June 1992

Inference for MA (1) processes with a root on or near the unit circle

Richard A. Davis

William T. M. Dunsmuir

BOND UNIVERSITY
School of Business

DISCUSSION PAPERS

"Inference for MA (1) Processes with a Root On or Near the Unit Circle"

by

Richard A. Davis & William T. M. Dunsmuir

DISCUSSION PAPER NO 36

JUNE 1992

University Drive,
Gold Coast,QLD,4229
AUSTRALIA

Inference for MA(1) Processes with a Root On or Near the Unit Circle*

Richard A. Davis
Colorado State University

and

William T. M. Dunsmuir
Bond University

Abstract

This paper considers maximum likelihood estimation for Gaussian MA(1) processes when the moving average parameter is on or near the unit circle. A unified framework for the derivation of the asymptotic distribution of the MLE of the moving average parameter is given. Of practical significance is the fact that the asymptotic distribution is surprisingly accurate even for small sample sizes and for values of the moving average parameter considerably far from the unit circle. In the latter case, where the normal limit approximation is thought to be reasonably good, the approximation based on our limiting results is often more accurate than the normal approximation. The theoretical results are applied to inference problems of hypothesis testing and confidence interval construction and is demonstrated by simulation to have superior properties than procedures based on a normal distribution approximation.

*This research supported in part by NSF DMS Grant No. 9100392.
AMS 1980 Subject Classification: 62M10; 62M09; 62M07
Key Words and Phrases. Moving average process, unit roots, noninvertible moving averages, maximum likelihood

1 Introduction

Inference for stationary time series models is usually based on asymptotic theory and, yet, for even the simplest of models the asymptotic theory is quite inadequate. For example in the MA(1), the moving average of order one,

$$Y_t = \epsilon_t - \theta_0 \epsilon_{t-1}, \tag{1.1}$$

where $\{\epsilon_t\} \sim \text{NID}(0, \sigma^2)$ ($\{\epsilon_t\}$ is a sequence of independent $N(0, \sigma^2)$ random variables) and $|\theta_0| < 1$, the asymptotic normal distribution $N(0, (1 - {\theta_0}^2)/T)$ (T is the sample size) for the maximum likelihood estimator is particularly inaccurate for values of the true parameter close to ± 1.

This "near invertible" case arises quite frequently in time series modelling especially when the original time series observed has been differenced at lag one to achieve stationarity. In particular when θ_0 is near 1 the asymptotic normal distribution is a poor approximation to the actual distribution of the MLE as is also clearly shown in Figures 2–5 below.

A series of authors have noted that when θ_0 is close to 1 the MLE has a high probability of being equal to 1. See, for example, Cooper and Thompson(1977), Ansley and Newbold(1980), and Dunsmuir(1981). These empirical observations prompted a number of authors to investigate alternative approximations to the exact distribution of the MLE in such cases. For example Cryer and Ledolter (1981), Anderson and Takemura(1986), Tanaka and Satchell(1989), Pesaran(1983) have all contributed to our understanding of the distributional properties of the MLE. In particular these authors have shown, in line with the earlier empirical evidence, that the MLE has a substantial probability of being exactly on the boundary of the invertibility region for both the case when θ_0, the true value, is equal to 1 and the case where it is strictly less than 1.

We will return to a more complete review of this literature shortly. Before doing so it is worth pointing out that the major contribution of this paper is to provide a unified theoretical framework for the derivation of the limit distribution of the MLE for cases when $\theta_0 = 1$ and when θ_0 is less than but close to 1. The main theorem presented below allows us to derive such a unified asymptotic result. Perhaps of more practical significance is that the asymptotic distribution turns out to be surprisingly accurate even for small sample sizes and for values of $|\theta_0|$ as small as .6 - see Section 3.

Before going into further details of the past theory and the present results it is worth exploring the reason why there is a need to have good approximations to the finite sample distributions of the MLE for the MA(1) when θ_0 is close to 1. In practice time series are often differenced one or more times at lag one in order to produce a series which is stationary. This has become common especially due to the influence of Box and Jenkins (1976). Sometimes overdifferencing occurs. In such cases, the usual asymptotic normal distribution results do not hold but, it is important to be able to test if the resulting series has a moving average component with a unit root. This is put forward as motivation by Tanaka and Satchell(1989) who refer to Plosser and Schwert(1977) and cite Christiano (1987) who "stresses the need to develop an asymptotic theory for these problems". Chan and Tsay (1992) also use the possibility of over-differencing as motivation for their procedures.

Often after differencing by the correct amount the moving average estimate is at or near to 1 indicating strong evidence that the process observed may have been generated by an IMA(0,1,1) model. In such cases a confidence interval or hypothesis test for θ_0 would be useful, yet inferences based on a poorly fitting limit distribution are inaccurate.

Another case in which there is a need to have an accurate basis for inference is when the data are conceived of as being generated by a so called structural model such as discussed in Harvey(1989). The simplest form of the structural model is the random walk signal plus noise model in which $Z_t = \mu_t + E_t$ and $\mu_t = \mu_{t-1} + V_t$. Here $\{E_t\}$ is an i.i.d. sequence independent of the sequence $\{V_t\}$ which is also i.i.d. If these random variables are normally distributed then this model is equivalent to the IMA(0,1,1) model. Of interest in such models is the null hypothesis that the signal μ_t is constant. This is equivalent to testing the hypothesis that $\theta_0 = 1$ in the differenced series $Y_t = Z_t - Z_{t-1}$. Examples of where such a hypothesis might occur are in the development of control charts for an industrial process which could have an underlying process level drifting according to the random walk but for which direct measurement is not available; rather the process level is observed subject to noise. Of course if the null hypothesis is rejected the practitioner would conclude that the underlying process is not stable and be led to developing appropriate control action.

We turn now to a more detailed review of available finite sample and asymptotic theory for the MA(1) case. We restrict our attention to this case for completeness and brevity, although extension of our results to higher order moving average or even mixed autoregressive-moving average models should follow along similar lines. The main theoretical contributions to our understanding of the finite sample distribution of the MLE appear to have commenced with Kang(1975) who provided the earliest indication of the tendency for both the MLE and the least squares estimates to occur on the boundary of the invertibility region. Dunsmuir(1981) points out that the concentrated likelihood always has a local minimum or maximum at $|\theta| = 1$ and reported empirical evidence that if a local maximum occurs at these boundary values then it was the global maximum. Cryer and Ledolter(1981) give a more detailed theoretical explanation for the tendency of the MLE to "pile up" at $|\theta| = 1$ and derive approximations to $P[\hat{\theta} = 1]$ for a general sample size. Davidson(1981a,b) has also investigated finite sample probabilities. The case when the true parameter $\theta_0 = 1$ has also been investigated by Sargan and Bhargava (1983) and by Pesaran(1983). Anderson and Takemura (1986) "develop, organize and generalize these results" and provide "a rigorous derivation of the limiting probabilities that the likelihood function attains a local maximum" at $\hat{\theta} = 1$ for values of $\theta_0 = 1$ and values where $\theta_0 < 1$. They also generalise their results to higher order mixed autoregressive-moving average models.

All of these authors investigate the discrete component of the finite sample distribution of the MLE or asymptotic approximations to it but none of them provide approximations to the continuous part corresponding to $\hat{\theta} < 1$. Tanaka and Satchell(1989) investigate both components of the distribution of the MLE including the continuous part but use an approximation that is inadequate. As we show later our avoidance of the approximation suggested by them is both feasible computationally and leads to considerably more accurate approximate distributions.

None of these previous authors have developed a complete asymptotic theory that is useful for values of θ_0 close to the unit circle as well as for values where $\theta_0 = 1$. Development of a unified theory covering all values of θ_0 and which gives an approximate distribution to both components of the distribution of $\hat{\theta}$ is the major contribution of this paper.

In Section 2 we establish the limiting behaviour of the MLE under a sequence of local values of θ_0 which converge to 1 at rate $1/T$ where T is the number of observations in the time series. This asymptotic theory provides a unified limit law for the discrete component and the absolutely continuous component of the actual finite sample distribution. The accuracy of this approximation is examined in detail for a range of values of θ_0 close to unity and for a range of sample sizes in Section 3. There the standard asymptotic normal distribution is also evaluated and shown, even for values of θ_0 as small as .6 and $T = 50$, to be a poor approximation compared to that afforded by the results of this paper.

Section 4 is devoted to illustrating the asymptotic results on problems of hypothesis testing. In these applications the theoretical results are used for calculating the power of the proposed test as well as for size. Section 5 is concerned with application of the theory to confidence interval construction.

2 Asymptotic theory

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

$$Y_t = \epsilon_t - \theta\epsilon_{t-1}, \tag{2.1}$$

where $\{\epsilon_t\} \sim \text{NID}(0, \sigma^2)$ and $|\theta| \leq 1$. Since we are interested in inference about θ when θ is at or near 1, we use the parameterization $\theta = \theta_T = 1 - \beta/T$ where $\beta \geq 0$ and T is the sample size. Inference about β and hence θ will be based on the observations $Y_1, \ldots, Y_T$ which come from model (2.1) with true parameter $\theta_0 = 1 - \gamma/T$, where $\gamma \geq 0$. In this section, we derive an expression for the concentrated likelihood as a function of β and show that its first and second derivatives converge in distribution to stochastic processes indexed by β. From this result, the limit random variable of the maximum likelihood estimator of β can be defined in terms of the two limit stochastic processes which, by our parameterization, is also the limit random variable of $T(1 - \hat{\theta})$ where $\hat{\theta} = (1 - \hat{\beta}/T)$ is the maximum likelihood estimator of θ.

