The Annals of Statistics
1999, Vol. 27, No. 1, 237–261

ON RESIDUAL EMPIRICAL PROCESSES OF STOCHASTIC REGRESSION MODELS WITH APPLICATIONS TO TIME SERIES

BY SANGYEOL LEE AND CHING-ZONG WEI

Seoul National University and Academia Sinica

Motivated by Gaussian tests for a time series, we are led to investigate the asymptotic behavior of the residual empirical processes of stochastic regression models. These models cover the fixed design regression models as well as general AR(q) models. Since the number of the regression coefficients is allowed to grow as the sample size increases, the obtained results are also applicable to nonlinear regression and stationary AR(∞) models. In this paper, we first derive an oscillation-like result for the residual empirical process. Then, we apply this result to autoregressive time series. In particular, for a stationary AR(∞) process, we are able to determine the order of the number of coefficients of a fitted AR(q_n) model and obtain the limiting Gaussian processes. For an unstable AR(q) process, we show that if the characteristic polynomial has a unit root 1, then the limiting process is no longer Gaussian. For the explosive case, one of our side results also provides a short proof for the Brownian bridge results given by Koul and Levental.

1. Introduction. Many of the statistical techniques available for the analysis of stationary time series have been designed for Gaussian processes, since the Gaussian assumption usually leads to a more tractable situation. For example, if $\{X_t\}$ is a Gaussian process, and $\{X_1, \ldots, X_n\}$ is observed, the best predictor $\hat{X}_{n+h}$ of X_{n+h}, $h \geq 1$, is a linear combination of $X_1, \ldots, X_n$. This, in general, is not true for non-Gaussian processes. However, in some applications, the process is expected to be non-Gaussian. The deconvolution problem arising in geographical analysis such as seismic exploration [cf. Lii and Rosenblatt (1982) and Rosenblatt (1985), pages 45 and 206] is an example. Thus, it is important to know whether a time series is Gaussian or not.

For the Gaussian test of a linear process, Subba Rao and Gabr (1980) and Hinich (1982) considered the frequency domain approach. Their method is based on the property that the bispectral density of a Gaussian process is identically zero. However, their tests do not perform well under symmetric alternatives since in these cases, the bispectral densities are also zero. For a related work, see Epps (1987).

For the time domain approach, the idea is to fit the time series by an AR(q) process through least squares methods. The residuals $\hat{\varepsilon}_t$ are then used to construct a related empirical process, which in turn is employed to form a test

Received March 1996; revised September 1998.

AMS 1991 *subject classifications*. Primary 60F17; secondary 62M10.

Key words and phrases. Gaussian tests, residual empirical process, stochastic regression model, stationary AR(∞) process, oscillation-like result, unstable AR(q) process, explosive process, Brownian bridge.

statistic, such as the Kolmogorov–Smirnov statistic. Boldin (1982) and Pierce (1985) considered the stationary AR(q) model,

$$X_t = \beta_1 X_{t-1} + \cdots + \beta_q X_{t-q} + \varepsilon_t, \qquad \{\varepsilon_t\} \sim \mathrm{iid}(0, \sigma^2),$$

and the process

$$\hat{Y}_n(u) = n^{-1/2} \sum_{t=1}^{n} \big[I(G(\hat{\varepsilon}_t) \le u) - u\big], \qquad u \in [0, 1], \tag{1.1}$$

where G is a true underlying distribution of $\{\varepsilon_t\}$ and $\{\hat{\varepsilon}_t\}$ are residuals. They showed that $\hat{Y}_n$ converges weakly to a Brownian bridge in $D[0, 1]$. For a general review of the residual empirical process in stationary AR(q) models, see Koul (1992), Chapter 7. Kreiss (1988) extended Boldin's result to the stationary AR(∞) process by fitting a long AR(q_n) process, where the order q_n depends on the sample size n.

In this paper, we consider the following stochastic regression model:

$$y_{nt} = \boldsymbol{\beta}_n' \mathbf{x}_{nt} + r_{nt} + \varepsilon_{nt}, \qquad 1 \le t \le n, \tag{1.2}$$

where $\boldsymbol{\beta}_n$ are $q_n \times 1$ unknown parameters, $\mathbf{x}_{nt}$ are observable $q_n \times 1$ random vectors and r_{nt} are random variables which may be viewed as "model bias." The setting includes classical fixed design regression, nonlinear regression, stationary AR(∞) and general AR(q) models. For example, if $q_n = q$, $r_{nt} = 0$, $\varepsilon_{nt} = \varepsilon_t$, $\boldsymbol{\beta}_n = \boldsymbol{\beta}$, $y_{nt} = X_t$ and $\mathbf{x}_{nt} = (X_{t-1}, \ldots, X_{t-q+1})'$, the model in (1.2) is an AR(q) model. If $\varepsilon_{nt} = \varepsilon_t$, $y_{nt} = X_t$, $\boldsymbol{\beta}_n = (\beta_1, \ldots, \beta_{q_n})'$, $\mathbf{x}_{nt} = (X_{t-1}, \ldots, X_{t-q_n+1})'$ and $r_{nt} = \sum_{j=q_n+1}^{\infty} \beta_j X_{t-j}$, then the model is an AR(∞) model. For discussion of other models which can be covered by (1.2), see Wei (1992). Note that the double array setting for the error terms is to allow the obtained results to be used to calculate the local power of a test. In this case, a sequence of contiguous alternatives is considered and the error terms ε depend on n as well.

