\infty} \frac{\beta^2 (\pi^2 k^2 + \gamma^2)}{(\pi^2 k^2 + \beta^2)^2} \bar{X}^2_k + \sum_{k=1}^{\infty} \ln \left( \frac{\pi^2 k^2}{\pi^2 k^2 + \beta^2} \right). \]

with \( \bar{X}_k = X_k / X_0 \), and \( Z_\gamma' \) and \( Z_\gamma'' \) denote the first and second derivatives of \( Z_\gamma \) with respect to \( \beta \).

(ii) \( L_T(\beta) - L_T(0) \overset{d}{\to} Z_\gamma(\beta) \) on \( C[0, \infty) \).

(iii) If \( \hat{\beta}_{LM} \) is the local maximizer (on the interval \([0, 2T]\)) of \( L_T \) closest to 0, then

\[ \hat{\beta}_{LM} \overset{d}{\to} \bar{\beta}_\gamma, \]

where \( \bar{\beta}_\gamma \) is the local maximizer (on the set \( \beta \geq 0 \)) of \( Z_\gamma \) closest to 0.

**Remark 2.1.** Note that we may write

\[ Z_\gamma(\beta) = \frac{1}{2} \int_0^\beta \tau Y_\gamma(\tau) d\tau, \]

where

\[ Y_\gamma(\beta) = 4 \sum_{k=1}^{\infty} \frac{\pi^2 k^2 + \gamma^2}{(\pi^2 k^2 + \beta^2)^2} \bar{X}^2_k - 4 \sum_{k=1}^{\infty} \frac{1}{\pi^2 k^2 + \beta^2}. \]

Using the \( Y_\gamma \) process, Theorem 2.1, (i) can be restated as

\[ (L_T'(\beta), L_T''(\beta)) \overset{d}{\to} \frac{1}{2} \left( \beta Y_\gamma(\beta), \beta Y_\gamma'(\beta) + Y_\gamma(\beta) \right). \]

In addition, Theorem 2.1, (ii) is immediate from (i) and the continuous mapping theorem.

In the case when \( Ec^2 < \infty \) the limit processes \( Z_\gamma \) and \( Y_\gamma \) have exactly the same structure as in the infinite variance case; see e.g. Davis and Dunsmuir (1996). The only difference is that the random variables \( \bar{X}_k \) have to be replaced by iid \( N(0, 1) \) random variables. In the infinite variance case, the \( \bar{X}^2_k \) have properties surprisingly similar to an iid \( \chi^2_1 \) sequence: they are uncorrelated, have mean 1 and an exponentially decreasing tail. (See Lemma 3.5 and Proposition 5.3 for details.)
Remark 2.2. We always take a càdlàg version of the SoS Lévy motion $M$. In this case, the integral in (2.7) can be interpreted pathwise as a Riemann-Stieltjes integral (see Young (1936), Dudley (1992) and Dudley and Norvaisa (1997)). This is due to the fact that $M$ has bounded $p$-variation, $p > \alpha$ (see Frisstedt and Taylor (1973)). For this reason, the quadratic variation in $M$, i.e., $\sum_{0 \leq \tau \leq 1} (\Delta M(\tau))^2$, converges a.s. Alternatively, the integral in (2.7) could be defined as an Itô integral (see Protter (1992)) or as an integral with respect to a SoS random measure (see Samorodnitsky and Taqqu (1994)). The definition of the pathwise integral will be made precise in the proof of Proposition 5.1.

Remark 2.3. We say that $\beta^*$ is a local maximizer of $L_T$ if there exists a $\delta > 0$ such that $L_T(\beta^*) \geq L_T(\beta)$ for all $\tau \in [0, 2T]$ with $|\tau - \beta^*| < \delta$. Clearly, if a local maximizer $\beta^*$ occurs on the interior of $(0, 2T)$, then $L_T'(\beta^*) = 0$ and $L_T''(\beta^*) \leq 0$. As seen from (2.3) and (2.4), $L_T'$ is a rational function of $\beta$ and hence can have at most a finite number of zeros, provided the likelihood is not identically constant. Assuming the latter, which with probability 1 will be the case for all $T$ large enough, there is at least one and at most a finite number of local maximizers of $L_T$ of which $\hat{\beta}_{LM}$ is defined as the minimum. If $L_T'(\beta)$ is constant on any interval, and hence identically constant, then $\hat{\beta}_{LM}$ is 0 according to our definition.

Remark 2.4. By the differentiability of $Z_\gamma$ and $Z'_\gamma$ with respect to $\beta$, a local maximizer $\beta^*$ of $Z_\gamma$ must satisfy $Z'_\gamma(\beta^*) = 0$ and $Z''_\gamma(\beta^*) \leq 0$. However, by (5.3) of the Appendix, $Z'_\gamma$ and $Z''_\gamma$ cannot have common zeros with probability one, so that $\beta^*$ is a local maximizer of $Z_\gamma$ if and only if $Z'_\gamma(\beta^*) = 0$ and $Z''_\gamma(\beta^*) < 0$. It follows that the local maximizers of $Z_\gamma$ are isolated and from the continuity of the sample paths of $Z'_\gamma$ and $Z''_\gamma$ and (5.3), the infimum of these local maximizers is also a local maximizer. (The existence of a local maximizer is ensured by (5.4) of the Appendix.) Consequently, with probability one,

$$\hat{\beta}_\gamma = \inf \{ \beta \geq 0 : \beta Y_\gamma(\beta) = 0 \text{ and } \beta Y'_\gamma(\beta) + Y_\gamma(\beta) < 0 \}.$$ 

Remark 2.5. The value of $\hat{\beta}_\gamma$ can be determined directly from the sample path of $Y_\gamma$ without explicit knowledge of the $Y'_\gamma$ process. Since $Z'_\gamma(0) = 0$ and $Z''_\gamma(0) = Y_\gamma(0)/2$, it is immediate from the preceding remark that $\hat{\beta}_\gamma = 0$ if and only if $Y_\gamma(0) < 0$. On the other hand, if $Y_\gamma(0) > 0$ (i.e., $\hat{\beta}_\gamma > 0$), then $Z'_\gamma(\hat{\beta}_\gamma) = 0$ and $Z''_\gamma(\hat{\beta}_\gamma) < 0$ which are equivalent to $Y_\gamma(\hat{\beta}_\gamma) = 0$ and $Y'_\gamma(\hat{\beta}_\gamma) < 0$. In other words, $\hat{\beta}_\gamma$ must be a down-crossing of 0 by the process $Y_\gamma$ and since $Y_\gamma(0) > 0$, it follows that $\hat{\beta}_\gamma$ must be the smallest zero of $Y_\gamma(\beta)$.

Remark 2.6. Theorem 2.1 covers the non-invertible case when $\theta_0 = 1$ by taking $\gamma = 0$.

Remark 2.7. The argument given below in Section 3 can be adapted to prove the corresponding theorem of Davis and Dunsmuir (1996) under slightly weaker conditions on the noise process. In particular, the iid assumption on $\{e_t\}$ can be relaxed to a stationary finite variance martingale.
difference sequence satisfying a Lindeberg-Feller condition for the CLT. The orthogonality of the noise is the critical ingredient of the proof.