As a function of the lag 1 correlation $\rho = \rho(\theta) = -\theta/(1+\theta^2)$, the concentrated likelihood is given by (see equation (10) of Anderson and Takemura (1986)),

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

where $\mathbf{Y}' = (Y_1, \ldots, Y_T)$ is the data vector and G is the covariance matrix of $\mathbf{Y}/(\mathrm{var}(Y_1))^{1/2}$. The concentrated likelihood as a function of β is then given by

$$L_T(\beta) = M(\rho(1-\beta/T)) \tag{2.3}$$

with

$$\begin{aligned} L_T'(\beta) &= \left(\frac{dM}{d\rho}\right)\left(\frac{d\rho}{d\theta}\right)\left(\frac{d\theta}{dT}\right) \\ &= \left(-\frac{dM}{d\rho}\right)\left(\frac{2\beta/T - \beta^2/T^2}{(2(1-\beta/T)+\beta^2/T^2)^2}\right)\left(\frac{-1}{T}\right) \\ &= \left(\frac{\beta}{2T^2}\right)(1-\beta/(2T))a^2(\beta,T)\left(\frac{dM}{d\rho}\right) \end{aligned} \tag{2.4}$$

where

$$a(\beta,T) = (1 - \beta/T + \beta^2/(2T^2))^{-1}.$$

Using equation (20) in Anderson and Takemura (1986), we have under the true $\theta_0 = 1-\gamma/T$, that

$$\frac{dM}{d\rho} \overset{d}{=} -\sum_{t=1}^{T}\frac{2d_t}{1+2p_T d_t} + \frac{T}{\sum_{t=1}^{T}\frac{(1+2q_T d_t)X_t^2}{1+2p_T d_t}}\sum_{t=1}^{T}\frac{2d_t(1+2q_T d_t)X_t^2}{(1+2p_T d_t)^2} \tag{2.5}$$

where $\{X_t\} \sim \mathrm{NID}(0,1)$, $d_t = \cos\frac{\pi t}{T+1}$,

$$\begin{aligned} p_T &= -(1-\beta/T)/(1+(1-\beta/T)^2), \\ &= -\frac{1}{2} + \frac{\beta^2}{4T^2}a(\beta,T), \\ q_T &= -\frac{1}{2} + \frac{\gamma^2}{4T^2}a(\gamma,T), \end{aligned} \tag{2.6}$$

and $\overset{d}{=}$ means equal in distribution as random functions (of β) on $C[0,\infty)$. (Here $C[0,\infty)$ denotes the space of continuous functions on $[0,\infty)$ and convergence is defined as uniform convergence on compact sets.)

Similarly, we find that

$$L_T''(\beta) \overset{d}{=} \left(\frac{\beta^2}{4T^4}\right)(1-\beta/(2T))^2 a^4(\beta,T)\left(\frac{d^2M}{d\rho^2}\right) + \left(\frac{b(\beta,T)}{2T^2}\right)\left(\frac{dM}{d\rho}\right), \tag{2.7}$$

where $b(\beta,T) \to 1$ as $T \to \infty$ uniformly on compacts sets.

The following theorem describes the joint limiting behavior of $L_T'(\beta)$ and $L_T''(\beta)$.

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 S denote the state space $S = \{\boldsymbol{f} = (f_1, f_2)$: where $f_1, f_2 \in C[0,\infty)$ and $f_1(0) = 0\}$. Then with $L_T(\beta)$ defined by (2.3),*

(i) $(L_T'(\beta), L_T''(\beta)) \xrightarrow{d} \left(\frac{\beta}{2}Y_\gamma(\beta), \frac{\beta}{2}Y_\gamma'(\beta) + \frac{1}{2}Y_\gamma(\beta)\right)$ *as* $T \longrightarrow \infty$,

where $\xrightarrow{d}$ *denotes weak convergence on S and*

$$Y_\gamma(\beta) = \sum_{k=1}^{\infty} \frac{4(\pi^2k^2 + \gamma^2)X_k^2}{(\pi^2k^2 + \beta^2)^2} - \sum_{k=1}^{\infty} \frac{4}{\pi^2k^2 + \beta^2}$$

with $\{X_k\} \sim$ NID $(0,1)$.

(ii) If $\hat\beta = \inf\{\beta \geq 0$: $L_T'(\beta) = 0$ and $L_T''(\beta) < 0\}$ (i.e., $\hat\beta$ is the local maximum of $L_T(\cdot)$ closest to 0), then

$$\hat\beta \xrightarrow{d} \tilde\beta_\gamma$$

where

$$\tilde\beta_\gamma = \inf\{\beta \geq 0 : \beta Y_\gamma(\beta) = 0 \quad \text{and} \quad \beta Y_\gamma'(\beta) + Y_\gamma(\beta) < 0\}.$$

REMARKS.

1. The existence of a local maximum $\hat\beta$ is assured by Anderson and Mentz (1980).

2. The theorem covers the non-invertible case when $\theta_0 \equiv 1$ by taking $\gamma = 0$.

3. It is established, in the course of the proof of the theorem (see also the appendix), that $\tilde\beta_\gamma = 0$ if and only if $Y_\gamma(0) < 0$ and if $Y_\gamma(0) > 0$, then $\tilde\beta_\gamma$ is the smallest zero of $Y_\gamma(\beta)$.

Corollary *Let $\hat\theta = (1 - \hat\beta/T)$ (i.e., $\hat\theta$ is the local maximum of the likelihood which is closest to the boundary at 1) and let P_γ denote the probability law under the model (2.1) with $\theta_0 = 1 - \gamma/T$. Then*

(a) $T(\hat\theta - 1) \xrightarrow{d} -\tilde\beta_\gamma$

(b) $P_\gamma[\hat\theta = 1] \to P[\tilde\beta_\gamma = 0] = P[(1/6 - W_1)/W_2 \geq \gamma^2]$,

where $W_1 = \sum_{k=1}^{\infty} \frac{X_k^2}{\pi^2k^2}$ *and* $W_2 = \sum_{k=1}^{\infty} \frac{X_k^2}{\pi^4k^4}$.

(c) For $x > 0$, $P_\gamma[\hat\beta > x] \to P[\tilde\beta_\gamma > x] = P[Y_\gamma(0) > 0, \tilde\beta_\gamma > x]$.

Proof of Corollary: (a) is immediate from (ii) of the theorem.

(b) Since with probability one, the sample paths of $Y(\beta)$ and $Y'(\beta)$ are continuous and cannot be zero on any interval, we have $\tilde\beta_\gamma = 0$ if and only if $Y_\gamma(0) < 0$ so that

$$\begin{aligned} P_\gamma[\hat\theta = 1] &= P_\gamma[\hat\beta = 0] \\ &\to P[\tilde\beta_\gamma = 0] \end{aligned}$$

$$
\begin{aligned}
&= P[Y_\gamma(0) < 0] \\
&= P\left[\sum_{k=1}^{\infty}\frac{X_k^2}{\pi^2k^2} + \gamma^2\sum_{k=1}^{\infty}\frac{X_k^2}{\pi^4k^4} \le \sum_{k=1}^{\infty}\frac{1}{\pi^2k^2}\right] \\
&= P\left[\frac{1/6 - W_1}{W_2} \ge \gamma^2\right].
\end{aligned}
$$

(c) It follows from the preceding argument that $\widetilde{\beta}_\gamma > 0$ if and only if $Y_\gamma(0) > 0\,(P[Y_\gamma(0) = 0] = 0)$. □

REMARK 4. The result given in (b) with $\gamma = 0$,

$$
P\left[\widetilde{\beta}_0 = 0\right] = P\left[W_1 \le 1/6\right] = .6575
$$

was previously derived in Anderson and Takemura (1986) and Tanaka and Satchell (1989).

Proof of Theorem 2.1: With $p_T = p_T(\beta) = \rho(1 - \beta/T)$ given in (2.6) we first show

$$
\left(T^{-2}\frac{dM}{d\rho}(p_T), T^{-4}\beta\frac{d^2M}{d\rho^2}(p_T)\right) \xrightarrow{d} \left(Y_\gamma(\beta), 2Y_\gamma'(\beta)\right) \quad \text{in} \quad C^2[0,\infty). \tag{2.8}
$$

As in Anderson and Takemura (1986), let k_T be a sequence of integers satisfying $k_T \to \infty$, $k_T/T \to 0$ and $k_T^2/T \to \infty$. Then for all $t = 1, \ldots, k_T$, and $\beta \in [0, M]$

$$
\begin{aligned}
2(T+1)^2(1 + 2p_T d_t) &= 2(T+1)^2\left(1 - d_t + \frac{1}{2}\frac{\beta^2}{T^2}d_t a(\beta,T)\right) \\
&= 2(T+1)^2\left(1 - \cos\left(\frac{\pi t}{T+1}\right)\right) + \left(\frac{T+1}{T}\right)^2\beta^2 d_t a(\beta,T) \\
&= \pi^2t^2 + t^4O(T^{-2}) + \beta^2\left(1 + o(1)\right) \\
&= \pi^2t^2 + t^2O\left(\frac{k_T}{T}\right)^2 + \beta^2\left(1 + o(1)\right) \\
&= (\pi^2t^2 + \beta^2)(1 + o(1)) \qquad (2.9)
\end{aligned}
$$

where $o(1)$ depends on k_T/T, T and M only. Since d_t is decreasing on $t = k_T + 1, \ldots, T$, we have

$$
\begin{aligned}
2(T+1)^2(1 - d_t) &\ge (T+1)^2\left(1 - \cos\frac{\pi k_T}{T+1}\right) \\
&= \pi^2k_T^2\left(1 + O\left(\frac{k_T}{T}\right)^2\right) \\
&= \pi^2k_T^2\left(1 + o(1)\right). \qquad (2.10)
\end{aligned}
$$

Also, for $t = 1, \ldots, k_T$, and all $\beta \in [0, M]$, we have from (2.9)

$$
\begin{aligned}
2(T+1)^2(1 + 2p_T d_t) - (\pi^2t^2 + \beta^2)d_t &= \left(\pi^2t^2 + \beta^2\right)\left(o(1) + 1 - d_t\right) \\
&= (\pi^2t^2 + \beta^2)o(1) \\
&= t^2o(1) \qquad (2.11)
\end{aligned}
$$

where $o(1)$ does not depend on t or β.