Let $\tilde{\boldsymbol{\beta}}_n$ denote the least squares estimate of $\boldsymbol{\beta}_n$ in (1.2) and $\tilde{\varepsilon}_{nt} = y_{nt} - \tilde{\boldsymbol{\beta}}_n' x_{nt}$ the tth residual. The corresponding residual empirical process is defined by

$$\tilde{Y}_n(u) = n^{-1/2} \sum_{t=1}^{n} \big[I(H_n(\tilde{\varepsilon}_{nt}) \le u) - u\big],$$

where H_n is the underlying distribution of $\{\varepsilon_{nt}\}$. The asymptotic behavior of any test based on $\tilde{Y}_n$ is determined by the asymptotic distribution of $\tilde{Y}_n$. One may expect this distribution is close to the empirical process based on true errors, that is,

$$Y_n(u) = n^{-1/2} \sum_{t=1}^{n} \big[I(H_n(\varepsilon_{nt}) \le u) - u\big].$$

In Section 2, we find that an additional term $n^{-1/2}\sum_{t=1}^{n}[H_n(x+(\tilde{\boldsymbol\beta}_n-\boldsymbol\beta_n)'\mathbf{x}_{nt})-H_n(x)]$ is needed. More precisely, we have the oscillation-like result

$$\begin{aligned}\sup_x\bigg|n^{-1/2}\sum_{t=1}^{n}\big[I(\tilde{\varepsilon}_{nt}\le x)-H_n\big(x+(\tilde{\boldsymbol\beta}_n-\boldsymbol\beta_n)'\mathbf{x}_{nt}\big)&\\ +H_n(x)-I(\varepsilon_{nt}\le x)\big]\bigg|\to_P 0.&\end{aligned}\tag{1.3}$$

In fact, a result (Theorem 2.2) which is more general than (1.3) is established with the aid of martingale tools. This result is the key tool for our further analysis.

In Section 3, we concentrate on the applications. For the fixed design case and related issues, one can see Durbin (1973) and Shorack and Wellner (1985), page 708. We will focus on autoregressive processes. More specifically, in Section 3.1, the limiting process of $\tilde{Y}_n$ of (3.10) under a sequence of contiguous alternatives is shown to be a Gaussian process with a drift. The result (Theorem 3.1) not only helps us to determine the order of the fitted AR(q_n) process but also provides us a mean in the study of the local power of tests. In Section 3.2, we investigate the case for an unstable AR(q) process. The limiting process of $\hat{Y}_n$ of (3.26) is the standard Brownian bridge unless there is a unit root 1. The unit root case is of importance since it covers the IAR models. Our result (Theorem 3.2) shows that in this case, the limiting process is no longer Gaussian. Finally, we consider the explosive AR(1) case. The Brownian bridge result given by Koul and Levental (1989) is reestablished by a short proof (see Remark 3.1).

2. Main results. Suppose that $(\Omega,\mathscr{F},P)$ is a probability space, $\{\mathscr{F}_{ni};0\le i\le n\}$ is a double array of sub-σ-fields of $\mathscr{F}$ such that $\mathscr{F}_{ni}\subset\mathscr{F}_{n,i+1}$, $i=0,\dots,n-1$ and $q:=q_n$ is a sequence of positive integers. Let us consider the stochastic regression model

$$y_{nt}=\boldsymbol\beta_n'\mathbf{x}_{nt}+r_{nt}+\varepsilon_{nt},\qquad t=1,\dots,n,\tag{2.1}$$

where $\boldsymbol\beta_n$ are unknown $q\times 1$ vectors, $\mathbf{x}_{nt}$ are observable $q\times 1$ random vectors that are $\mathscr{F}_{n,t-1}$-measurable and r_{nt} are random variables not necessarily observable. The errors $\{\varepsilon_{nt};1\le t\le n\}$ are iid random variables with common distribution H_n, which has zero mean and finite variance. Also assume that ε_{nt} is $\mathscr{F}_{nt}$-measurable and independent of $\mathscr{F}_{n,t-1}$.

Given the observations $(\mathbf{x}_{n1},y_1),\dots,(\mathbf{x}_{nn},y_n)$, the least squares estimate of $\boldsymbol\beta_n$ is given by

$$\tilde{\boldsymbol\beta}_n=\left(\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}'\right)^{-1}\sum_{t=1}^{n}\mathbf{x}_{nt}y_{nt}.\tag{2.2}$$

In view of (2.1), $\tilde{\boldsymbol\beta}_n=\boldsymbol\beta_n+\boldsymbol\gamma_{n1}+\boldsymbol\gamma_{n2}$, where

$$\boldsymbol\gamma_{n1}=\left(\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}'\right)^{-1}\sum_{t=1}^{n}\mathbf{x}_{nt}r_{nt}\tag{2.3}$$

and

$$\boldsymbol{\gamma}_{n2} = \left(\sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}' \right)^{-1} \sum_{t=1}^{n} \mathbf{x}_{nt}\varepsilon_{nt}. \tag{2.4}$$

Note that if we set $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_n + \boldsymbol{\gamma}_{n2}$, then $\hat{\beta}_n$ is the least squares estimate when $r_{nt} = 0$ for all n, t.

In the following, we would like to obtain the limiting process of the residual empirical process

$$\tilde{E}_n(x) = n^{-1/2} \sum_{t=1}^{n} \big[I(\tilde{\varepsilon}_{nt} \le x) - H_n(x) \big], \qquad x \in R, \tag{2.5}$$

where $\tilde{\varepsilon}_{nt} = y_{nt} - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_{nt}$, $t = 1, \ldots, n$. To tackle this problem, we first introduce the residual empirical process $\hat{E}_n$ when the model bias is zero. Precisely,

$$\hat{E}_n(x) = n^{-1/2} \sum_{t=1}^{n} \big[I(\hat{\varepsilon}_{nt} \le x) - H_n(x) \big], \qquad x \in R, \tag{2.6}$$

where $\hat{\varepsilon}_{nt} = y_{nt} - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_{nt}$. Our approach is to reduce the limiting distribution problem of $\tilde{E}_n$ to that of $\hat{E}_n$ and then investigate the conditions for the limiting distribution of $\tilde{E}_n$ to exist. We start with two lemmas, which do not require the validity of (2.1).