**Corollary 2.2** Let \( \hat{\theta}_{LM} = (1 - \hat{\theta}_{LM}/T) \) (i.e., \( \hat{\theta}_{LM} \) is the local maximizer of the likelihood which is closest to the boundary at 1) and let \( P_\gamma \) denote the probability law under the model (2.1) with \( \theta_0 = 1 - \gamma/T \). Then

\[
(a) \quad T(\hat{\theta}_{LM} - 1) \xrightarrow{d} -\bar{\beta}_\gamma.
\]

\[(b) \quad P_\gamma[\hat{\theta}_{LM} = 1] \to P[(1/6 - W_1)/W_2 \geq \gamma^2] = P[\bar{\beta}_\gamma = 0].
\]

where \( W_1 = \sum_{k=1}^{\infty} \frac{\hat{X}_k^2}{\pi^2 k^2} \) and \( W_2 = \sum_{k=1}^{\infty} \frac{\hat{X}_k^2}{\pi^4 k^4} \).

\[(c) \quad \text{For all } x > 0, P_\gamma[\hat{\theta}_{LM} > x] \to P[\bar{\beta}_\gamma > x] = P[Y_\gamma(0) > 0, \bar{\beta}_\gamma > x].
\]

**Proof:** (a) is immediate from Theorem 2.1,(ii).

(b) There is a local maximizer of \( L_T \) at 0 (\( \hat{\theta}_{LM} = 1 \)) if and only if \( L'_T(0) < 0 \) or \( L''_T(0) = 0 \) and \( L_T(\tau) \leq L_T(0) \) for all \( \tau \in [0, \delta) \) for some \( \delta > 0 \). Now from the weak convergence in Theorem 2.1,(i),

\[
\limsup_{T \to \infty} P_\gamma[L''_T(0) = 0] \leq P(Y_\gamma(0) = 0) = 0,
\]

where the equality follows from the fact that the distribution of \( Y_\gamma(0) \) is continuous; see Corollary 5.2. Consequently,

\[
P_\gamma[\hat{\theta}_{LM} = 1] = P_\gamma[\hat{\beta}_{LM} = 0] = P_\gamma[L'_T(0) < 0] + o(1)
\]

\[
\to P[Y_\gamma(0) < 0]
\]

\[
= P\left[\sum_{k=1}^{\infty} \frac{\hat{X}_k^2}{\pi^2 k^2} + \gamma^2 \sum_{k=1}^{\infty} \frac{\hat{X}_k^2}{\pi^4 k^4} < \sum_{k=1}^{\infty} \frac{1}{\pi^2 k^2}\right]
\]

\[
= P\left[\frac{1/6 - W_1}{W_2} > \gamma^2\right].
\]

Using Remark 2.5, it follows that \( P[\bar{\beta}_\gamma = 0] = P[Y_\gamma(0) < 0] \) as claimed.

(c) The distribution of \( \bar{\beta}_\gamma \) is continuous except at 0, since if \( P[\bar{\beta}_\gamma = c] > 0 \) for some nonzero constant \( c \), then \( P[Y_\gamma(c) = 0] > 0 \). But this is impossible since \( Y_\gamma(c) \), like \( Y_\gamma(0) \), must also have a continuous distribution. It follows from Remark 2.5 that \( \bar{\beta}_\gamma > 0 \) if and only if \( Y_\gamma(0) > 0 \), from which (c) is now immediate. \( \Box \)

**Remark 2.8.** The result given in (b) with \( \gamma = 0 \) gives the probability of a pile up at 1, i.e.,

\[
\lim_{T \to \infty} P[\hat{\theta} = 1] = P[\hat{\beta}_0 = 0] = P[W_1 \leq 1/6].
\]
In the finite variance case, this probability is equal to .6575; see Anderson and Takemura (1986) and Tanaka and Satchell (1989). As noted in Section 4, these probabilities decrease gradually with decreasing \( \alpha \). For example, the pile-up probability is .62 for \( \alpha = 1.0 \) and .61 for \( \alpha = .75 \).

Theorem 2.1(ii) suggests that the global maximizer of \( L_T \) might converge in distribution to the global maximizer of \( Z_T \). Of course, in general, convergence on \( C[0, \infty) \) does not necessarily imply convergence of the corresponding maximizers without additional assumptions on the underlying functions. An additional argument was given in Davis, Chen, and Dunsmuir (1995) to establish such convergence in the finite variance case. They also argued that the limit distributions for the two estimators \( \hat{\theta}_{LM} \) and \( \hat{\theta}_{MLE} \) are different. These results, using essentially the same arguments, can be extended to the infinite variance stable case.

3 Proof of Theorem 2.1

With \( p_T = p_T(\beta) = \rho(1 - \beta/T) \) given in (2.5), it suffices to show that for any fixed positive \( N \)

\[
\left( T^{-3} \frac{dM}{d\rho}(p_T), T^{-4} \frac{d^2 M}{d\rho^2}(p_T) \right) \overset{d}{\to} \left( Y_\gamma(\beta), 2Y'_\gamma(\beta) \right) \quad \text{in} \quad C^2[0, N].
\]

Before embarking on the proof of this result, we introduce some notation and consider a preliminary lemma. Define

\[
\sigma_T^2(0) = \frac{1 + \theta_0^2}{T + 1} \sum_{t=0}^{T} \epsilon_t^2
\]

and, for \( t = 1, \ldots, T \), put

\[
\tilde{Y}_t = \sigma_T^{-1}(0) Y_t
\]

and

\[
\tilde{U}_{t,T} = (2/(T + 1))^{1/2}(1 + q_T d_T)^{-1/2} \sum_{s=1}^{T} \tilde{Y}_s \sin \omega_{st},
\]

where \( \omega_t = \pi t/(T + 1) \).

**Lemma 3.1** The random variables \( \tilde{U}_{1,T}, \ldots, \tilde{U}_{T,T} \) are uncorrelated with mean 0 and variance 1.

**Proof:** First note that by the symmetry of the distribution of \( \epsilon_t \), \( E(\tilde{\sigma}^{-1}(0) \epsilon_t) = 0 \) and \( E(\tilde{\sigma}^{-2}(0) \epsilon_s \epsilon_t) = 0 \) whenever \( s \neq t \). On the other hand, it follows that

\[
E(\tilde{\sigma}^{-2}(0) \epsilon_t^2) = (T + 1)^{-1} E \left( \tilde{\sigma}^{-2}(0) \sum_{s=0}^{T} \epsilon_s^2 \right) = (1 + \theta_0^2)^{-1},
\]

from which we conclude that the variables \( \tilde{Y}_1, \ldots, \tilde{Y}_T \) have mean zero and the covariance function of an MA(1) process, namely,

\[
\text{Cov}(\tilde{Y}_s, \tilde{Y}_t) = \begin{cases} 
1, & \text{if } s = t, \\
-\frac{\theta_0}{1 + \theta_0^2}, & \text{if } |s - t| = 1, \\
0, & \text{otherwise}.
\end{cases}
\]
The result now follows from an application of the spectral decomposition of the covariance matrix of the \( \tilde{Y}_t \) as given in Anderson and Takemura (1986).