We first consider the individual terms which comprise $\frac{dM}{d\rho}(p_T)$ in (2.5). In particular, we show

$$\frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t}{1+2p_T d_t}-\sum_{t=1}^{k_T}\frac{4}{\pi^2t^2+\beta^2}\to 0 \quad \text{uniformly on } \beta\in[0,M], \tag{2.12}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1+2q_T d_t}{1+2p_T d_t}X_t^2-\frac{1}{T}\sum_{t=1}^{T}X_t^2\xrightarrow{P}0, \tag{2.13}$$

and

$$\frac{1}{(T+1))^2}\sum_{t=1}^{T}\frac{2d_t(1+2q_T d_t)X_t^2}{(1+2p_T d_t)^2}-\sum_{t=1}^{k_T}\frac{4(\pi^2t^2+\gamma^2)X_t^2}{(\pi^2t^2+\beta^2)^2}\xrightarrow{P}0, \tag{2.14}$$

where $\xrightarrow{P}$ is convergence in probability with respect to the uniform metric on $C[0,M]$.

Using (2.10), there exists a positive constant C such that for all $\beta\in[0,M]$ and T large

$$\begin{aligned}
\left|\frac{1}{(T+1)^2}\sum_{t=k_T+1}^{T}\frac{2d_t}{1+2p_T d_t}\right| &\le \frac{1}{(T+1)^2}\sum_{t=k_T+1}^{T}\frac{2}{\left(1-d_t+\frac{\beta^2}{2T^2}d_t a(\beta,T)\right)} \\
&\le \sum_{t=k_T+1}^{T}\frac{1}{(T+1)^2(1-d_t)-C} \\
&\le \sum_{t=k_T+1}^{T}\frac{1}{\pi^2k_T^2(1+o(1))-C} \\
&\le \frac{T}{\pi^2k_T^2(1+o(1))-C} \\
&\to 0.
\end{aligned} \tag{2.15}$$

Now from (2.9) and (2.11), we have for all $\beta\in[0,M]$ and T large

$$\begin{aligned}
\left|\frac{1}{(T+1)^2}\sum_{t=1}^{k_T}\frac{2d_t}{1+2p_T d_t}-\sum_{t=1}^{k_T}\frac{4}{\pi^2t^2+\beta^2}\right| &= \sum_{t=1}^{k_T}\frac{t^2o(1)}{(T+1)^2(1+2p_T d_t)(\pi^2t^2+\beta^2)} \\
&\le \sum_{t=1}^{k_T}\frac{t^2o(1)}{(\pi^2t^2+\beta^2)(1+o(1))\pi^2t^2} \\
&\le o(1)\sum_{t=1}^{k_T}\frac{1}{t^2} \\
&\to 0
\end{aligned}$$

as $T\to\infty$. This combined with (2.15) proves (2.12).

The difference in (2.13) is for all $\beta\in[0,M]$ bounded by

$$\begin{aligned}
\left|\frac{1}{T}\sum_{t=1}^{T}\frac{(1+2q_T d_t-(1+2p_T d_t))X_t^2}{1+2p_T d_t}\right| &\le \frac{1}{2T^3}\sum_{t=1}^{T}\frac{|\gamma^2a(\gamma,T)-\beta^2a(\beta,T)|X_t^2}{1+2p_T d_t} \\
&\le \frac{(\text{const})}{2(T+1)^3}\sum_{t=1}^{T}\frac{X_t^2}{1+2p_T d_t}.
\end{aligned} \tag{2.16}$$

Now, from (2.9), there exists a constant C such that

$$\frac{1}{2(T+1)^3}\sum_{t=1}^{k_T}\frac{X_t^2}{1+2p_T d_t} \le \frac{1}{T+1}\sum_{t=1}^{k_T}\frac{X_t^2}{\pi^2 t^2(1+o(1))}$$
$$\xrightarrow{P} 0$$

and using the argument for (2.15), the bound in (2.16) converges to zero in probability uniformly for $\beta \in [0, M]$. This proves (2.13).

Turning to (2.14), we have

$$\begin{aligned}
\frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t(1+2q_T d_t)X_t^2}{(1+2p_T d_t)^2} &= \frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t(1-d_t+\frac{\gamma^2}{2T^2}d_t a(\gamma,T))X_t^2}{(1+2p_T d_t)(1-d_t+\frac{\beta^2}{2T^2}d_t a(\beta,T))} \\
&= \frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t X_t^2}{1+2p_T d_t} \\
&\quad + \frac{1}{2(T+1)^2}\frac{(\gamma^2 a(\gamma,T)-\beta^2 a(\beta,T))}{T^2}\sum_{t=1}^{T}\frac{2d_t^2 X_t^2}{(1+2p_T d_t)^2} \\
&=: A + B.
\end{aligned}$$

Using the same argument given for (2.12), one readily shows that

$$\sup_{0<\beta\le M} |A - \sum_{t=1}^{k_T}\frac{4X_t^2}{\pi^2 t^2+\beta^2}| \xrightarrow{P} 0.$$

For β, we have (cf. (2.15))

$$\begin{aligned}
\frac{1}{2(T+1)^4}\sum_{t=k_T+1}^{T}\frac{2d_t X_t^2}{(1+2p_T d_t)^2} &\le \sum_{t=k_T+1}^{T}\frac{X_t^2}{((T+1)^2(1-d_t-C))^2} \\
&\le \sum_{t=k_T+1}^{T}\frac{X_t^2}{(\pi^2 k_T^2(1+o(1))-C)^2}
\end{aligned}$$

which has expectation converging to 0. Using (2.9), it follows that for all $\beta \in [0, M]$

$$\begin{aligned}
\left|\frac{1}{2(T+1)^4}\sum_{t=1}^{k_T}\frac{2d_t^2 X_t^2}{(1+2p_T d_t)^2}\right. &- \left.\sum_{t=1}^{k_T}\frac{4X_t^2}{(\pi^2 t^2+\beta^2)^2}\right| \\
&\le \sum_{t=1}^{k_T}\frac{|2d_t^2(\pi^2t^2+\beta^2)^2 - 8(T+1)^4(1+2p_T d_t)^2|X_t^2}{2(T+1)^4(1+2p_T d_t)^2(\pi^2t^2+\beta^2)^2} \\
&\le \sum_{t=1}^{k_T}\frac{t^2 o(1)(\pi^2t^2+\beta^2)(1+o(1))X_t^2}{(\pi^2t^2+\beta^2)^4(1+o(1))^2} \\
&\le o(1)\sum_{t=1}^{k_T}\frac{X_t^2}{t^4}
\end{aligned}$$

which has expectation converging to zero. We conclude that

$$\sup_{0\le\beta\le M} |B| \xrightarrow{P} 0$$

from which (2.14) easily follows.

Now using the right-hand side of (2.5) as $\frac{dM}{d\rho}(p_T)$, relations (2.12) – (2.15) imply

$$\frac{1}{T^2}\frac{dM}{d\rho}(p_T) \xrightarrow{P} Y_\gamma(\cdot). \tag{2.17}$$

For the second derivative of the concentrated likelihood,

$$\frac{d^2M}{d\rho^2}(p_T) \stackrel{d}{=} \sum_{t=1}^{T}\frac{4d_t^2}{(1+2p_Td_t)^2} - \frac{T}{\sum_{t=1}^{T}\left(\frac{1+2q_Td_t}{1+2p_Td_t}\right)X_t^2}\sum_{t=1}^{T}\frac{8d_t^2(1+2q_Td_t)X_t^2}{(1+2p_Td_t)^3}$$

$$+T\left[\sum_{t=1}^{T}\frac{(1+2q_Td_t)}{1+2p_Td_t}X_t^2\right]^{-2}\left[\sum_{t=1}^{T}\frac{(1+2q_Td_t)2d_tX_t^2}{(1+2p_Td_t)^2}\right]^2$$

It follows, using the same style of argument as above to handle each of the terms in $\frac{d^2M}{d\rho^2}(p_T)$, that

$$\begin{aligned}\frac{\beta}{T^4}\frac{d^2M}{d\rho^2}(p_T) &\xrightarrow{P} \sum_{t=1}^{\infty}\frac{16\beta}{(\pi^2t^2+\beta^2)^2} - \sum_{t=1}^{\infty}\frac{32(\pi^2t^2+\gamma^2)\beta X_t^2}{(\pi^2t^2+\beta^2)^3} \\ &\qquad +0\cdot\sum_{t=1}^{\infty}\frac{4(\pi^2t^2+\gamma^2)\beta X_t^2}{(\pi^2t^2+\beta^2)^2} \\ &= 2Y_\gamma'(\beta).\end{aligned}$$

Combining this convergence with (2.17) gives joint convergence in (2.8) as asserted. Finally, from (2.4), (2.7), and (2.8), we conclude

$$\begin{aligned}(L_T'(\beta), L_T''(\beta)) &= \left(\frac{\beta}{2T^2}\frac{dM}{d\rho}(p_T), \frac{\beta^2}{4T^4}\frac{d^2M}{d\rho}(p_T) + \frac{1}{2T^2}\frac{dM}{d\rho}(p_T)\right) + o_p(1) \\ &\xrightarrow{d} \left(\frac{\beta}{2}Y_\gamma(\beta), \frac{\beta}{2}Y_\gamma'(\beta) + \frac{1}{2}Y_\gamma(\beta)\right)\end{aligned}$$

in $C^2[0,\infty)$ and hence, in S since S is a closed subset of $C^2[0,\infty)$ and $L_T'(0) = 0$. This completes the proof of part (i).