LEMMA 2.1. *For each n, let $R_n = \max_{1 \le t \le n} |r_{nt}|$ and let $\boldsymbol{\gamma}_{n1}$ be the random variable in* (2.3). *Then*,

$$\max_{1 \le t \le n} \left| \boldsymbol{\gamma}_{n1}' \mathbf{x}_{nt} \right| \le (nq)^{1/2} R_n \max_{1 \le t \le n} \left\{ \mathbf{x}_{nt}' \left(\sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}' \right)^{-1} \mathbf{x}_{nt} \right\}^{1/2}.$$

PROOF. Put $A_n = \sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}'$. By the Schwarz inequality,

$$\begin{aligned}
\left| \boldsymbol{\gamma}_{n1}' \mathbf{x}_{nt} \right| &\le \left\| \mathbf{x}_{nt}' A_n^{-1/2} \right\| \left\| A_n^{-1/2} \sum_{t=1}^{n} \mathbf{x}_{nt} r_{nt} \right\| \\
&\le n^{1/2} R_n \left(\mathbf{x}_{nt}' A_n^{-1} \mathbf{x}_{nt} \right)^{1/2} \left(\sum_{t=1}^{n} \left\| A_n^{-1/2} \mathbf{x}_{nt} \right\|^2 \right)^{1/2} \\
&= (nq)^{1/2} R_n \left(\mathbf{x}_{nt}' A_n^{-1} \mathbf{x}_{nt} \right)^{1/2}.
\end{aligned}$$

This completes the proof. □

LEMMA 2.2. *Suppose that for each n, ε_{nt}, $t = 1, \ldots, n$, are iid random variables with common distribution H_n, and ε_{nt}^* are the random variables* *satisfying*

$$\begin{aligned}\alpha_n := \sup_x \Big| n^{-1/2} \sum_{t=1}^{n} \big[I(\varepsilon_{nt}^* \le x) - H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^*) \\ + H_n(x) - I(\varepsilon_{nt} \le x) \big] \Big| \\ = o_P(1).\end{aligned} \tag{2.7}$$

Let $\{\eta_n\}$ *be a sequence of nonnegative random variables with* $n^{1/2}\eta_n = o_P(1)$. *If* $\sup_{n,x} |H_n'(x)| < \infty$, *then*

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in \mathscr{C}_n} \sup_x \Big| n^{-1/2} \sum_{t=1}^{n} \big[I(\varepsilon_{nt}^* \le x + \theta_t) - H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^* + \theta_t) \\ + H_n(x) - I(\varepsilon_{nt} \le x) \big] \Big| \\ = o_P(1),\end{aligned}$$

where $\mathscr{C}_n = \{\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)' \in R^n; \max_{1 \le t \le n} |\theta_t| \le \eta_n\}$.

PROOF. For $x \in R$ and $\boldsymbol{\theta} \in R^n$, set

$$\begin{aligned}S_n(\boldsymbol{\theta}, x) = \sum_{t=1}^{n} \big[I(\varepsilon_{nt}^* \le x + \theta_t) - H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^* + \theta_t) \\ + H_n(x) - I(\varepsilon_{nt} \le x) \big].\end{aligned}$$

Since the mapping $y \to H_n(y)$ and $y \to I(\varepsilon_{nt}^* \le y)$ are nondecreasing, $S_n(\boldsymbol{\theta}, x)$ has an upper bound

$$\begin{aligned}&\sum_{t=1}^{n} \big[I(\varepsilon_{nt}^* \le x + \eta_n) - H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^* + \eta_n) + H_n(x + \eta_n) - I(\varepsilon_{nt} \le x + \eta_n) \big] \\ &\quad + \sum_{t=1}^{n} \big[H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^* + \eta_n) H_n(x + \varepsilon_{nt} - \varepsilon_{nt}^* + \theta_t) \big] \\ &\quad + \sum_{t=1}^{n} \big[I(\varepsilon_{nt} \le x + \eta_n) - H_n(x + \eta_n) + H_n(x) - I(\varepsilon_{nt} \le x) \big]\end{aligned}$$

Similarly, $S_n(\boldsymbol{\theta}, x)$ has a lower bound that is the same as above with η_n replaced by $-\eta_n$. Thus, from (2.7) and Taylor's series expansion we have that

$$\sup_{\boldsymbol{\theta} \in \mathscr{C}_n} \sup_x \big| n^{-1/2} S_n(\boldsymbol{\theta}, x) \big| \le \alpha_n + n^{1/2} \eta_n \sup_{n,x} |H_{n,x}'(x)| + \omega_n(\eta_n),$$

where

$$\omega_n(a) = \sup_{x\in R,\,|y|\le a}\left|\sum_{t=1}^{n}\left[I(\varepsilon_{nt} \le x + y) - H_n(x+y) + H_n(x) - I(\varepsilon_{nt} \le x)\right]\right|. \tag{2.8}$$

Since $\omega_n(a) \to 0$ in probability as $a \to 0$ [cf. Stute (1982)], the lemma is established by our assumptions. □

REMARK 2.1. When $\varepsilon_{nt}^* = \varepsilon_{nt}$, (2.7) holds automatically. From Lemma 2.2, we have that

$$\sup_{\boldsymbol{\theta}\in\mathscr{C}_n}\sup_x\left|n^{-1/2}\sum_{t=1}^{n}\left[I(\varepsilon_{nt} \le x + \theta_t) - H_n(x+\theta_t) + H_n(x) - I(\varepsilon_{nt} \le x)\right]\right| = o_P(1).$$

This perturbation result is useful in establishing the result given by Koul and Levental (1989). For the details, see Remark 3.1.

The following theorem provides us conditions under which $\tilde{E}_n$ and $\hat{E}_n$ have the same limiting distribution.