Observe that \( \frac{dM}{dp}(p_r) \) can be rewritten as

\[
\frac{dM}{dp} = -\sum_{t=1}^{T} \frac{2d_t}{1 + 2p_r d_t} \left( \frac{1}{T} \sum_{t=1}^{T} \frac{1 + 2q_r d_t}{1 + 2p_r d_t} \tilde{U}_{t,T}^2 \right)^{-1} \sum_{t=1}^{T} \frac{2d_t(1 + 2q_r d_t)}{(1 + 2p_r d_t)^2} \tilde{U}_{t,T}^2 .
\]

If \( k_T \) is a sequence of integers satisfying \( k_T \to \infty \), \( k_T/T \to 0 \) and \( k_T^2/T \to \infty \), then the analogues of (2.12), (2.13), and (2.15) in Davis and Dunsmuir (1996) given by

\[
(3.1) \quad \frac{1}{(T+1)^2} \sum_{t=1}^{T} \frac{2d_t}{1 + 2p_r d_t} - \sum_{t=1}^{k_T} \frac{4}{\pi^2 t^2 + \beta^2} \to 0 \text{ uniformly on } \beta \in [0, N],
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{1 + 2q_r d_t}{1 + 2p_r d_t} \tilde{U}_{t,T}^2 - \frac{1}{T} \sum_{t=1}^{T} \tilde{U}_{t,T}^2 \to 0 ,
\]

and

\[
(3.2) \quad \frac{1}{(T+1)^2} \sum_{t=1}^{T} \frac{2d_t(1 + 2q_r d_t)}{(1 + 2p_r d_t)^2} \tilde{U}_{t,T}^2 - \sum_{t=1}^{k_T} \frac{4(\pi^2 t^2 + \gamma^2)}{(\pi^2 t^2 + \beta^2)^2} \tilde{U}_{t,T}^2 \to 0 ,
\]

remain valid. (The notation \( \to_d (\tilde{U}_{t,T}^2) \) refers to convergence in probability (distribution) with respect to the uniform metric on \( C[0,N] \).) The proof given in Davis and Dunsmuir (1996) only requires the \( \tilde{U}_{t,T} \) to have a uniformly bounded second moment.

The remainder of the proof of (i) is devoted primarily to establishing the following limits,

\[
(3.2) \quad \frac{1}{T} \sum_{t=1}^{T} \tilde{U}_{t,T}^2 \to 1 ,
\]

and

\[
(3.3) \quad \sum_{t=1}^{k_T} \frac{4(\pi^2 t^2 + \gamma^2)}{(\pi^2 t^2 + \beta^2)^2} \tilde{U}_{t,T}^2 \to \sum_{t=1}^{\infty} \frac{4(\pi^2 t^2 + \gamma^2)}{(\pi^2 t^2 + \beta^2)^2} \tilde{X}_t^2 .
\]

Relations (3.2) and (3.3) are direct consequences of Lemmas 3.4 and 3.5 below. The proof of (3.3) is identical to the argument given for (2.16) in Davis and Dunsmuir (1996) which essentially only uses the joint convergence of \( \tilde{U}_{t,T} \) to \( \tilde{X}_t^2 \) and the property that the \( \tilde{X}_t^2 \) have uniformly bounded expectations (see Proposition 3.5 below). In addition, Proposition A.1 of Davis and Dunsmuir (1996) has the exact analogue (see Proposition 5.4) and so the remainder of the proof of Theorem 2.1 is the same as the argument given on pp. 15–18 of Davis and Dunsmuir (1996) and hence is omitted.

The proof of (3.2) and the joint weak convergence of the \( \tilde{U}_{t,T} \) is treated in a series of lemmas given below.
Lemma 3.2 For \( r, s = 1, \ldots, T \), let \( r_{s,t} \) be the covariance between \( W_s \) and \( W_t \), where

\[
W_s = \left( \frac{2}{(T + 1)} \right)^{1/2} \sum_{j=1}^{T} (1 + 2q_T d_j)^{-1/2} Z_j \left( \sin \omega_{sj} - \left( 1 - \frac{\gamma}{T} \right) \sin \omega_{(s+1)j} \right),
\]

and \( Z_1, \ldots, Z_t \) are iid \( N(0,1) \) random variables. Then, as \( T \to \infty \),

\[
r_{s,s} = 2(1 + o(1)),
\]

and

\[
|r_{s,t}| \leq O(k_T T^{-1} + T k_T^{-2}),
\]

where the terms \( o(1) \) and \( O(k_T T^{-1} + T k_T^{-2}) \) are uniform in \( 1 \leq s < t \leq T \).

Proof: We have

\[
r_{s,t} = \frac{2}{T + 1} \sum_{j=1}^{T} (1 + 2q_T d_j)^{-1} \left\{ \left( \sin \omega_{sj} - \left( 1 - \frac{\gamma}{T} \right) \sin \omega_{(s+1)j} \right) \left( \sin \omega_{tj} - \left( 1 - \frac{\gamma}{T} \right) \sin \omega_{(t+1)j} \right) \right\}
\]

\[
= \frac{2}{T + 1} \sum_{j=1}^{T} (1 + 2q_T d_j)^{-1} \left\{ \left( \sin \omega_{sj} \left( 1 - d_j + \frac{\gamma}{T} d_j \right) - \left( 1 - \frac{\gamma}{T} \right) \cos \omega_{sj} \sin \omega_j \right) \left( \sin \omega_{tj} \left( 1 - d_j + \frac{\gamma}{T} d_j \right) - \left( 1 - \frac{\gamma}{T} \right) \cos \omega_{tj} \sin \omega_j \right) \right\}. \tag{3.4}
\]

Since \( 2(T + 1)^2 (1 + 2q_T d_j) \to (\pi^2 j^2 + \gamma^2) \) uniformly for \( j \in \{1, \ldots, k_T\} \) (see (2.9) in Davis and Dunsmuir (1996)), the absolute value of the sum in (3.4) truncated at \( k_T \) is

\[
\leq \left( \text{const} \right) (T + 1)^2 \sum_{j=1}^{k_T} (\pi^2 j^2 + \gamma^2)^{-1} \left( \frac{\pi^2 j^2}{(T + 1)^2} + \frac{\gamma}{T} + \frac{\pi j}{T + 1} \right)^2
\]

\[
\leq \left( \text{const} \right) (T + 1)^2 \sum_{j=1}^{k_T} (\pi^2 j^2 + \gamma^2)^{-1} \frac{j^2}{(T + 1)^2}
\]

\[
= O(k_T T^{-1}),
\]

where the value of \( \text{const} \) may change from line to line.

We now turn to the remaining part of the sum in (3.4) for values of \( j \) between \( k_T \) and \( T \). For \( j \geq k_T \),

\[
\frac{1 - d_j}{1 + 2q_T d_j} = \left( 1 + \frac{d_j \gamma^2 a(\gamma, T)}{2T^2(1 - d_j)} \right)^{-1} = 1 + O(k_T^{-2})
\]

and

\[
T^{-2}(1 - d_j)^{-1} = O(k_T^{-2}) = o(T^{-1}).
\]
Using these relations and the property that $\sin^2 \omega_j = 1 - d_j^2 = (1 - d_j)(1 + d_j)$, the sum in (3.4) for $j \geq k_T$ is then

\[
\sim \frac{2}{T+1} \sum_{j=k_T}^{T} (1 - d_j)^{-1} \left\{ (\sin \omega_{s_j}(1 - d_j + O(T^{-1})) - \cos \omega_{s_j} \sin \omega_j (1 + O(T^{-1}))) \\
- (\sin \omega_{t_j}(1 - d_j + O(T^{-1})) - \cos \omega_{t_j} \sin \omega_j (1 + O(T^{-1}))) \right\}
\]

\[
= \frac{2}{T+1} \sum_{j=k_T}^{T} \sin \omega_{s_j} \sin \omega_{t_j} (1 - d_j + O(T^{-1})) + \frac{2}{T+1} \sum_{j=k_T}^{T} \cos \omega_{s_j} \cos \omega_{t_j} (1 + d_j)(1 + O(T^{-1}))
\]

\[
- \frac{2}{T+1} \sum_{j=k_T}^{T} (\cos \omega_{s_j} \sin \omega_{t_j} - \cos \omega_{t_j} \sin \omega_{s_j}) \sin \omega_j (1 + T^{-1})(1 - d_j)^{-1}
\]

\[
= \frac{2}{T+1} \sum_{j=k_T}^{T} \left( \cos \omega_{(s-t)j} + d_j \cos \omega_{(s+t)j} \right) + \frac{2}{T+1} \sum_{j=k_T}^{T} \sin \omega_{(s+t)j} \sin \omega_j + O(k_T^{-2}T)
\]

\[
= \frac{2}{T+1} \sum_{j=k_T}^{T} \left( \cos \omega_{(s-t)j} + \cos \omega_{(s+t-1)j} \right) + O(k_T^{-2}T)
\]

\[
= \frac{2}{T+1} \sum_{j=1}^{T+1} \left( \cos \omega_{(s-t)j} + \cos \omega_{(s+t-1)j} \right) + O(k_T^{-2}T).
\]