Let $\widetilde{S}$ be the subset of S defined by

$$\widetilde{S} = \{\boldsymbol{f} = (g, g')\text{: where } g(\cdot) \text{ and } g'(\cdot) \in C[0,\infty),\ g(\beta) = 0,\ g'(\beta) < 0 \text{ for some } \beta \ge 0, \text{ and } |g(x)| + |g'(x)| \ne 0 \text{ for all } x \ge 0\}.$$

For $\boldsymbol{f} = (f_1, f_2) \in S$, define the mapping $h : S \to [0,\infty)$ by

$$h(\boldsymbol{f}) = \begin{cases} \inf\{\beta \ge 0 : f_1(\beta) = 0 \text{ and } f_2(\beta) < 0\}, & \text{if } \{\beta \ge 0 : f_1(\beta) = 0 \text{ and } f_2(\beta) < 0\} \ne \phi, \\ 0, & \text{if } \{\beta \ge 0 : f_1(\beta) = 0 \text{ and } f_2(\beta) < 0\} = \phi.\end{cases}$$

We show h is continuous at every $\boldsymbol{f} \in \widetilde{S}$. To this end, suppose $\boldsymbol{f} = (g, g') \in \widetilde{S}$ and let $\boldsymbol{f}_n = (f_n^{(1)}, f_n^{(2)}) \to (g, g')$. If $h(\boldsymbol{f}) = 0$, then $g(0) = 0$ and $g'(0) < 0$. But $f_n^{(1)}(0) = 0$ and $f_n^{(2)}(0) \to g'(0) < 0$, whence $h(\boldsymbol{f}_n) = 0$ for all n large. Now suppose $h(\boldsymbol{f}) = \beta > 0$. Since $g'(\beta) < 0$, there exists a neighborhood $(\beta - \delta, \beta + \delta)$ such that $g(\beta - \delta) > 0 > g(\beta + \delta) > 0$ and $g'(x) < 0$ for all $x \in (\beta - \delta, \beta + \delta)$. Since $(f_n^{(1)}, f_n^{(2)})$ converges to (g, g') on compact sets, we have for all n sufficiently large

$$f_n^{(1)}(\beta - \delta) > 0 > f_n^{(1)}(\beta + \delta), \qquad \text{and} \qquad f_n^{(2)}(x) < 0 \qquad \text{for all} \quad x \in (\beta - \delta, \beta + \delta).$$

Thus there exists $\beta_n \in (\beta - \delta, \beta + \delta)$ such that $f_n^{(1)}(\beta_n) = 0$ and $f_n^{(2)}(\beta_n) < 0$. So $h(\boldsymbol{f}_n) \leq \beta_n \leq \beta + \delta$ is a bounded sequence. Let β^* be a limit point of some subsequence $h(\boldsymbol{f}_{n_j})$ and since $\delta > 0$ was arbitrary $\beta^* \leq \beta$. If $\beta^* < \beta$, then by uniform convergence, we must have

$$g(\beta^*) = 0 \qquad \text{and} \qquad g'(\beta^*) \leq 0.$$

By definition of $\beta = h(\boldsymbol{f})$, this implies that $g'(\beta^*) = 0$ which contradicts $\boldsymbol{f} \in S$ (since $|g(x)| + |g'(x)| > 0$ for all $x \geq 0$). This forces $\beta^* = \beta$ and establishes the continuity of h at $\boldsymbol{f}$.

In the appendix we show that $P\left[\left(\frac{\beta}{2}Y_\gamma(\beta), \frac{\beta}{2}Y_\gamma'(\beta) + \frac{1}{2}Y_\gamma(\beta)\right) \in \widetilde{S}\right] = 1$ and hence by the continuous mapping theorem $\hat{\beta} = h(L_T', L_T'') \xrightarrow{d} \widetilde{\beta}_\gamma$. □

3 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 actual finite sample distribution of the MLE. The accuracy of the standard asymptotic normal distribution will also be evaluated.

Tanaka and Satchell(1989) provide an approximate solution to the limit distribution of the MLE. Their method does not yield accurate results without ad hoc adjustments and only applies to deriving an approximation when $\theta_0 = 1$. In contrast the corollary to Theorem 2.1 applies to all $\gamma > 0$ thereby allowing approximation to $P_{\theta_0}[\hat{\theta} \leq x]$ for $\theta_0 \leq 1$.

It is easy to get replicates of $\widetilde{\beta}_\gamma$ in (a) of the Corollary from the expression for $Y_\gamma(\beta)$. This means that the distribution of $\widetilde{\beta}_\gamma$ is readily obtained via simulation techniques. The procedure we have adopted is as follows:

Step 1. The infinite series for $Y_\gamma(\beta)$ is truncated at some large integer N.

Step 2. For the truncated series, which we shall continue to call $Y_\gamma(\beta)$, we compute $Y_\gamma(0)$. If $Y_\gamma(0) < 0$, then the replicate of $\tilde{\beta}_\gamma$ is 0.

Step 3. If $Y_\gamma(0) > 0$, then the replicate of $\tilde{\beta}_\gamma$ is defined to be the smallest non-negative zero of $Y_\gamma(\beta)$.

Although we took $N = 25,000$ in all of our simulations, reasonably accurate results can be attained with N as small as 1000. In Step 3, we used the IMSL root finder ZREAL to compute $\tilde{\beta}_\gamma$. The function $Y_\gamma(\cdot)$ is quite smooth so that locating the first zero-crossing is rather straightforward. With $N = 25,000$ and $\gamma = 0$, it took approximately 60 seconds to compute 100 replicates of $\tilde{\beta}_0$ on an HP 720 workstation. For $N = 1000$, the time was reduced to 2.4 seconds. For all of the limit results reported below we generated 100,000 replicates of $\tilde{\beta}_0$ and 10,000 replicates of $\tilde{\beta}_\gamma$ for $\gamma > 0$.

In order to compare the limit distribution with the finite sample distribution of the MLE, it was necessary to generate replicates of the MLE. We used the innovations algorithm (see Brockwell and Davis (1991)) to compute the reduced likelihood. We then computed two different estimates of θ: (i) we maximized the likelihood using a nonlinear optimization program and (ii) we found the largest local maximum by evaluating the reduced likelihood at $\theta = 1, 1-.0001, 1-.0002$, etc. until a local maximum was achieved. Since the results for the MLE and the largest local maximum were nearly the same for the sample sizes and parameter values we tried, we will refer to either estimate as the MLE (and denote it by $\hat{\theta}$) throughout the remainder of the paper. Results reported below are based on 100,000 replicates of the MLE.

Figure 1 compares the sampling distribution of $T(\hat{\theta} - 1)$ with the distribution of the limit random variable $-\tilde{\beta}_0$ when $\theta_0 = 1$ ($\gamma = 0$) for sample sizes $T = 25, 50, 100$. 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 MLE and the limit approximation are in good agreement (see Table 3.1).

Table 3.1. Quantiles of $T(\hat{\theta} - 1)$ (empirically estimated) when $\theta_0 = 1$.

	p			
T	.01	.05	.1	.2
10	-11.818	-6.481	-4.545	-2.834
25	-11.320	-6.505	-4.657	-2.870
50	-11.375	-6.565	-4.745	-2.900
100	-11.170	-6.500	-4.670	-2.870
Limit	-11.253	-6.522	-4.736	-2.892

In Figures 2–5, the sampling distribution of $T(\hat{\theta} - 1)$ is plotted together with the distribution of the limit random variable $-\tilde{\beta}_\gamma$ and the normal approximation for the case $T = 50$ and $\theta_0 = .9, .8, .7, .6$ (i.e. $\gamma = 2.5, 5, 10, 15$), respectively. The normal approximation is based on the asymptotic normal distribution for the MLE given by $N(0, (1-\theta^2)/T)$. As is quite clear from these figures, the approximation based on the empirical distribution of $-\tilde{\beta}_\gamma$ with $\gamma = (1-\theta_0)T$ is very accurate and considerably outperforms the asymptotic normal distribution. As θ_0 moves away from the boundary, the limit and empirical distributions begin to approach the normal approximation. For the range of θ_0's displayed though, the limit distribution based on $\tilde{\beta}_\gamma$ provides a much better approximation to the distribution of the MLE. This is reaffirmed in Table 3.2 which gives a comparison of the quantiles and the jump at 0 of the 3 distributions for a variety of θ_0's.

Table 3.2. Comparison of quantile estimation of $T(\hat{\theta}-1)$ and estimation of $P_{\theta_0}[\hat{\theta}=1]$ using the limit distribution (see (a) and (b) of the Corollary) and the normal approximation. The sample size is 50 and the exact values are computed via simulation.