THEOREM 2.1. *Recall the definition of R_n in Lemma* 2.1. *Suppose that* (2.1) *and the conditions given in Lemma* 2.2 *hold*. *If*:

(a) $\sup_{n,x}|H_n'(x)| < \infty$;
(b) $R_n = o_P(n^{-1/2})$;
(c) $(nq)^{1/2}R_n \max_{1\le t\le n}\{\mathbf{x}_{nt}'(\sum_{t=1}^n \mathbf{x}_{nt}\mathbf{x}_{nt}')^{-1}\mathbf{x}_{nt}\}^{1/2} = o_P(1)$;
(d) $\sup_x|n^{-1/2}\sum_{t=1}^n[I(\hat{\varepsilon}_{nt} \le x) - H_n(x + \boldsymbol{\gamma}_{n2}'\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x)]| = o_P(1)$,

then $\sup_x|\tilde{E}_n(x) - \hat{E}_n(x)| = o_P(1)$.

PROOF. Note that $\tilde{\varepsilon}_{nt} = \hat{\varepsilon}_{nt} - \boldsymbol{\gamma}_{n1}'\mathbf{x}_{nt}$ and

$$\begin{aligned}
&\sup_x|\tilde{E}_n(x) - \hat{E}_n(x)| \\
&\quad\le \sup_x\left|n^{-1/2}\sum_{t=1}^{n}\left[I(\hat{\varepsilon}_{nt} \le x) - H_n(x + \varepsilon_{nt} - \hat{\varepsilon}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x)\right]\right| \\
&\qquad + \sup_x\left|n^{-1/2}\sum_{t=1}^{n}\left[I(\hat{\varepsilon}_{nt} \le x + \boldsymbol{\gamma}_{n1}'\mathbf{x}_{nt}) - H_n(x + \varepsilon_{nt} - \hat{\varepsilon}_{nt} + \boldsymbol{\gamma}_{n1}'\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x)\right]\right| \\
&\qquad + \sup_x\left|n^{-1/2}\sum_{t=1}^{n}\left[H_n(x + \varepsilon_{nt} - \hat{\varepsilon}_{nt} + \boldsymbol{\gamma}_{n1}'\mathbf{x}_{nt}) - H_n(x + \varepsilon_{nt} - \hat{\varepsilon}_{nt})\right]\right| \\
&\quad = I_{n1} + I_{n2} + I_{n3} \quad \text{(say)}.
\end{aligned}$$

Observe that

$$I_{n1} \le \sup_x \left| n^{-1/2} \sum_{t=1}^{n} \big[I(\hat{\varepsilon}_t \le x) - H_n(x + \boldsymbol{\gamma}'_{n2}\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x) \big] \right|$$
$$+ \sup_x \left| n^{-1/2} \sum_{t=1}^{n} \big[H_n(x + \varepsilon_{nt} - \hat{\varepsilon}_{nt}) - H_n(x + \boldsymbol{\gamma}'_{n2}\mathbf{x}_{nt}) \big] \right|.$$

The first term in the right-hand side of the above inequality converges in probability to zero by (d). Since $\varepsilon_{nt} - \hat{\varepsilon}_{nt} = \boldsymbol{\gamma}'_{n2}\mathbf{x}_{nt} - r_{nt}$, the second term is bounded by $n^{1/2} \sup_{n,x} |H'_n(x)| \max_{1\le t\le n} |r_{nt}|$, which also converges in probability to zero in view of (a) and (b). Therefore, $I_{n1} = o_P(1)$. This and (a) in turn imply that Lemma 2.2 is applicable with $\varepsilon^*_{nt} = \hat{\varepsilon}_{nt}$.

For I_{n2}, we first note that Lemma 2.1 and (c) give that $\max_{1\le t\le n} |\boldsymbol{\gamma}'_{n1}\mathbf{x}_{nt}| = o_P(n^{-1/2})$. This and Lemma 2.2 immediately prove that $I_{n2} = o_P(1)$. Together with (a), the order on $|\boldsymbol{\gamma}'_{n1}\mathbf{x}_{nt}|$ also shows that $I_{n3} = o_P(1)$. □

Now we are ready to study the limiting distribution of $\hat{E}_n$. Let $E_n(x) = n^{-1/2} \sum_{t=1}^{n} [I(\varepsilon_{nt} \le x) - H_n(x)]$, the empirical process based on errors. According to our analysis (cf. Section 3.2), the limiting distribution of $\hat{E}_n$ depends not only on E_n but also on the term $n^{-1/2} \sum_{t=1}^{n} [H(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{nt}) - H_n(x)]$. This leads us to consider the following oscillation-like result:

$$\sup_x \left| n^{-1/2} \sum_{t=1}^{n} \big[I(\hat{\varepsilon}_{nt} \le x) - H_n(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x) \big] \right| = o_P(1), \tag{2.9}$$

where the oscillation terms $(\hat{\boldsymbol{\beta}}_n - \beta_n)'\mathbf{x}_{nt}$ depend not only on n but also on t. In general, we may expect that there exists a sequence of nonstochastic matrices Λ_n, such that $\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n) = O_P(1)$. Let $\delta_{nt} = \varepsilon_{nt}$, $\boldsymbol{\alpha}'_n = -\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$ and $\mathbf{z}_{nj} = (\Lambda'_n)^{-1}\mathbf{x}_{nj}$. Then we reformulate (2.9) in a more general form as follows [cf. (2.20)].

THEOREM 2.2. *For each n, let $\{\mathscr{F}_{nj}; 1 \le j \le n\}$ be a family of nondecreasing sequences of sub-σ-fields of $\mathscr{F}$ and $\{\varepsilon_{nj}; 1 \le j \le n\}$ be iid random variables with the distribution H_n such that ε_{nj} is independent of $\mathscr{F}_{n,\,j-1}$. Suppose that δ_{nj}, $1 \le j \le n$ are the random variables such that*

$$\delta_{nj} = \varepsilon_{nj} + \boldsymbol{\alpha}'_n\mathbf{z}_{nj} + \rho_{nj}, \tag{2.10}$$

where $\boldsymbol{\alpha}_n$ and $\mathbf{z}_{nj}$ are $q = q_n$ dimensional random vectors, $\mathbf{z}_{nj}$ is $\mathscr{F}_{n,\,j-1}$-measurable and ρ_{nj} are random variables with $\max_{1\le j\le n} |\rho_{nj}| = o_P(n^{-1/2})$.