Applying the formula,

\[
\sum_{j=1}^{T+1} \cos \omega_{kj} = \cos \left( \frac{\pi k}{2} \right) + \frac{\pi k}{2(T+1)} \sin \left( \frac{\pi k}{2} \right) / \left( 2 \sin \left( \frac{\pi k}{2(T+1)} \right) \right)
\]

\[
= \begin{cases} 
T + 1, & \text{if } k = 0 \text{ or } k \text{ even with } k = m(T + 1), \; m = \pm 1, \pm 2, \ldots, \\
-1, & \text{if } k \text{ is odd}, \\
0, & \text{otherwise},
\end{cases}
\]

(see Gradshteyn and Ryzhik (1980), p. 30) this last sum is $2 + O(k_T^{-2}T)$ for $s = t$ and $O(k_T^{-2}T)$ for $s \neq t$ uniformly for $s, t \in \{1, \ldots, T\}$. This combined with our analysis for the sum of the first $k_T$ terms comprising $r_{s,t}$ completes the proof of the lemma. \hfill \Box

**Lemma 3.3** If $r_{s,t}$ is the covariance function defined in Lemma 3.2, then

\[
(3.5) \quad T^{-2/\alpha} \sum_{s=1}^{T} \sum_{t=1}^{T} \epsilon_s \epsilon_t r_{s,t} = 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 + o_p(1) \xrightarrow{d} X_0^2,
\]

where $X_0^2$ is a positive $\alpha/2$-stable random variable with Laplace–Stieltjes transform

\[
E \exp \left\{ -\lambda X_0^2 \right\} = \exp \left\{ -\lambda^{\alpha/2}2^{\alpha} E|Z|^{\alpha} \right\}
\]

and $Z$ is a $N(0, 1)$ random variable.
Proof: We first show the convergence in distribution part of (3.5). If $Z_1, \ldots, Z_T$ are iid $N(0,1)$ random variables, independent of the $\{\epsilon_t\}$ sequence, then the Laplace–Stieltjes transform of $2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2$ is seen to be
\[
E \exp \left\{ -\lambda 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 \right\} = E \exp \left\{ i\lambda^{1/2}2T^{-1/\alpha} \sum_{t=1}^{T} \epsilon_t Z_t \right\} 
\]
\[
= E \exp \left\{ -\lambda^{\alpha/2}2\alpha T^{-1} \sum_{t=1}^{T} |Z_t|^\alpha \right\},
\]
which by the strong law of large numbers
\[
\rightarrow \exp \left\{ -\lambda^{\alpha/2}2\alpha \mathbb{E}|Z|^\alpha \right\} = E \exp \left\{ -\lambda X_0^2 \right\}.
\]

As for the remaining part of (3.5), we have
\[
T^{-2/\alpha} \sum_{s=1}^{T} \sum_{t=1}^{T} \epsilon_s \epsilon_t r_{s,t} + o_p(1) = 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 + 2T^{-2/\alpha} \sum_{1 \leq s < t \leq T} \epsilon_t \epsilon_s r_{s,t},
\]
so that it suffices to show the second term is $o_p(1)$. Using the symmetry of the distribution of the $\epsilon_t$, it follows that
\[
\text{Var} \left( T^{-2/\alpha} \sum_{1 \leq s < t \leq T} \epsilon_s \epsilon_t r_{s,t} \right| |\epsilon_t|, t = 1, \ldots, T) 
\leq T^{-4/\alpha} \sum_{1 \leq s < t \leq T} \epsilon_t^2 \epsilon_s^2 r_{s,t}^2 
\leq \text{(const)}(kT^{-1} + Tk_T^2)^2 \left( T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 \right)^2
\]
\[= o_p(1), \]
and hence the second sum in (3.6) must be $o_p(1)$ as claimed.

Lemma 3.4 We have
\[
T^{-2/\alpha} \sum_{t=1}^{T} U_{t,T}^2 = 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 + o_p(1)
\]
and
\[
T^{-1} \sum_{t=1}^{T} \tilde{U}_{t,T}^2 \overset{P}{\to} 1.
\]

Proof: Write
\[
U_{k,T} = (2/(T + 1))^{1/2}(1 + 2q_k d_k)^{-1/2} \sum_{s=1}^{T} \epsilon_s \left( \sin \omega_{sk} - \left( 1 - \frac{\gamma}{T} \right) \sin \omega_{(s+1)k} \right) - A_{k,T},
\]
where
\[
A_{k,T} = (2/(T + 1))^{1/2}(1 + 2q_k d_k)^{-1/2} \left( 1 - \frac{\gamma}{T} \right) \epsilon_0 \sin \omega_k.
\]
(Note for fixed k, $A_{k,T} = o_p(1)$.) Now, by Lemma 3.3,

$$T^{-2/\alpha} \sum_{k=1}^{T}(U_{k,T} + A_{k,T})^2 = T^{-2/\alpha} \sum_{s=1}^{T} \sum_{t=1}^{T} \epsilon_s \epsilon_t r_{s,t}$$

$$= 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 + o_p(1).$$

It can easily be seen from (3.1) that $\sum_{k=1}^{T} T^{-2}(1 + 2q_k d_k)^{-1} = O(1)$, and hence

$$T^{-2/\alpha} \sum_{k=1}^{T} A_{k,T}^2 \leq \frac{2}{T^{1+2/\alpha}} \sum_{k=1}^{T} (1 + 2q_k d_k)^{-1} \epsilon_0^2 = T^{1-2/\alpha} O_p(1) = o_p(1)$$

from which (3.7) now follows.