θ_0		p: .01	.05	.1	.2	$P_\gamma[\hat{\theta}=1]$
.7	Exact	−29.065	−24.030	−21.695	−19.020	.054
	Limit	−29.407	−24.472	−21.897	−19.019	.060
	Normal	−27.750	−23.308	−21.472	−19.249	.001
.8	Exact	−22.455	−17.750	−16.650	−13.260	.131
	Limit	−22.704	−17.972	−15.699	−13.226	.141
	Normal	−19.872	−16.980	−15.438	−13.570	.009
.9	Exact	−15.810	−11.265	−9.365	−7.360	.331
	Limit	−15.913	−11.245	−9.373	−7.240	.347
	Normal	−12.172	−10.071	−8.951	−7.594	.052
.95	Exact	−12.765	−8.190	−6.385	−4.520	.521
	Limit	−12.791	−8.088	−6.317	−4.395	.542
	Normal	−7.637	−6.133	−5.330	−4.358	.129
.99	Exact	−11.475	−6.680	−4.790	−2.980	.647
	Limit	−11.180	−6.495	−4.734	−2.887	.663
	Normal	−2.820	−2.141	−1.779	−1.339	.308

While we've seen that the limit distribution of $-\tilde{\beta}_\gamma$ provides an extremely accurate approximation to the lower quantiles of $T(\hat{\theta}-1)$, improvement can be made near $x=0$ for larger quantiles and, in particular, at the jump point $x=0$. To improve the estimate of $P_{\theta_0}[\hat{\theta}_0=1]$, we use an alternate γ given by $\gamma' = T(1+8T^{-3/2})(1-\theta_0)$. The additional term in $T^{-3/2}$ was derived by calibrating $P[\tilde{\beta}_\gamma=0]$ with the values published in Anderson and Takemura(1986) and Cryer and Ledolter (1981). Use of this adjustment leads to considerably improved results as is clear from Table 3.3. Although the use of the additional factor $(1+8T^{-3/2})$ in γ' has not been derived on theoretical grounds the improved approximation obtained through its use indicates that such "higher" order terms are worth investigating in more detail. It is also important to note that $P[\tilde{\beta}_\gamma=0]$ can be computed easily for any value of $\gamma \geq 0$ from the distribution of $(1/6-W_1)/W_2$ (see (b) of the

Corollary). It is much easier to estimate the distribution of this random variable via simulation than to generate replicates of $\tilde{\beta}_\gamma$ and compute the proportion of such replicates which are equal to 0.

Table 3.3. Comparison of the approximation to $P_{\theta_0}[\hat{\theta} = 1]$ using (b) of the Corollary with $\gamma = T(1-\theta_0)$ and $\gamma' = (1+8T^{-3/2})\gamma$. Exact values are from Table 2 in Cryer and Ledolter (1981).

	$T = 10$			$T = 25$			$T = 50$			$T = 100$		
θ_0	Exact	γ'	γ	Exact	γ'	γ	Exact	γ'	γ	Exact	γ'	γ
.5	.280	.273	.343	.076	.082	.095	.013	.014	.015	.001	.001	.001
.6	.353	.342	.417	.119	.128	.141	.025	.026	.028	.002	.003	.003
.7	.440	.436	.496	.193	.201	.219	.055	.059	.062	.007	.007	.008
.8	.533	.533	.571	.319	.324	.343	.130	.135	.141	.027	.027	.028
.9	.609	.610	.634	.513	.523	.533	.333	.337	.343	.136	.139	.141
1.0	.637	.659	.659	.649	.659	.659	.653	.659	.659	.655	.659	.659

4 Applications to Hypothesis Testing

In this section the distribution for $\tilde{\beta}_\gamma$ is used as the basis for hypothesis testing. Initially tests of the hypotheses $H_0 : \theta = 1$ versus $H_A : \theta < 1$ will be derived based on the MLE $\hat{\theta}$. Since the standard asymptotic normal theory does not apply under the null hypothesis a test cannot be constructed using the normal distribution. However a test can be constructed using the above results. In this test H_0 will be rejected at level α if $\hat{\theta} < 1 - b_\alpha/T$ where b_α is the $(1-\alpha)$ quantile of $\tilde{\beta}_0$, i.e.

$$\alpha = P[\tilde{\beta}_0 > b_\alpha]$$

Table 4.1 gives the values of b_α for various α. For example, H_0 is rejected at level .05 if $\hat{\theta} < 1 - 6.522/T$. Thess critical values are .8696 and .9348 for $T = 50$ and $T = 100$ respectively.

Table 4.1. $(1-\alpha)$ quantiles for the distribution of $\tilde{\beta}_0$.

	α			
	.01	.025	.05	.1
b_α	11.253	8.478	6.522	4.736

Using these b_α as the basis for a nominal α level test the actual probability of Type I error (i.e. $P_1[\hat{\theta} < 1 - b_\alpha/T]$) achieved for various sample sizes is calculated via simulation and reported in Table 4.2

As can be seen from Table 4.2 the actual and nominal type I error probabilities are in close agreement. This is not unexpected given the closeness of the approximation of the limit distribution discussed in Section 3.

Table 4.2. Simulated type I errors using cutoffs given in Table 4.1.

T	α = .01	.025	.05	.1
10	.0116	.0257	.0493	.0934
25	.0103	.0256	.0497	.0970
50	.0104	.0254	.0508	.1002
100	.0097	.0246	.0493	.0975

The asymptotic theory of Section 2 also allows us to calculate the nominal power of the above test against local alternatives of the form:

$$H_A : \theta = \theta_A$$

where $\theta_A = 1 - \gamma/T$. The power of a test based on on a b_α cutoff value can be approximated by $P[\tilde{\beta}_\gamma > b_\alpha]$. This limit approximation is compared with the actual power of the test in Table 4.3 for $\theta_A = .7, .8, .9, .95, .99$ which, with $T = 50$ corresponds to $\gamma = 15, 10, 5, 2.5, .5$. Again the limit approximations are quite good.

While the power does not increase very rapidly as θ_A moves away from 1 the above test has an asymptotic power equal to 1 against any fixed local alternative. This property is not shared by the procedure proposed in Chan and Tsay (1992).

Table 4.3. Power ($P_{\theta_A}[\hat{\theta} < 1 - b_\alpha/T]$) calculations for $T = 50$ using the limit approximation for tests with b_α given in Table 4.1. Exact values are computed via simulation.

θ_A		α = .01	.025	.05	.10
.7	Exact	.688	.814	.873	.907
	Limit	.650	.789	.858	.898
.8	Exact	.326	.532	.666	.760
	Limit	.317	.504	.641	.741
.9	Exact	.050	.137	.258	.405
	Limit	.050	.136	.247	.385
.95	Exact	.017	.045	.095	.185
	Limit	.017	.043	.093	.177
.99	Exact	.011	.026	.053	.102
	Limit	.010	.024	.050	.100

The asymptotic results of Section 2 can also be used to test hypotheses of the form $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_A$ where $\theta_A < \theta_0 < 1$. The critical region for such a test is given by $\hat{\theta} < 1 - b_\alpha/T$ where now the b_α are calculated from

$$P[\tilde{\beta}_{\gamma_0} > b_\alpha]$$

with $\gamma_0 = T(1 - \theta_0)$. As above, the power of the test can be approximated by

$$P[\tilde{\beta}_{\gamma_A} > b_\alpha]$$

where $\gamma_A = T(1 - \theta_A)$.

This test can also be compared with one based on the asymptotic normal distribution. The critical region for such a test has the form

$$\hat{\theta} < \theta_0 - z_\alpha[(1 - \theta_0^2)/T]^{.5}$$

where z_α is the $(1 - \alpha)$ quantile of the standard normal distribution. Likewise the power of this test is calculated as $P[U < \theta_0 - z_\alpha[(1 - \theta_0^2)/T]^{.5}$ with $U \sim \text{NID}(\theta_A, (1 - {\theta_A}^2)/T)$.

The actual size and power of the two test procedures is reported in Table 4.4 for several combinations of (θ_0, θ_A). With the exception of the case $(\theta_0, \theta_A) = (.90, .80)$ $\alpha = .05$, the size of the test based on the normal approximation is much too large making power comparisons difficult. On the other hand, the size calculation of the tests based on the limit distribution are close to their respective nominal values.

Table 4.4. Actual size and power calculations for testing $H_0 : \theta = \theta_0$ vs $H_A : \theta = \theta_A$ using the limit distribution and the normal approximation when $T = 50$.

		α			
(θ_0, θ_A)		.025		.05	
		Size	Power	Size	Power
(.99,.95)	Limit	.0277	.0468	.0532	.0954
	Normal	.2367	.3637	.0611	.1102
(.95,90)	Limit	.0275	.0814	.0519	.1570
	Normal	.0893	.2365	.1101	.2881
(.90,.80)	Limit	.0265	.2122	.0535	.3269
	Normal	.0541	.3414	.0538	.3404

5 Applications to Confidence Intervals

A comparison between confidence intervals for θ_0 based on the asymptotic normal distribution and those based on the results of Theorem 2.1 is now made. One-sided intervals based on $\tilde{\beta}_\gamma$ can be approximately constructed as follows. Let $1-\alpha$ be the required confidence level and set

$$\theta_U = 1 - b_{1-\alpha}^{-1}(\hat{\beta})/T$$

where $b_{1-\alpha}(\gamma)$ is the $1-\alpha$ quantile of $\tilde{\beta}_\gamma$, i.e.

$$1-\alpha = P[\tilde{\beta}_{\gamma_0} \leq b_{1-\alpha}(\gamma_0)].$$

Then a $100(1-\alpha)\%$ one-sided confidence interval for θ_0 is $[-1, \theta_U]$.