Further, assume that $\sup_{n,\,x} |H''_n(x)| < \infty$ and that there exists a distribution G with uniformly bounded second derivative such that

$$\sup_x \big| H'_n(x) - G'(x) \big| \to 0 \quad \text{as } n \to \infty, \tag{2.11}$$

and that for some positive real sequences b_n *and* c_n,

$$\|\boldsymbol{\alpha}_n\| = O_P(1), \tag{2.12}$$

$$n^{-1/2}\left\|\sum_{j=1}^{n} \mathbf{z}_{nj}\right\| = O_P(1), \tag{2.13}$$

$$n^{-1/2}\sum_{j=1}^{n} \|\mathbf{z}_{nj}\|^2 = o_P(1), \tag{2.14}$$

$$\sum_{j=1}^{n} \|\mathbf{z}_{nj}\| = O_P(b_n), \tag{2.15}$$

$$\max_{1\le j\le n} \|\mathbf{z}_{nj}\| = O_P(c_n). \tag{2.16}$$

Finally, assume that there exists a sequence $\{\zeta_n\}$ *of positive integers such that*

$$\lim_{n\to\infty} \zeta_n = \infty, \tag{2.17}$$

$$\lim_{n\to\infty} (nq)^{1/2} c_n/\zeta_n = 0, \tag{2.18}$$

$$\lim_{n\to\infty} \zeta_n^q n^{1/2} \exp\{-nB/(q^{1/2}b_n + n^{1/2})\} = 0 \quad \text{for all } B > 0. \tag{2.19}$$

Then, $\Gamma_n := \sup_x |\Gamma_n(x)| = o_P(1)$, *where*

$$\Gamma_n(x) = \left| n^{-1/2} \sum_{j=1}^{n} \left[I(\delta_{nj} \le x) - H_n(x - \boldsymbol{\alpha}_n' \mathbf{z}_{nj}) + H_n(x) - I(\varepsilon_{nj} \le x)\right]\right|. \tag{2.20}$$

PROOF. Put $\varepsilon_{nj}^* = \varepsilon_{nj} + \boldsymbol{\alpha}_n'\mathbf{z}_{nj}$. By Taylor's series expansion, we have that

$$\begin{aligned}
\Gamma_n &\le \sup_x \left| n^{-1/2} \sum_{j=1}^{n} \left[I(\delta_{nj} \le x) - H_n(x - \boldsymbol{\alpha}_n' \mathbf{z}_{nj} - \rho_{nj}) + H_n(x) - I(\varepsilon_{nj} \le x)\right]\right| \\
&\quad + n^{1/2} \max_{1\le j\le n} |\rho_{nj}| \sup_{n,\,x} |H_n'(x)| \\
&= \sup_x \left| n^{-1/2} \sum_{j=1}^{n} \left[I(\varepsilon_{nj}^* \le x - \rho_{nj}) - H_n(x + \varepsilon_{nj} - \varepsilon_{nj}^* - \rho_{nj}) \right.\right. \\
&\qquad\qquad \left.\left. + H_n(x) - I(\varepsilon_{nj} \le x)\right]\right| + o_P(1),
\end{aligned}$$

where the last equality follows from (2.11) and the assumption on ρ_{nj}. Hence, in view of Lemma 2.2, it suffices to show that

$$\begin{aligned}
&\sup_x \left| n^{-1/2} \sum_{j=1}^{n} \left[I(\varepsilon_{nj} \le x - \boldsymbol{\alpha}_n'\mathbf{z}_{nj}) - H_n(x - \boldsymbol{\alpha}_n'\mathbf{z}_{nj}) \right.\right. \\
&\qquad\qquad \left.\left. + H_n(x) - I(\varepsilon_{nj} \le x)\right]\right| = o_P(1).
\end{aligned} \tag{2.21}$$

Toward this end, fix $d > 0$ and partition the real line by the points $-\infty = x_{n0} < \cdots < x_{n,N_n} = \infty$ such that $N_n = [n^{1/2}d^{-1}]$ and $H_n(x_{ni}) = i/N_n$ for $i = 0, \dots, N_n$. Then we have for $x \in (x_{nr}, x_{n,r+1}]$,

$$|H_n(x_{ni}) - H_n(x)| \le n^{-1/2}d, \qquad i = r, r+1. \tag{2.22}$$

Furthermore, by (2.11) and the proposition in the Appendix, we have that for all sufficiently large n and $x \in (x_{nr}, x_{n,r+1}]$,

$$|H'_n(x_{ni}) - H'_n(x)| \le d, \qquad i = r, r+1. \tag{2.23}$$

Observe that for any $x \in (x_{nr}, x_{n,r+1}]$, $\Gamma_n(x)$ in (2.20) is bounded by $I_{n1}(x) + I_{n2}(x) + I_{n3}(x)$, where

$$\begin{aligned} I_{n1}(x) = \max_{i=r,r+1}\Bigg| n^{-1/2}\sum_{j=1}^{n}\big[& I(\varepsilon_{nj} \le x_{ni} - \boldsymbol{\alpha}'_n \mathbf{z}_{nj}) - H_n(x_{ni} - \boldsymbol{\alpha}'_n \mathbf{z}_{nj}), \\ & + H_n(x_{ni}) - I(\varepsilon_{nj} \le x_{ni})\big]\Bigg| \end{aligned}$$

$$I_{n2}(x) = \max_{i=r,r+1}\Bigg| n^{-1/2}\sum_{j=1}^{n}\big[H_n(x_{ni} - \boldsymbol{\alpha}'_n \mathbf{z}_{nj}) - H_n(x - \boldsymbol{\alpha}'_n \mathbf{z}_{nj})\big]\Bigg|,$$

$$I_{n3}(x) = \max_{i=r,r+1}\Bigg| n^{-1/2}\sum_{j=1}^{n}\big[I(\varepsilon_{nj} \le x_{ni}) - H_n(x_{ni}) + H_n(x) - I(\varepsilon_{nj} \le x)\big]\Bigg|.$$