Finally, from Lemma 3.3 and (3.7), we obtain

$$T^{-1} \sum_{1}^{T} \tilde{U}_{i,T}^2 \sim \frac{T^{-2/\alpha} \sum_{i=1}^{T} U_{i,T}^2}{2T^{-2/\alpha} \sum_{i=1}^{T} \epsilon_i^2} \xrightarrow{P} 1.$$  

\[ \Box \]

**Lemma 3.5** For any fixed positive integer k,

(i) $(2T^{-2/\alpha} \sum_{t=0}^{T} \epsilon_t^2, T^{1/2-1/\alpha} U_{1,T}, \ldots, T^{1/2-1/\alpha} U_{k,T}) \xrightarrow{d} (X_0^2, X_1, \ldots, X_k)$, where $(X_0^2, X_1, \ldots, X_k)'$ has a mixed stable distribution with joint Laplace–Fourier transform given by

$$E \exp \left\{ -\lambda_0 X_0^2 + i \sum_{j=1}^{k} \lambda_j X_j \right\} = E \exp \left\{ -\int_{0}^{1} \left[ \lambda_0^{1/2} 2Z + 2 \sum_{j=1}^{k} \frac{\lambda_j - \pi j \cos(\pi x j) + \gamma \sin(\pi x j)}{(\pi j^2 + \gamma^2)^{1/2}} \right] dx \right\}$$

with $Z \sim N(0, 1)$,

(ii) $(\tilde{U}_{1,T}, \ldots, \tilde{U}_{k,T}) \xrightarrow{d} (\tilde{X}_1, \ldots, \tilde{X}_k)$ where $\tilde{X}_j = X_j/X_0$, and

(iii) $E \tilde{X}_j^2 = 1$ and $E(\tilde{X}_j^2 \tilde{X}_k^2) = 1$ for $j \neq k$. Moreover, the $\tilde{X}_j$ have uniformly exponentially decreasing tails, i.e., $\sup_j P(|\tilde{X}_j| > x) \leq c \exp\{-c'x\}$ for $x > 0$ and appropriate constants $c, c'$. In particular, the $\tilde{X}_j$ have uniformly bounded moments of all orders, i.e., for all $p > 0$, $\sup_j E|\tilde{X}_j|^p < \infty$.

**Proof:** (i) Since for each fixed $j$, $2(2T + 1)^2(1 + 2q_k d_k) \rightarrow \pi^2 j^2 + \gamma^2$ we have, using (3.12) and the expansion, $\sin \omega_{s,j} - \sin \omega_{s+1,j} = -(\pi j/T) \cos \omega_{s+1,j} \times o(T^{-1})$, that

$$T^{1/2-1/\alpha} U_{j,T} = T^{1-1/\alpha} 2(\pi j^2 + \gamma^2)^{-1/2} \sum_{s=1}^{T} \epsilon_s \left( \sin \omega_{s,j} - \left( 1 - \frac{\gamma}{T} \right) \sin \omega_{s+1,j} \right) + o_p(1)$$

$$= T^{1-1/\alpha} 2(\pi j^2 + \gamma^2)^{-1/2} \sum_{s=1}^{T} \epsilon_s \left( -\frac{\pi j}{T} \cos \omega_{s+1,j} + \frac{\gamma}{T} \sin \omega_{s+1,j} \right) + o_p(1).$$
If $Z_1, \ldots, Z_T$ are iid $N(0,1)$ random variables, independent of the $\epsilon_t$ sequence, then the joint Laplace–Fourier transform of the vector in (i) is given by

$$E \exp \left\{ -\lambda_0 2T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 + i \sum_{j=1}^{k} \lambda_j T^{1/2 - 2/\alpha} U_{j, T} \right\}$$

$$= E \exp \left\{ iT^{-1/\alpha} \sum_{s=1}^{T} \epsilon_s \left( \lambda_{0}^{1/2} Z_s + 2T \sum_{j=1}^{k} \lambda_j \left( -\frac{\pi j}{T} \cos \omega_{(s+1)j} + \frac{\gamma}{T} \sin \omega_{(s+1)j} \right) \right) + o_p(1) \right\}$$

and, after taking expectations with respect to the $\epsilon_t$ sequence first, is equal to

$$E \exp \left\{ -T^{-1} \sum_{s=1}^{T} \left| \lambda_{0}^{1/2} Z_s + 2T \sum_{j=1}^{k} \lambda_j (\pi^2 j^2 + \gamma^2)^{-1/2} (-\pi j \cos \omega_{(s+1)j} + \gamma \sin \omega_{(s+1)j}) \right|^\alpha + o_p(1) \right\}.$$

Now the exponent inside the expectation (excluding the $o_p(1)$ term) has mean

$$-T^{-1} \sum_{s=1}^{T} E \left| \lambda_{0}^{1/2} Z_s + 2T \sum_{j=1}^{k} \lambda_j (\pi^2 j^2 + \gamma^2)^{-1/2} (-\pi j \cos \omega_{(s+1)j} + \gamma \sin \omega_{(s+1)j}) \right|^\alpha$$

$$\rightarrow - \int_{0}^{1} E \left| \lambda_{0}^{1/2} Z + 2 \sum_{j=1}^{k} \lambda_j \frac{-\pi j \cos(\pi xj) + \gamma \sin(\pi xj)}{(\pi^2 j^2 + \gamma^2)^{1/2}} \right|^\alpha dx,$$

and variance converging to 0. Consequently, the exponent must converge to its mean in probability from which (i) now follows.

(ii) This part is immediate by the continuous mapping theorem and the conclusion of part (i).

(iii) We show that $\sup_{k,T} P(\bar{U}_{k,T}^2 > x) \leq c \exp\{-c'x\}$ for appropriate constants $c, c'$. First note that we can write

$$\bar{U}_{k,T} = \frac{\sum_{s=0}^{T} \epsilon_s a_{s,T}}{\left(\sum_{s=0}^{T} \epsilon_s^2\right)^{1/2}},$$

where $a_{s,T}$ are constants that are bounded in absolute value, uniformly in $s$ and $T$. Conditionally upon $|\epsilon_s|$, $s = 0, \ldots, T$, $\bar{U}_{k,T}^2$ is a Rademacher quadratic form with uniformly (in $k$ and $T$) bounded variance. It follows for example from Pisier and Zinn (1977), p. 292, that

$$\sup_{k,T} E P \left( \bar{U}_{k,T}^2 > x \mid |\epsilon_s|, s = 0, \ldots, T \right) \leq c \exp\{-c'x\},$$

where $c, c'$ are positive constants. It follows now, that all power moments of $\bar{U}_{k,T}$ are uniformly bounded and that the $\bar{X}_k^2$ have exponential tails. Since $\bar{U}_{k,T} \overset{d}{=} \bar{X}_k^2$ and $E \bar{U}_{k,T}^2 = 1$, we may conclude that $E \bar{X}_k^2 = 1$. The proof of the uncorrelatedness of the $\bar{X}_k^2$ is given in Proposition 5.2.

□
4 Accuracy of the Asymptotic Distribution

The accuracy of the asymptotic distribution derived in Theorem 2.1 will be evaluated in this section by comparison with the finite sample distribution of the \( \hat{\theta}_{LM} \) estimated by simulation. Replicates of \( \hat{\beta}_\gamma \) in Corollary 2.2, (a) are easy to compute from realizations of the sample for \( Y_\gamma(\beta) \) (see Remark 2.5). The procedure we have adopted to generate replicates of \( \hat{\beta}_\gamma \) is as follows:

**Step 1.** For fixed large integers \( K \) and \( N \), the random variables, \( X_0, X_1, \ldots, X_N \) were simulated using the approximation to the integrals in (2.6) and (2.7) given by

\[
X_t = 2 \sum_{s=1}^{K} \frac{-\pi t \cos(\pi ts/K) + \gamma \sin(\pi ts/K)}{(\pi^2 t^2 + \gamma^2)^{1/2}} Z_s, \quad t = 1, 2, \ldots, N,
\]

and

\[
X_0^2 = 2 \sum_{s=1}^{K} Z_s^2,
\]

where \( Z_1, \ldots, Z_K \) are iid SoS random variables.

**Step 2.** The infinite series for \( Y_\gamma \) was truncated at \( N \).

**Step 3.** For the truncated series, which we shall continue to call \( Y_\gamma \), we computed \( Y_\gamma(0) \). If \( Y_\gamma(0) < 0 \), then the replicate of \( \hat{\beta}_\gamma \) was set to 0.