To obtain the upper limit of the interval in this procedure, an extended version of Table 3.2 must be constructed in order to compute the function $b_{1-\alpha}(\gamma_0)$. This function is then inverted to get $b_{1-\alpha}^{-1}$. For example, if $\hat{\theta}$ was observed as .776, then with $T = 50$, $\hat{\beta} = 11.2$ and $b_{.95}^{-1}(11.2) = 5$ and $b_{.99}^{-1}(11.2) = .5$. Hence the one-sided 95% interval is $[-1, 1-5/50]$ or $[-1, .9]$ and the one-sided 95% interval is $[-1, 1-.5/50]$ or $[-1, .99]$.

To construct a lower confidence bound for θ_0, one could use $[\theta_L, 1]$ where $\theta_L = 1 - b_\alpha^{-1}(\hat{\beta})/T$. However, since for a wide range of γ's, the distribution of $\tilde{\beta}_\gamma$ has substantial mass at 0 (greater than most commonly used values of α), $b_\alpha(\gamma)$ will be identically 0 for such γ's. In this case, the interval $[\theta_L, 1]$ degenerates into the single point 1.

There are a number of possible methods for constructing intervals based on a normal approximation to the distribution of $\hat{\theta}$. Three such possibilities are based on the following quantities, each of which has a standard normal limit distribution provided $\theta_0 < 1$:

$$\begin{aligned} Z_1 &= \frac{T^{\frac{1}{2}}(\hat{\theta}-\theta_0)}{(1-{\theta_0}^2)^{\frac{1}{2}}} \\ Z_2 &= \frac{T^{\frac{1}{2}}(\hat{\theta}-\theta_0)}{(1-\hat{\theta}^2)^{\frac{1}{2}}} \\ Z_3 &= (l''(\hat{\theta})/2)^{\frac{1}{2}}(\hat{\theta}-\theta_0) \end{aligned}$$

where

$$l(\theta) = -M(\rho(\theta)).$$

Use of an estimate of the asymptotic variance in Z_3 creates a statistic whose distribution does not appear to have a discrete component at $|\theta| = 1$, as is illustrated in Dunsmuir(1981) for example. Use of Z_1 leads to the following two-sided confidence interval. Let $c = z_{\alpha/2}^2/T$ then the $100(1-\alpha)\%$ interval is $[\theta_L(1), \theta_U(1)]$ where

$$\theta_L(1) = \max(-1, \frac{\hat{\theta} - \sqrt{c^2 + c(1-\hat{\theta}^2)}}{1+c})$$

and

$$\theta_U(1) = \min\left(\frac{\hat\theta + \sqrt{c^2 + c(1-\hat\theta^2)}}{1+c}, 1\right)$$

Note that if $\hat\theta = 1$ then the interval reduces to $[(1-c)/(1+c), 1]$.

Table 5.1. Coverage probabilities of upper confidence intervals based on the 4 methods: $\tilde\beta_\gamma$, Z_1, Z_2, and Z_3. Sample size is 50 and tabulated results are based on 1000 replicates.

		$1-\alpha$		
θ_0		.99	.95	.90
	β_γ	.990	.953	.894
.6	Z_1	.982	.882	.885
	Z_2	.990	.960	.905
	Z_3	.992	.967	.913
	β_γ	.990	.947	.896
.7	Z_1	.984	.929	.877
	Z_2	.992	.961	.914
	Z_3	.993	.972	.920
	β_γ	.989	.946	.881
.8	Z_1	.978	.928	.874
	Z_2	.992	.962	.922
	Z_3	.992	.977	.934
	β_γ	.990	.949	.885
.9	Z_1	.965	.918	.862
	Z_2	.995	.969	.933
	Z_3	.994	.982	.945
	β_γ	.989	.948	.883
.95	Z_1	.928	.869	.831
	Z_2	.992	.960	.916
	Z_3	.994	.978	.938

Intervals based on Z_2 and Z_3, denoted by $[\theta_L(2), \theta_U(2)]$ and $[\theta_L(3), \theta_U(3)]$ respectively, take the form

$$\begin{aligned}
\theta_L(2) &= \max(-1, \hat\theta - z_{\alpha/2}\sqrt{(1-\hat\theta^2)/T}) \\
\theta_U(2) &= \min(\hat\theta + z_{\alpha/2}\sqrt{(1-\hat\theta^2)/T}, 1) \\
\theta_L(3) &= \max(-1, \hat\theta - z_{\alpha/2}\sqrt{l''(\hat\theta)/(2T)})
\end{aligned}$$

and

$$\theta_U(3) \quad = \quad \min(\hat{\theta} + z_{\alpha/2}\sqrt{l''(\hat{\theta})/(2T)}, 1)$$

One advantage of the latter interval is that it will not degenerate to a single point as frequently as the interval based on Z_2. For example when $\hat{\theta} = 1$ the interval $[\theta_L(2), \theta_U(2)]$ reduces to the single point 1.

Table 5.1 compares the coverage characteristics of upper one-sided confidence intervals using the above 4 methods. As is evident from the table, the method based on $\tilde{\beta}_\gamma$ has extremely good coverage properties even when the true θ_0 is as small as .6. It is also interesting to note that the simple method based on Z_2 outperforms the other two normal based procedures.

Appendix

Let $X(\beta) = \frac{\beta}{2}Y_\gamma(\beta)$, $X'(\beta) = \frac{\beta}{2}Y'_\gamma(\beta) + \frac{1}{2}Y_\gamma(\beta)$, where $Y_\gamma(\beta)$ is the process defined in the statement of Theorem 2.1. In order to complete the proof of Theorem 2.1, it remains to show $(X(\cdot), X'(\cdot)) \in \tilde{S}$ *a.s.*, i.e.

$$(A1) \qquad P[X(\beta) = X'(\beta) = 0 \text{ for some } \beta \geq 0] = 0$$

and

$$(A2) \qquad P[X(\beta) = 0, X'(\beta) < 0 \text{ for some } \beta \geq 0] = 1.$$

Proof of (A1): First note that $X'(0) = \frac{1}{2}Y_\gamma(0) \neq 0$ *a.s.*, since $Y_\gamma(0)$ has a continuous distribution. For notational simplicity define $A(\phi) = X(\phi^{1/2})$, i.e.

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

and

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

where $\{X_t\}$ ~NID(0, 1).

Conditional on $X_2, X_3, \ldots$, let $C = C(X_2, X_3, \ldots,)$ denote the set of X_1 for which $A(\phi) = A'(\phi)$ for some $\phi \geq 0$. Since

$$P[X_1 \in C \mid X_2, X_3, \ldots] = P[X(\beta) = X'(\beta) = 0 \text{ for some } \beta \geq 0 \mid X_2, X_3, \ldots],$$

(A1) will follow once we show the left-hand side above is 0 a.s. For each $X_1 \in C$, there exists $\phi = \phi(X_1)$ such that $A(\phi) = A'(\phi) = 0$. Solving $A(\phi) = 0$ for X_1 and substituting its value into the equation $A'(\phi) = 0$, we obtain, after some simplification,

$$g(\phi) = \sum_{k=1}^{\infty} \left(\frac{2}{(\pi^2k^2 + \phi)^2} - \frac{4}{\pi^2k^2 + \phi}X_k^2\right) - 4\sum_{k=2}^{\infty} \left(\frac{\pi^2k^2 + \gamma^2}{(\pi^2k^2 + \phi)^2} - \frac{\pi^2k^2 + \gamma^2}{(\pi^2k^2 + \phi)^3}\right) = 0.$$

Extending the definition of $g(\cdot)$ to the region $\phi > -\pi^2$, one checks that $g(\cdot)$ is analytic (as a function of a real variable) on this region. Since $g(\phi) \to \infty$ as $\phi \downarrow -\pi^2$, $g(\cdot)$ is not identically zero on $\phi > -\pi^2$ and hence the zeros of $g(\cdot)$ must be isolated. This implies $\{\phi(x_1), x_1 \in C\}$ is countable and since each ϕ in this set completely determines the value of X_1^2 through the relation $A(\phi) = 0$, C must also be countable. It now follows that

$$P[X_1 \in C \mid X_2, X_3, \ldots] = 0 \quad a.s.$$

as asserted. □

Proof of (A2): We first show that

$$(A3) \qquad P[Y_\gamma(\beta) \geq 0 \text{ for all } \beta \geq 0] = 0.$$

If this probability is greater than $\epsilon > 0$, then for all β large

$$P[Y_\gamma(\beta) \geq 0] > \epsilon.$$

But

$$\begin{aligned} EY_\gamma(\beta) &= -\sum_{k=1}^{\infty} \frac{4}{\pi^2 k^2 + \beta^2} + 4 \sum_{k=1}^{\infty} \frac{\pi^2 k^2 + \gamma^2}{(\pi^2 k^2 + \beta^2)^2} \\ &= -4 \sum_{k=1}^{\infty} \frac{(\beta^2 - \gamma^2)}{(\pi^2 k^2 + \beta^2)^2} \end{aligned}$$

and

$$\begin{aligned} \operatorname{Var}(Y_\gamma(\beta)) &= 12 \sum_{k=1}^{\infty} \frac{(\pi^2 k^2 + \gamma^2)^2}{(\pi^2 k^2 + \beta^2)^4} \\ &\leq 12 \sum_{k=1}^{\infty} \frac{1}{(\pi^2 k^2 + \beta^2)^2} \end{aligned}$$

for $\beta \geq \gamma$. Since $\beta^2 \sum_{k=1}^{\infty} \frac{1}{(\pi^2 k^2 + \beta^2)^2} \to \frac{1}{4}$, we have by Chebyshev's inequality

$$\begin{aligned} P[Y_\gamma(\beta) \geq 0] &\leq \frac{\operatorname{Var}(Y_\gamma(\beta))}{(EY_\gamma(\beta))^2} \\ &\to 0 \end{aligned}$$

as $\beta \to \infty$, a contradiction.