By Taylor's series expansion, we have that

$$\begin{aligned} |I_{n2}(x)| \le\ & n^{1/2}\max_{i=r,r+1}|H_n(x_{ni}) - H_n(x)| \\ & + n^{-1/2}\|\boldsymbol{\alpha}_n\|\left\|\sum_{j=1}^{n}\mathbf{z}_{nj}\right\|\max_{i=r,r+1}|H'_n(x_{ni}) - H'_n(x)| \\ & + n^{-1/2}\|\boldsymbol{\alpha}_n\|^2\sum_{j=1}^{n}\|\mathbf{z}_{nj}\|^2\sup_{n,x}|H''_n(x)| \\ =\ & d + dO_P(1) + o_P(1), \end{aligned} \tag{2.24}$$

where the last inequality follows from (2.11), (2.13), (2.14), (2.22) and (2.23). This in turn implies that $\sup_x |I_{n2}(x)| = o_P(1)$ because d can be chosen arbitrarily small. Meanwhile, $\sup_x |I_{n3}(x)| = o_P(1)$ since $\sup_x |I_{n3}(x)| \le \omega_n(n^{-1/2}d) = o_P(1)$, where

$$\omega_n(a) = \sup_{\{0\le u,\,\nu\le 1;\,|u-\nu|\le a\}}\Bigg| n^{-1/2}\sum_{j=1}^{n}\big[I(H_n(\varepsilon_{nj}) \le u) - u + \nu - I(H_n(\varepsilon_{nj}) \le \nu)\big]\Bigg|$$

[cf. Stute (1982)]. Therefore, (2.21) follows once we verify

$$\begin{aligned}\Gamma_n^* := \max_{0\le r\le N_n}\Bigg| n^{-1/2}\sum_{j=1}^{n}\big[I(\varepsilon_{nj}\le x_{nr}-\boldsymbol{\alpha}_n'\mathbf{z}_{nj})\\ -H_n(x_{nr}-\boldsymbol{\alpha}_n'\mathbf{z}_{nj})+H_n(x_{nr})-I(\varepsilon_{nj}\le x_{nr})\big]\Bigg| = o_P(1).\end{aligned}\tag{2.25}$$

For any $\gamma > 0$, in view of (2.12), (2.15) and (2.16), there exists $K > 0$ such that $P(\bigcup_{i=1}^3 S_i^c) < \gamma$ for all sufficiently large n, where

$$S_1 = \{\|\boldsymbol{\alpha}_n\| \le K\},$$

$$S_2 = \left\{\sum_{j=1}^{n}\|\mathbf{z}_{nj}\| \le K b_n\right\},$$

$$S_3 = \left\{\max_{1\le j\le n}\|\mathbf{z}_{nj}\| \le K c_n\right\}.$$

Then for $\lambda > 0$,

$$\begin{aligned}P(\Gamma_n^* > \lambda) &\le P(\Gamma_n^* > \lambda, \cap_{i=1}^3 S_i) + \gamma\\ &\le P\left(\max_{0\le r\le N_n,\,\mathbf{y}\in\mathscr{U}}\left| n^{-1/2}\sum_{j=1}^{n} d_j(x_{nr},\mathbf{y})\right| > \lambda,\, S_2\cap S_3\right) + \gamma,\end{aligned}$$

where $\mathscr{U} = \{\mathbf{y}\in R^q; \|\mathbf{y}\| \le K\}$ and for $x\in R$, $\mathbf{y}\in\mathscr{U}$,

$$d_j(x,\mathbf{y}) = I(\varepsilon_{nj}\le x+\mathbf{y}'\mathbf{z}_{nj}) - H_n(x+\mathbf{y}'\mathbf{z}_{nj}) + H_n(x) - I(\varepsilon_{nj}\le x).$$

In order to verify (2.25), it suffices to show that

$$\sup_{0\le r\le N_n,\,\mathbf{y}\in\mathscr{U}}\left| n^{-1/2}\sum_{j=1}^{n} d_j(x_{nr},\mathbf{y})\right| I(S_2\cap S_3) = o_P(1),\tag{2.26}$$

since γ can be taken arbitrarily small.

Partition the rectangle $[-K, K]^q$ in R^q by subrectangles generated by the lattices $V_q = \{(y_{1j_1},\dots,y_{qj_q});\ 0\le j_1,\dots,j_q\le \zeta_n\}$, where $y_{ij} = -K + 2Kj/\zeta_n$, $i = 1,\dots,q$, $j = 0,\dots\zeta_n$. Let $\mathscr{C}$ denote the class of all the subrectangles C such that $C\cap\mathscr{U}\ne\varnothing$. Note that the cardinal number k_n of $\mathscr{C}$ is at most ζ_n^q. Let $\mathscr{C} = \{C_s; s = 1,\dots,k_n\}$. For $\mathbf{y}\in\mathscr{C}_s$, let $w_{js}^+ = \sup_{\mathbf{y}\in\mathscr{C}_s}\mathbf{y}'\mathbf{z}_{nj}$ and $w_{js}^- = \inf_{\mathbf{y}\in\mathscr{C}_s}\mathbf{y}'\mathbf{z}_{nj}$. Note that both w_{js}^+ and w_{js}^- are $\mathscr{F}_{n,\,j-1}$ measurable. Now, for $\mathbf{y}\in\mathscr{C}_s$, we have that $L_j(x,\mathbf{y}) \le d_j(x,\mathbf{y}) \le U_j(x,\mathbf{y})$, where

$$\begin{aligned}U_j(x,\mathbf{y}) = &\{I(\varepsilon_{nj}\le x+w_{js}^+) - H_n(x+w_{js}^+) + H_n(x) - I(\varepsilon_{nj}\le x)\}\\ &+ \{H_n(x+w_{js}^+) - H_n(x+\mathbf{y}'\mathbf{z}_{nj})\},\end{aligned}$$

and $L_j(x,\mathbf{y})$ is the same as $U_j(x,\mathbf{y})$ with w_{js}^+ replaced by w_{js}^-. On $S_2\cap S_3$,

$$|H_n(x+w_{js}^+) - H_n(x+\mathbf{y}'\mathbf{z}_{nj})| \le \sup_{n,\,x}|H_n'(x)|\,|w_{js}^+ - \mathbf{y}'\mathbf{z}_{nj}| = O(c_n q^{1/2}/\zeta_n),$$

which is $o(n^{-1/2})$ by (2.11) and (2.18). Similarly, on $S_2 \cap S_3$,

$$\left|H_n\left(x + w_{js}^-\right) - H_n\left(x + y' z_{nj}\right)\right| = o(n^{-1/2}).$$