**Step 3.** If \( Y_\gamma(0) > 0 \), then the replicate of \( \hat{\beta}_\gamma \) was defined as the smallest non-negative zero of \( Y_\gamma(\beta) \).

In all of simulations we took \( K = N = 1000 \). The results were not appreciably different with larger values of \( K \) and \( N \). In Step 3, we used the IMSL root finder DZREAL to compute \( \hat{\beta}_\gamma \). The smoothness of the sample paths of \( Y_\gamma \) makes it relatively straightforward to locate the first zero-crossing. All of the limit results reported below are based on 10,000 replicates of \( \hat{\beta}_\gamma \).

In order to compare the limit distribution with the finite sample distribution of the LM estimator, it was necessary to generate replicates of \( \hat{\theta}_{LM} \). The estimate \( \hat{\theta}_{LM} \) was computed by evaluating the reduced likelihood (computed using the innovations algorithm as described in Brockwell and Davis (1991)) at \( \theta = 1, 1 - .001, 1 - .002 \), etc. until a local maximum was achieved. Results reported below are also based on 10,000 replicates of the \( \hat{\theta}_{LM} \).

Figure 1 compares the sampling distribution of \( T(\hat{\theta}_{LM} - 1) \) with the distribution of the limit random variable \( -\beta_0 \) when \( \theta_0 = 1 \) (\( \gamma = 0 \)) for sample sizes \( T \in \{25, 50\} \) and \( \alpha = 1 \). These distributions are only plotted for \( x < 0 \) since they all take the value 1 at \( x = 0 \). The limit distribution provides a remarkably good approximation for sample sizes as small as 25 and 10 (not shown) and is virtually exact for \( T = 100 \). As expected from this figure, the lower quantiles of the sampling distribution of the LM estimator and the limit approximation are in good agreement. The
Figure 1: Comparison of sampling cdf's with limit cdf ($\theta_0 = 1.0, \alpha = 1.0$).

.05 and .1 quantiles of the distribution of $T(\hat{\theta}_{LM} - 1)$ for sample sizes $T = 10, 25, 50, 100$ and the corresponding limit distribution are displayed in Table 4.1 for $\alpha = .75, 1.0, 1.5, 2.0$. The tabulated values for the case $\alpha = 2.0$ are taken from Davis and Dunsmuir (1996).

In Figure 2, the limit distribution of $T(\hat{\theta}_{LM} - 1)$ is plotted for the 4 values of $\alpha = .75, 1.0, 1.5, 2.0$. As seen from this figure and Table 4.1 the limit distribution does not vary a great deal with $\alpha$. On the other hand, the pile-up probabilities, labelled as P-U in Table 4.1, corresponding to $P(\hat{\theta}_{LM} = 1)$ increase gradually with $\alpha$. Notice that there is good agreement between the sample and limiting values of the pile up and that the pile-up effect increases with increasing $\alpha$.

Table 4.1 Quantiles of $T(\hat{\theta}_{LM} - 1)$ and $-\bar{\beta}_0$ together with the pile-up (P-U) probabilities corresponding to $P(\hat{\theta}_{LM} = 1)$.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\alpha = .75$</th>
<th></th>
<th>$\alpha = 1.0$</th>
<th></th>
<th>$\alpha = 1.5$</th>
<th></th>
<th>$\alpha = 2.0$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.05</td>
<td>.10</td>
<td>P-U</td>
<td>.05</td>
<td>P-U</td>
<td>.05</td>
<td>P-U</td>
<td>.05</td>
</tr>
<tr>
<td>10</td>
<td>-5.63</td>
<td>-5.05</td>
<td>.597</td>
<td>-7.76</td>
<td>-4.87</td>
<td>.602</td>
<td>-6.66</td>
<td>-4.50</td>
</tr>
<tr>
<td>25</td>
<td>-6.80</td>
<td>-4.02</td>
<td>.594</td>
<td>-6.17</td>
<td>-4.07</td>
<td>.606</td>
<td>-6.20</td>
<td>-4.50</td>
</tr>
<tr>
<td>50</td>
<td>-5.20</td>
<td>-3.90</td>
<td>.605</td>
<td>-5.50</td>
<td>-4.10</td>
<td>.617</td>
<td>-6.25</td>
<td>-4.50</td>
</tr>
<tr>
<td>100</td>
<td>-5.20</td>
<td>-3.80</td>
<td>.615</td>
<td>-5.50</td>
<td>-4.20</td>
<td>.616</td>
<td>-5.80</td>
<td>-4.30</td>
</tr>
<tr>
<td>Limit</td>
<td>-5.13</td>
<td>-3.89</td>
<td>.607</td>
<td>-5.60</td>
<td>-4.20</td>
<td>.617</td>
<td>-5.93</td>
<td>-4.43</td>
</tr>
</tbody>
</table>
In Figure 3, the sampling distribution of $T(\hat{\theta}_{LM} - 1)$ with $T = 50$ is plotted together with the distribution of the limit random variable $-\beta_\gamma$ for the cases $\alpha = 1.0$, and $\theta_0 = .95, .9, .8, .7$ (i.e. $\gamma = 2.5, 5, 10, 15$), respectively. As is clear from these graphs, the approximation based on the empirical distribution of $-\bar{\beta}_\gamma$ with $\gamma = (1 - \theta_0)T$ is very accurate. This is reaffirmed in Table 4.2 which gives a comparison of the quantiles and the pile up at 0 for the cdf's plotted in Figures 3 and 4. Note that the pile up remains for values of $\theta_0$ as small as .7 ($\gamma = 15$) and these pile-up probabilities for $T = 50$ are well approximated by their limiting counterparts.

The quantiles reported in the bottom row of Table 4.1 can be used for calculating the cutoff value in testing $H_0 : \theta_0 = 1$ vs. $H_1 : \theta_0 < 1$. The null hypothesis is rejected at level $\delta$ if $\hat{\theta}_{LM} < 1 + b_\delta/T$ where $b_\delta$ is the $\delta$-quantile of $-\beta_0$. For example, if $T = 50$ and $\alpha = 1$, then $H_0$ is rejected at level .05 if $\hat{\theta}_{LM} < 1 - 5.60/50 = .888$ and at level .10 if $\hat{\theta}_{LM} < 1 - 4.20/50 = .916$. The limiting power of the test for the sequence of local alternatives $\theta_T = 1 - \gamma/T$ is given by $P(\bar{\beta}_\gamma > -b_\delta)$.

Alternatively, the likelihood ratio test can be used to test $H_0 : \theta_0 = 1$ vs. $H_1 : \theta_0 < 1$. First note that the logarithm of the likelihood ratio statistic is in fact $L_T(\bar{\beta}_{LM}) - L_T(0)$, which by applying the continuous mapping theorem to Theorem 2.1, converges in distribution to $Z_\gamma(\bar{\beta}_\gamma)$. In particular, under $H_0$, the asymptotic cutoff value for the likelihood ratio is the $1 - \delta$ quantile of $\bar{\beta}_0$. These values are easy to tabulate via simulation. For further details about the likelihood ratio test and its superior performance to the test based on the MLE in the Gaussian case, see Davis et al (1995).
Figure 3: Comparison of sample and limit cdf's for $\theta = .7, .8, .9, .95$ ($\gamma = 15, 10, 5, 2.5$) and $\alpha = 1.0$.