Now the event $Y_\gamma(0) < 0$ is contained in the event described in (A2) so suppose $Y_\gamma(0) > 0$. By (A3), there exists a $\beta > 0$ such that $Y_\gamma(x) > 0$ for all $x < \beta$ and $Y_\gamma(\beta) = 0$. Certainly $Y_\gamma'(\beta) \leq 0$ and since $Y_\gamma'(\beta) = 0$ is precluded by (A1), we must have $Y_\gamma'(\beta) < 0$ which establishes (A2). □

ACKNOWLEDGEMENT. We are indebted to Yi-Ching Yao for his invaluable assistance in the proof of (A1).

REFERENCES

Anderson, T. W. and Raul P. Mentz (1980). On the Structure of the Likelihood Function of Autoregressive and Moving Average Models. *Journal of Time Series Analysis* **1**, 83-94.

Anderson, T. W. and Akimichi Takemura (1986). Why Do Noninvertible Estimated Moving Averages Occur? *Journal of Time Series Analysis* **7**, 235-254.

Box, G. E. P. and G. J. Jenkins (1976). *Time Series Analysis: Forecasting and Control, revised edition.* Holder-Day, San Francisco.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods, 2nd Edition.* Springer-Verlag, New York.

Cryer, J. D. and J. Ledolter (1981). Small-Sample Properties of the Maximum Likelihood Estimator in the First-Order Moving Average Model. *Biometrika* **68**, 691-694.

Davidson, J. (1981a). Problems with the Estimation of Moving Average Processes. *Journal of Econometrics* **16**, 295-310.

Davidson, J. (1981b). Small Sample Properties of Estimators of the Moving Average Process. In *Proceedings of the Econometric Society European Meeting*, 1979, edited by E. G. Chasratsis, pp. 27-62.

Dunsmuir, W.T.M. (1981) Estimation for Stationary Time Series When Data Are Irregularly Spaced or Missing. In *Applied Time Series II*, edited by David F. Findley, Academic Press, New York.

Kang, K. M. (1975). A Comparison of Estimators for Moving Average Processes. Unpublished, Australian Bureau of Statistics.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Pesaran, M. H. (1983). A Note on the Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors with Roots on the Unit Circle. *Australian Journal of Statistics* **25**, 442-448.

Sargan, J.D. and Bhargava, A. (1983). Maximum Likelihood Estimation of Regression Models with First-Order Moving Average Errors when the Root Lies on the Unit Circle. *Econometrics* **51**, 799-820.

Tanaka, K. and S. E. Satchell (1989). Asymptotic Properties of the Maximum-Likelihood and Nonlinear Least-Squares Estimators for Noninvertible Moving Average Models. *Econometric Theory* **5**, 333-353.