Therefore, the left term in (2.26) is bounded by $II_{n1} + II_{n2} + o(1)$, where

$$II_{n1} = \max_{1 \le s \le k_n} \max_{0 \le r \le N_n} \left| n^{-1/2} \sum_{j=1}^{n} e_j(x_{nr}, w_{js}^+) \right| I(S_2 \cap S_3),$$

$$II_{n2} = \max_{1 \le s \le k_n} \max_{0 \le r \le N_n} \left| n^{-1/2} \sum_{j=1}^{n} e_j(x_{nr}, w_{js}^-) \right| I(S_2 \cap S_3)$$

and

$$e_j(x, y) = I(\varepsilon_{nj} \le x + y) - H_n(x + y) + H_n(x) - I(\varepsilon_{nj} \le x).$$

In the following, we only provide a proof for $II_{n1} = o_P(1)$ since II_{n2} can be handled similarly.

Define

$$\tilde{II}_{n1} = \max_{1 \le s \le k_n} \max_{0 \le r \le N_n} \left| n^{-1/2} \sum_{j=1}^{n} \tilde{e}_j(x_{nr}, w_{js}^+) \right|,$$

where

$$\tilde{e}_j := \tilde{e}_j(x_{nr}, w_{js}^+) = e_j(x_{nr}, w_{js}^+) I\left(\sum_{i=1}^{j} \|\mathbf{z}_{ni}\| \le K b_n \right).$$

Note that $\{\tilde{e}_j, \mathscr{F}_{nj}\}$ is a martingale difference sequence with $|\tilde{e}_j| \le 1$ a.s. for all j. Further,

$$\begin{aligned} E\left(\tilde{e}_j^2 \mid \mathscr{F}_{n,\, j-1}\right) &\le \left| H_n(x_{nr} + w_{js}^+) - H_n(x_{nr}) \right| \left| I\left(\sum_{i=1}^{j} \|\mathbf{z}_{ni}\| \le K b_n \right) \right. \\ &\le \sup_{n,\, x} |H_n'(x)|\, |w_{js}^+| I\left(\sum_{i=1}^{j} \|\mathbf{z}_{ni}\| \le K b_n \right). \end{aligned}$$

Consequently,

$$\sum_{j=1}^{n} E\left(\tilde{e}_j^2 \mid \mathscr{F}_{n,\, j-1}\right) \le \theta q^{1/2} b_n,$$

where $\theta = K^2 \sup_{n,\, x} |H_n'(x)|$. By Bernstein's inequality for martingales [cf. Shorack and Wellner (1986), page 809], we have that for all $\lambda > 0$,

$$\begin{aligned} P(\tilde{II}_{n1} > \lambda) &\le 2 k_n (N_n + 1) \exp\left\{ -n\lambda^2 / 2(\theta q^{1/2} b_n + n^{1/2} \lambda / 3) \right\} \\ &= O\left(\zeta_n^q n^{1/2} \exp\left\{ -nB / (q^{1/2} b_n + n^{1/2}) \right\} \right), \qquad B > 0, \end{aligned}$$

which goes to 0 as $n \to \infty$ due to (2.19). Here, note that

$$P(\tilde{e}_j \neq e_j(x_{nr}, w_{nj}^+) \quad \text{for some } j \leq n \text{ on } S_2 \cap S_3)$$
$$\leq P\left(\sum_{j=1}^{n} \|\mathbf{z}_{nj}\| > Kb_n \quad \text{on } S_2 \cap S_3\right) = 0.$$

Hence, $P(II_{n1} > \lambda) = P(\tilde{II}_{n1} > \lambda) \to 0$ as $n \to \infty$. This completes our proof. $\square$

REMARK 2.2. (a) When the contiguous alternatives $H_n = (1 - \gamma n^{-1/2})G + \gamma n^{-1/2}F$ are considered, then $\sup_x(|G''(x)| + |F''(x)|) < \infty$ would ensure (2.11).

(b) If condition (2.13) is replaced by $n^{-1/2}\|\sum_{j=1}^{n} \mathbf{z}_{nj}\| = o_P(1)$, then condition (2.11) can be replaced by $\sup_{n,x}|H_n'(x)| < \infty$ and $\sup_{n,x}|H_n''(x)| < \infty$. This is because in argument (2.24), $|I_{n2}(x)|$ is then bounded by $d + o_P(1) + o_P(1)$. The rest of the arguments are the same.

When the number of regression parameters are fixed and there is no model bias, the following corollary provides some easy-to-check conditions.

COROLLARY 2.1. *Assume that q is a fixed number and ρ_{ni} are equal to* 0. *If H_n satisfy* (2.11), *and if $\|\boldsymbol{\alpha}_n\| = O_P(1)$ and $\sum_{j=1}^{n} \|\mathbf{z}_{nj}\|^2 = O_P(1)$, then $\Gamma_n = o_P(1)$.*

PROOF. Note that by the Schwarz inequality, $\|\sum_{j=1}^{n} \mathbf{z}_{nj}\| \leq \sum_{j=1}^{n} \|\mathbf{z}_{nj}\| \leq n^{1/2}(\sum_{j=1}^{n} \|\mathbf{z}_{nj}\|^2)^{1/2}$. Thus (2.13) holds. Moreover,

$$\max_{1 \leq j \leq n} \|\mathbf{z}_{nj}\| \leq \left(\sum_{j=1}^{n} \|\mathbf{z}_{nj}\|^2\right)^{1/2} = O_P(1).$$

Now, applying Theorem 2.2 with $b_n = n^{1/2}$, $c_n = 1$ and $\zeta_n = n$, we obtain $\Gamma_n = o_P(1)$. $\square$

We now return to the original stochastic regression model (2.1).