Table 4.2 Quantiles of $T(\hat{\theta}_{LM} - 1)$ for $T = 50$ for values of $\theta_0 = .7, .8, .9, .95$ ($\gamma = 15, 10, 5, 2.5$) and $\alpha = 1.0$ and the corresponding limiting values. The last column is the probability of a pile up.

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$T = 50$</th>
<th>Limit</th>
<th>.05</th>
<th>.1</th>
<th>$P_\gamma[\hat{\theta} = 1]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.7</td>
<td>-21.75</td>
<td>-21.59</td>
<td>-16.15</td>
<td>-14.00</td>
<td>.113</td>
</tr>
<tr>
<td>.8</td>
<td>-15.72</td>
<td>-14.06</td>
<td>-10.00</td>
<td>-8.25</td>
<td>.276</td>
</tr>
<tr>
<td>.9</td>
<td>-9.53</td>
<td>-8.22</td>
<td>-7.15</td>
<td>-5.60</td>
<td>.465</td>
</tr>
<tr>
<td>.95</td>
<td>-6.67</td>
<td>-5.47</td>
<td>.464</td>
<td>.464</td>
<td>.464</td>
</tr>
</tbody>
</table>
5 Appendix

Denote

\[
\phi_k(x) = (\pi^2 k^2 + \gamma^2)^{-1/2}(-\pi k \cos(kx) + \gamma \sin(kx)).
\]

First, we give a joint a.s. representation of the \(X_k, k \geq 0\). In what follows, we assume that \(S\alpha S\) Lévy motion \(M\) has a càdlàg Lévy–Itô representation; see Itô (1969), Resnick (1986) and Samorodnitsky and Taqqu (1994), Proposition 3.11.1. Defining \(\Delta M(x) = M(x) - M(x-),\) we have

\[
M(x) = \lim_{\delta \downarrow 0} \sum_{|\Delta M(x)| > \delta} \Delta M(x)
= \lim_{\delta \downarrow 0} \sum_{\{N|N_i|^{-1/\alpha}|>\delta, U_i \leq x\}} N_i \Gamma_i^{-1/\alpha}, \quad x \in [0, 1],
\]

where \(\{N_i\}\) are iid \(N(0,1)\) random variables, \(\{\Gamma_i\}\) are the points of a homogeneous Poisson process on \((0,\infty)\) and \(\{U_i\}\) are iid uniform random variables on \((0,1)\). Moreover, the sequences \(\{N_i\}, \{\Gamma_i\}\) and \(\{U_i\}\) are independent. The convergence of the sums above to \(M\) is uniform on \([0,1]\). Recall that we assume that \(M(1) \overset{d}{=} \epsilon_1\).

**Proposition 5.1** Assume \(M\) has Lévy-Itô representation. Then the quantities

\[
X_k = 2 \int_0^1 \phi_k(\pi x) dM(x)
\]

can be interpreted as pathwise Riemann-Stieltjes integrals with representation

\[
(5.1) \quad 2 \sum_{i=1}^{\infty} \phi_k(\pi U_i) N_i \Gamma_i^{-1/\alpha},
\]

and \(X_k^2\) is the quadratic variation of \(M\) on \([0,1]\) with pathwise representation

\[
(5.2) \quad 2 \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha} N_i^2.
\]

**Proof:** We know from Lemma 3.5 and its proof that \(X_k^2\) is the weak limit of \(2T^{-2/\alpha} \sum_{i=1}^{T} \epsilon_i^2\), and the \(X_k\) are the weak limits of \(2T^{-1/\alpha} \sum_{i=1}^{T} \epsilon_i \phi_k(\omega_i)\). A possible representation of the random variables \(T^{-1/\alpha} \epsilon_i\) is given by \(M(t/T) - M((t-1)/T), t = 1, \ldots, T\) for every fixed \(T\), and from now on we will assume this representation. The stable process \(M\) has \(p\)-variation of order \(p > \alpha\); see Fristedt and Taylor (1973). Hence the a.s. limit of \(T^{-2/\alpha} \sum_{i=1}^{T} \epsilon_i^2\) exists and coincides with the quadratic variation of \(M\) on \([0,1]\). The representation (5.2) is a consequence of the Lévy-Itô representation for the jumps of \(M\).

For every fixed \(k\), the quantities \(2T^{-1/\alpha} \sum_{i=1}^{T} \epsilon_i \phi_k(\omega_i)\) are Riemann-Stieltjes sums whose limit exists and coincides with the value of the Riemann-Stieltjes integral \(2 \int_0^1 \phi_k(\pi x) dM(x)\). This follows from a classical paper by Young (1936) on Riemann-Stieltjes integration and from its generalizations.
due to Dudley (1992) and Dudley and Norvaiša (1997). Using the Lévy-Itô representation of $M$, it follows that the Riemann-Stieltjes integral $\int_0^1 \phi_k(\pi x)dM(x)$ is the pathwise limit of

$$
\lim_{\delta \downarrow 0} \int_0^1 \phi_k(\pi x) d\left( \sum_{i=1}^{N(\delta)} N_i \Gamma_i^{-1/\alpha} I_{\{U_i \leq \epsilon\}} \right) = \lim_{\delta \downarrow 0} \sum_{i=1}^{N(\delta)} \phi_k(\pi U_i) N_i \Gamma_i^{-1/\alpha},
$$

where $N(\delta) = \#\{i : |\Gamma_i^{-1/\alpha} N_i| > \delta\} \to \infty$. Here we made use of the fact that the Lévy-Itô series converges uniformly on compact intervals and of the theorem on term by term integration in Young (1936), p. 269. This concludes the proof of the lemma.

The following is an immediate consequence of Proposition 5.1. First recall from (2.8) the definition of the process $Y_\gamma$.

**Corollary 5.2** For every $\beta \geq 0$, the distribution of $Y_\gamma(\beta)$ is continuous. In particular, $P(Y_\gamma(\beta) = 0) = 0$.

**Proof:** Assume the representations (5.1) and (5.2) for all $X_k$, $k \geq 0$. The random variable $Y_\gamma(\beta)$, given $N_2, N_3, \ldots, \{\Gamma_i\}$ and $\{U_i\}$, is a rational function of $N_1$ and hence has a continuous distribution.

**Proposition 5.3** The random variables $\tilde{X}_k^2$ are uncorrelated.

**Proof:** We exploit two facts: the $\tilde{X}_k^2$ are the weak limits of $\tilde{U}_{k,T}^2$ as $T \to \infty$, and they are uniformly integrable; see Lemma 3.5 and its proof. Recall the definition of $\phi_k$ and write

$$
\tilde{e}_t = \epsilon_t / \left( \sum_{s=1}^{T} \epsilon_s^2 \right)^{1/2}, \quad t = 1, \ldots, T.
$$

For every $T$, the $\tilde{e}_t$ are exchangeable and conditionally independent given their absolute values. Moreover,

$$
\left( T^{-4/\alpha} \sum_{t=1}^{T} \epsilon_t^4, T^{-2/\alpha} \sum_{t=1}^{T} \epsilon_t^2 \right) \overset{d}{\to} (Z_{\alpha/4}, Z_{\alpha/2}),
$$

where the limit consists of an $\alpha/4$-stable and an $\alpha/2$ stable random variable. It follows that $\sum_{t=1}^{T} \epsilon_t^4 \overset{d}{\to} Z_{\alpha/4}/Z_{\alpha/2}^2$ and since these quantities are bounded by 1, the means also converge. From these properties, we obtain for $t \neq s$,

$$
E\tilde{e}_t^2 = \frac{1}{T} E\sum_{t=1}^{T} \epsilon_t^4 \sim T^{-1} c_1,
$$

$$
E\tilde{e}_t^2 \epsilon_s^2 = \frac{1}{T(T-1)} \sum_{1 \leq t \neq s \leq T} E\tilde{e}_t^2 \epsilon_s^2 \sim T^{-2} (1 - c_1),
$$