Figure 1: Comparison of Sampling CDF with Limit CDF

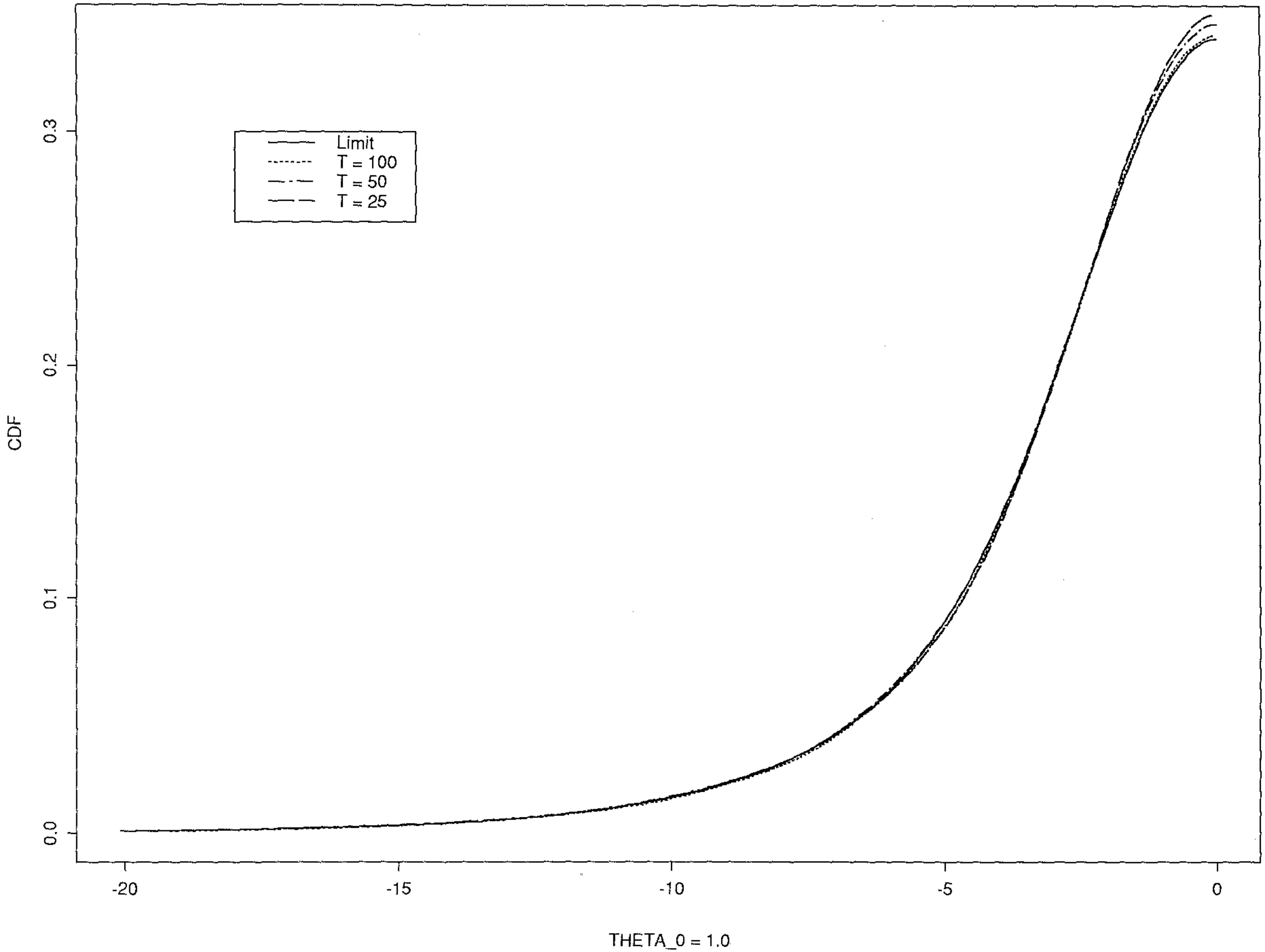

Figure 2: Comparison of Normal and Limit CDF's to Empirical

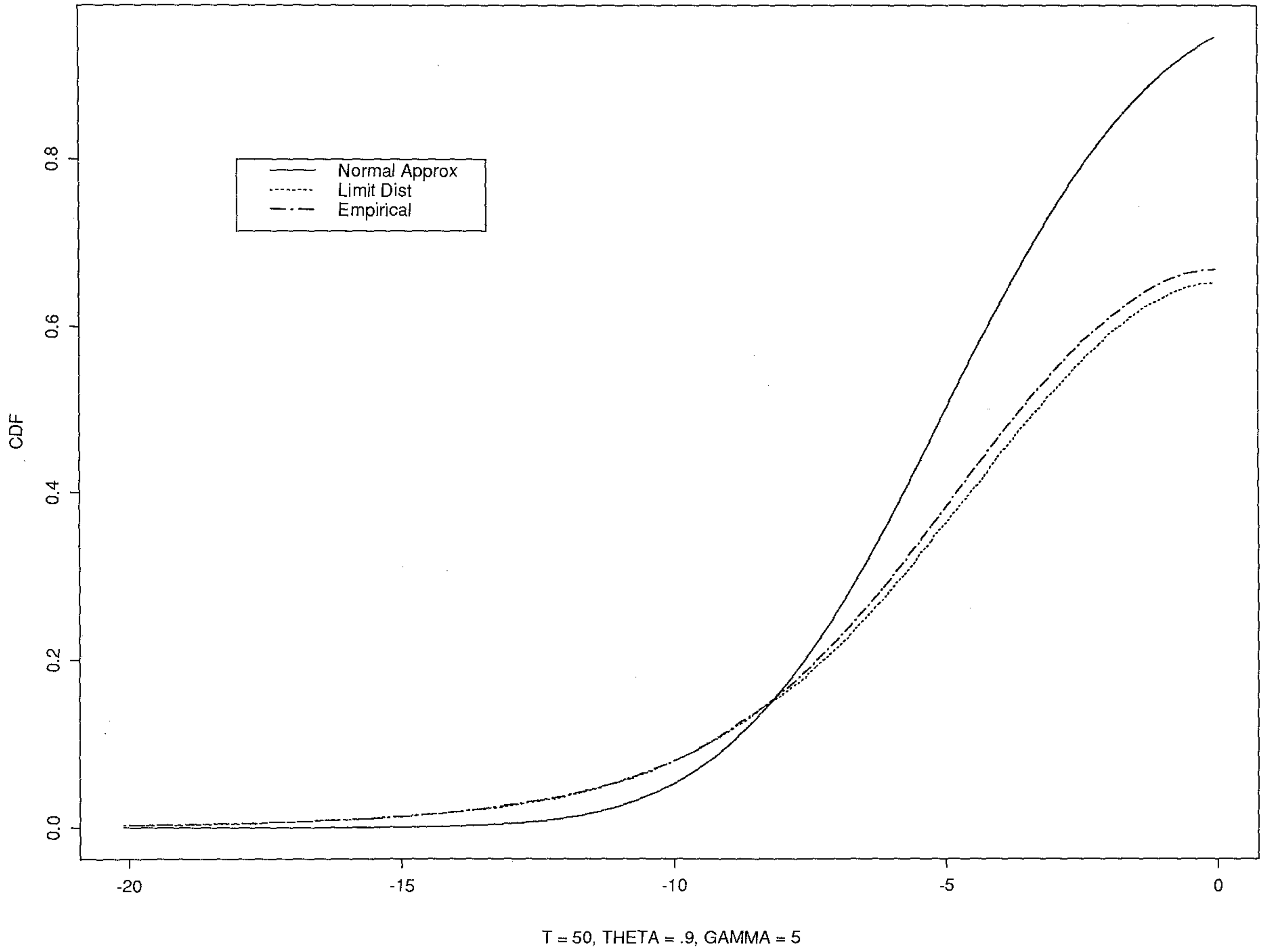

Figure 3: Comparison of Normal and Limit CDF's to Empirical

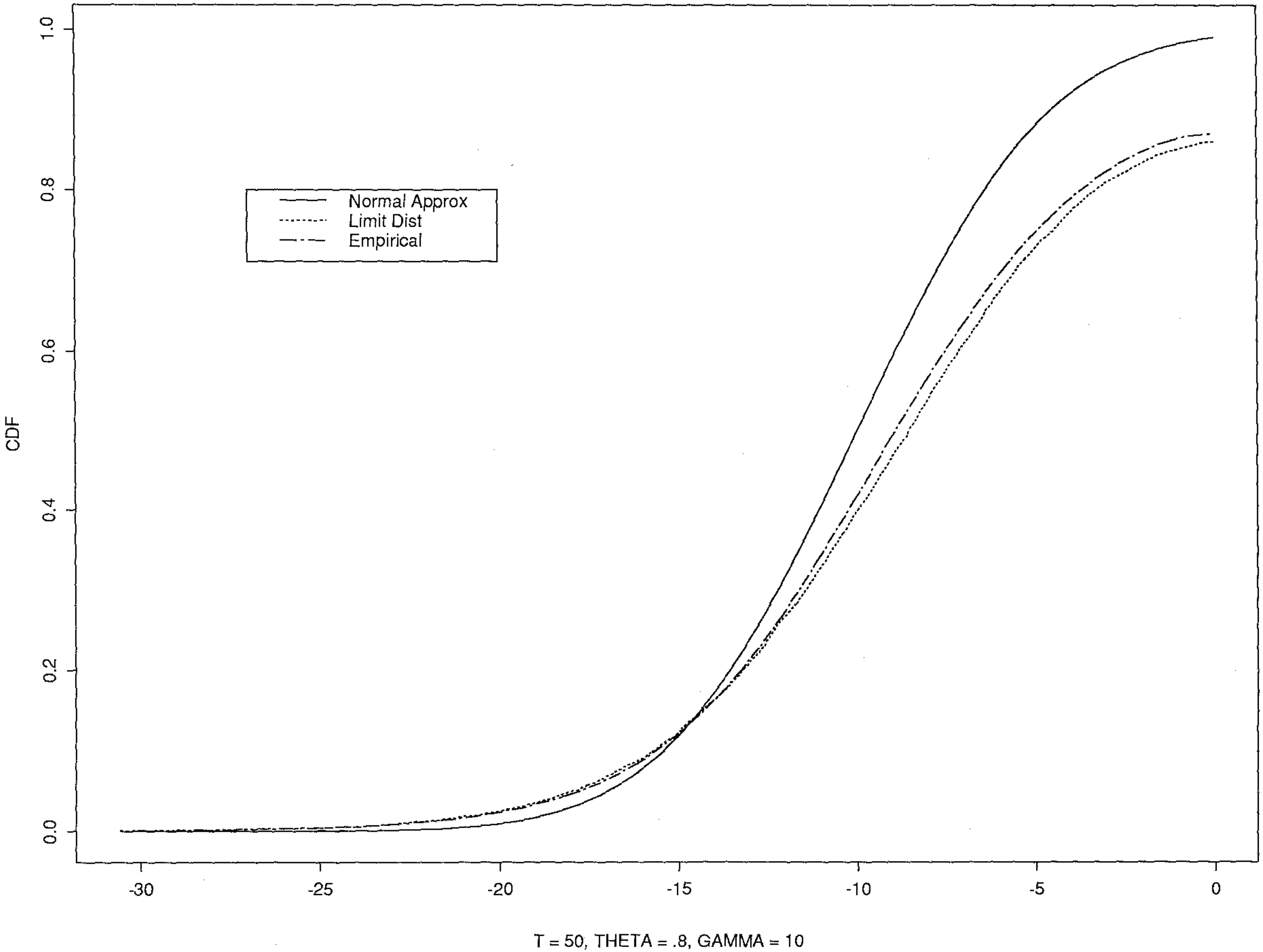

Figure 4: Comparison of Normal and Limit CDF's to Empirical

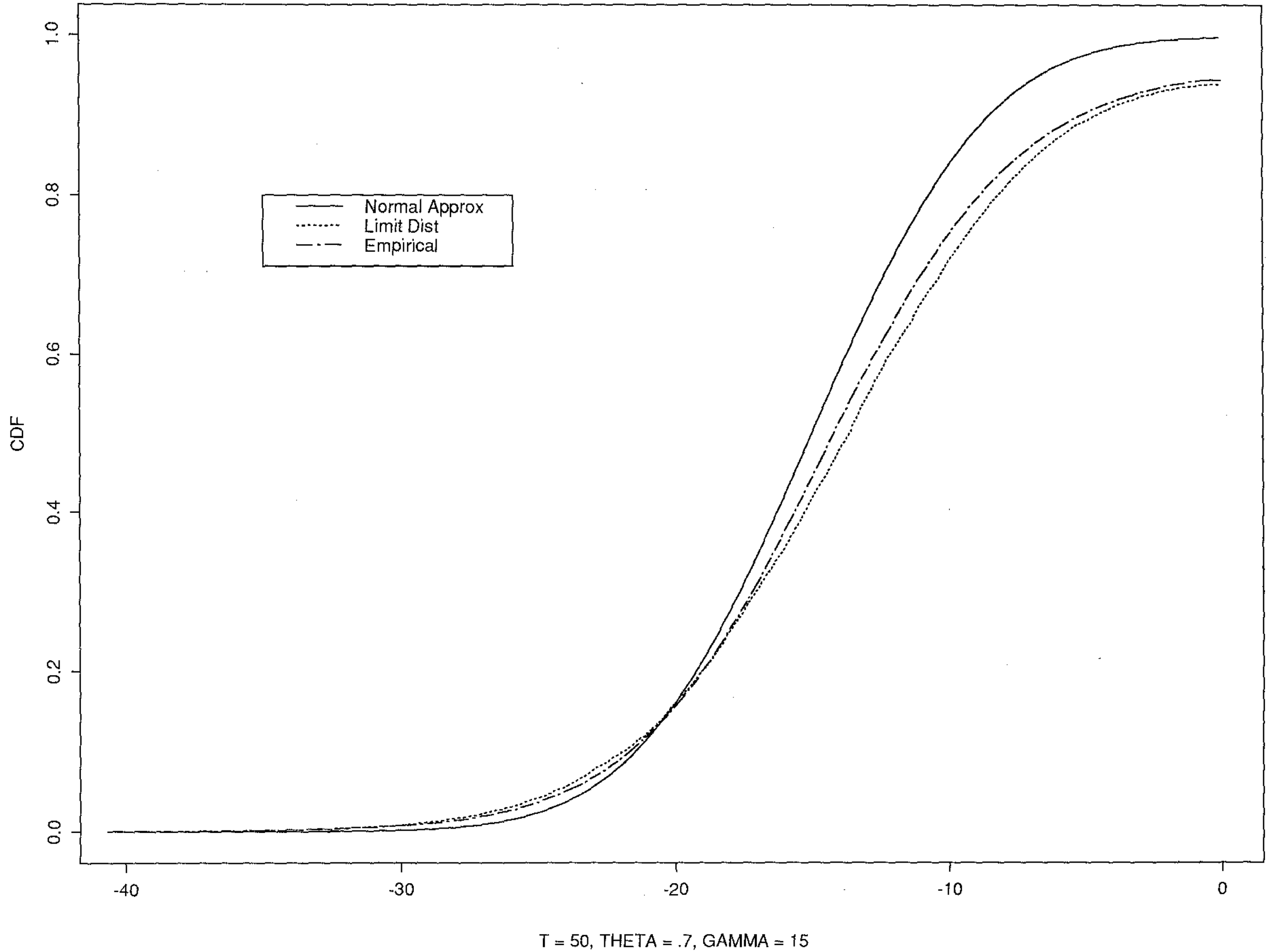

Figure 5: Comparison of Normal and Limit CDF's to Empirical