19
where \( c_1 = E(Z_{\alpha_4}/Z_{\alpha_2}^2) \). Then for \( k \neq l \),

\[
4^{-1}E\tilde{X}_k^2\tilde{X}_l^2
\]

\[
= \lim_{T \to \infty} E \left( \sum_{t=1}^T \tilde{\varepsilon}_t^2 \phi_k^2(\omega_t) \right) \left( \sum_{t=1}^T \tilde{\varepsilon}_t^2 \phi_l^2(\omega_t) \right) + 2 \lim_{T \to \infty} \sum_{1 \leq t \neq s \leq T} E\tilde{\varepsilon}_t^2\tilde{\varepsilon}_s^2 \phi_k(\omega_t)\phi_k(\omega_s)\phi_l(\omega_t)\phi_l(\omega_s)
\]

\[
= \lim_{T \to \infty} \sum_{t=1}^T E\tilde{\varepsilon}_t^4 \phi_k^2(\omega_t)\phi_l^2(\omega_t) + \lim_{T \to \infty} \sum_{1 \leq t \neq s \leq T} E\tilde{\varepsilon}_t^2\tilde{\varepsilon}_s^2 \phi_k(\omega_t)\phi_l(\omega_t)\phi_l(\omega_s)
\]

\[
+ 2(1 - c_1) \lim_{T \to \infty} \left( T^{-1} \sum_{t=1}^T \phi_k(\omega_t)\phi_l(\omega_t) \right)^2
\]

\[
= c_1 \lim_{T \to \infty} T^{-1} \sum_{t=1}^T \phi_k^2(\omega_t)\phi_l^2(\omega_t) + (1 - c_1) \lim_{T \to \infty} \left( T^{-1} \sum_{t=1}^T \phi_k^2(\omega_t) \right) \left( T^{-1} \sum_{t=1}^T \phi_l^2(\omega_t) \right)
\]

\[
+ 2(1 - c_1) \left( \int_0^1 \phi_k(\pi x)\phi_l(\pi x) dx \right)^2
\]

\[
= J_1 + J_2 + J_3 .
\]

Notice that \( J_3 = 0 \) by orthogonality of \( \phi_k(\pi x) \) and \( \phi_l(\pi x) \). Moreover, it is not difficult to check that

\[
J_2 = (1 - c_1) \int_0^1 \phi_k^2(\pi x)dx \int_0^1 \phi_l^2(\pi x)dx = 4^{-1}(1 - c_1) .
\]

Straightforward calculation also yields

\[
J_1 = c_1 \int_0^1 \phi_k^2(\pi x)\phi_l^2(\pi x) dx = 4^{-1}c_1 ,
\]

which concludes the proof. \( \Box \)

In order to complete the proof of Theorem 2.1, we need the following proposition which is the analogue of Proposition A.1 of Davis and Dunsmuir (1996).

**Proposition 5.4** Let \( Z'_\gamma(\beta) = \beta Y'_\gamma(\beta)/2, \ Z''_\gamma(\beta) = (\beta Y''_\gamma(\beta) + Y'_\gamma(\beta))/2, \) where \( Y'_\gamma(\beta) \) is the process defined (2.8). Then \( (Z'_\gamma, Z''_\gamma) \in S \) a.s., i.e.,

\[
P[Z'_\gamma(\beta) = Z''_\gamma(\beta) = 0 \text{ for some } \beta \geq 0] = 0
\]

and

\[
P[Z'_\gamma(\beta) = 0, \ Z''_\gamma(\beta) < 0 \text{ for some } \beta \geq 0] = 1.
\]

**Proof:** The proof of (5.3) and (5.4) mirror the arguments given for (A.1) and (A.2) in Davis and Dunsmuir (1996). As noted in Corollary 5.2, \( 2Z''_\gamma(0) = Y'_\gamma(0) \neq 0 \) a.s., so that it suffices to
consider $\beta > 0$ only in both (5.3) and (5.4). We start with (5.3). For notational simplicity define $A(\phi) := Y_\gamma(\phi^{1/2})$, i.e.,

$$A(\phi) = -4 \sum_{k=1}^{\infty} \frac{1}{\pi^2 k^2 + \phi} + 4 \sum_{k=1}^{\infty} \frac{\pi^2 k^2 + \gamma^2}{(\pi^2 k^2 + \phi)^2} \tilde{X}_k^2$$

and

$$A'(\phi) = 4 \sum_{k=1}^{\infty} \frac{1}{(\pi^2 k^2 + \phi)^2} - 8 \sum_{k=1}^{\infty} \frac{\pi^2 k^2 + \gamma^2}{(\pi^2 k^2 + \phi)^3} \tilde{X}_k^2.$$

Using the a.s. representation given in Proposition 5.1, write

$$X_0^2 = N_1^2 F_0 + G_0$$

and

$$X_k = N_1 F_k + G_k,$$

where $F_k, G_k$ are functions of $U = \{U_i\}$, $\Gamma = \{\Gamma_i\}$, and $\{N_j, j \geq 2\}$. Let $C = C(F_k, G_k, k \geq 0)$ denote the set of $N_1$ such that $A(\phi) = A'(\phi) = 0$ for some $\phi \geq 0$. Note that if $N_1 \in C$, then there exists a $\phi = \phi(N_1)$ such that $A(\phi) = A'(\phi) = 0$. Solving the equation $X_0^2 A(\phi) = 0$ for $N_1$, we see that $N_1$ must satisfy the quadratic equation

$$N_1^2 B_1(\phi) + N_1 B_2(\phi) + B_3(\phi) = 0,$$

where the $B_i$, conditional on $\{F_k, G_k, k \geq 0\}$, are analytic functions of $\phi$. Let $A_2(\phi)$ be equal to $A'(\phi)$, where $N_1$ is substituted with either of the solutions to the above quadratic equation. It follows that $A_2$ is an analytic function of $\phi$ and is not identically 0. Consequently, the set of $\phi$ satisfying $A_2(\phi)$ is countable and does not depend on $N_1$. This implies that the set $\{\phi(N_1), N_1 \in C\}$ is countable and each $\phi$ in this set uniquely determines at most 2 possible values of $N_1$. This, in turn, implies that $C$ is countable and since $N_1$ has a continuous distribution and is independent of $\{F_k, G_k, k \geq 0\}$, we conclude that

$$P[N_1 \in C|F_k, G_k, k \geq 0] = 0 \text{ a.s.}$$

Since the left-hand side is equal to $P[Z'(\beta) = Z''(\beta) = 0 \text{ for some } \beta > 0 | F_k, G_k, k \geq 0]$, (5.3) now follows.

The proof of (5.4) is completely analogous to (A.2) in Davis and Dunsmuir (1996): the uniform boundedness of $E\tilde{X}_k^4$ (which follows from Lemma 3.5.(iii)) and the orthogonality of the $\tilde{X}_k^2$ (which is guaranteed by Proposition 5.4) are all that is required.

\[\Box\]

**Acknowledgment.** This paper was written when the second named author was visiting the Department of Statistics at Colorado State University in Fort Collins. He would like to thank his colleagues for their excellent hospitality.
REFERENCES