COROLLARY 2.2. *Let $\boldsymbol{\gamma}_{n2}$ be the same in* (2.4) *and $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_n + \boldsymbol{\gamma}_{n2}$. Suppose that H_n satisfies* (2.11). *In addition, assume that there exists a sequence of $q \times q$ nonstochastic matrices Λ_n, such that*:

(i) $\|\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)\| = O_P(1)$;

(ii) $n^{-1/2}\|\sum_{j=1}^{n}(\Lambda_n')^{-1}\mathbf{x}_{nj}\| = O_P(1)$, $n^{-1/2}\sum_{j=1}^{n}\|(\Lambda_n')^{-1}\mathbf{x}_{nj}\|^2 = o_P(1)$, $\sum_{j=1}^{n}\|(\Lambda_n')^{-1}\mathbf{x}_{nj}\| = O_P(b_n)$ *and* $\max_{1 \leq j \leq n}\|(\Lambda_n')^{-1}\mathbf{x}_{nj}\| = O_P(c_n)$,

with some positive sequences b_n and c_n satisfying (2.15)–(2.19). *Then if* $\max_{1 \leq t \leq n}|r_{nt}| = o_P(n^{-1/2})$,

$$\sup_x \Bigg| n^{-1/2} \sum_{t=1}^{n} \big[I(\hat{\varepsilon}_{nt} \leq x) - H_n(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \leq x)\big]\Bigg| = o_P(1). \tag{2.27}$$

PROOF. Note that $\hat{\varepsilon}_{nt} = \varepsilon_{nt} - (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{n,t-1} + r_{nt}$. Let $\boldsymbol{\alpha}_n = -\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)$, $\mathbf{z}_{nj} = (\Lambda_n')^{-1}\mathbf{x}_{nj}$ and $\rho_{nj} = r_{nj}$. In view of Theorem 2.2, we obtain (2.27). □

COROLLARY 2.3. *Assume that q is a fixed number and r_{nt} in* (2.1) *are equal to* 0 . *Let $\hat{\boldsymbol{\beta}}_n$ be the least squares estimate of $\boldsymbol{\beta}_n$. If H_n satisfies* (2.11), *and if there exists a sequence of $q \times q$ nonstochastic matrices Λ_n, such that $\|\Lambda_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)\| = O_P(1)$ and $\sum_{j=1}^n \|(\Lambda_n')^{-1}\mathbf{x}_{nj}\|^2 = O_P(1)$, then*

$$\sup_x \left| n^{-1/2} \sum_{t=1}^n \big[I(\hat{\varepsilon}_{nt} \le x) - H_n(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{nt}) + H_n(x) - I(\varepsilon_{nt} \le x) \big] \right| = o_P(1).$$

The corollary is a direct result of Corollary 2.1.

3. Applications.

In this section, we apply the results in Section 2 to autoregressive processes. For a stationary AR(∞) process, the limiting process of the residual empirical process is shown to be a Gaussian process with a drift. For an unstable process, the limiting process is a standard Brownian bridge unless the characteristic polynomial has a unit root 1. A final remark gives a shorter proof for the explosive AR(1) results given by Koul and Levental (1989).

3.1. *Stationary AR (∞) time series models.* Let us consider a stationary AR(∞) process of the form

$$X_t - \sum_{\nu=1}^{\infty} \beta_\nu X_{t-\nu} = \varepsilon_t, \tag{3.1}$$

where $\{\varepsilon_t\}$ are iid random variables with mean zero and unknown variance $\sigma^2 > 0$, and the function $A(z) = 1 - \sum_{\nu=1}^{\infty} \beta_\nu z^\nu$ is analytic on an open neighborhood of the closed unit disk D in the complex plane and has no zeroes on D. The AR(∞) process in (3.1) has been given considerable attention in research areas such as econometrics and control theory. The related literature is quite extensive: see, for example, Shibata (1980).

Note that $\{X_t\}$ has the linear representation

$$X_t = \sum_{\nu=0}^{\infty} a_\nu \varepsilon_{t-\nu}, \tag{3.2}$$

where a_ν are uniquely determined by the relation $\theta(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu = A^{-1}(z)$ for $|z| \le 1$. Moreover, there exist constants $\rho \in (0,1)$ and $C > 0$, such that

$$\max\{|\beta_\nu|, |a_\nu|\} \le C\rho^\nu \quad \text{for all } \nu \ge 0 \tag{3.3}$$

[cf. Brockwell and Davis (1990), Proposition 3.2]. Since by the decomposition theorem in Cramér [(1974), page 213], $\{X_t\}$ is Gaussian if and only if $\{\varepsilon_t\}$ is Gaussian, the Gaussian test on $\{X_t\}$ is equivalent to that on $\{\varepsilon_t\}$. This allows us to set up the null and alternative hypotheses based on $\{\varepsilon_t\}$ rather than on $\{X_t\}$ themselves. The null and alternatives under consideration are as follows:

(3.4) $$K_0\colon \{\varepsilon_t\} \sim \Phi(\cdot/\sigma),$$

(3.5) $$K_n\colon \{\varepsilon_t\} \sim H_n(\cdot/\sigma) := (1-\gamma/n^{1/2})\Phi(\cdot/\sigma) + (\gamma/n^{1/2})H(\cdot/\sigma),$$

where Φ denotes the standard normal distribution and H is a distribution function with mean 0 and variance 1. Let $\{P_{n0}\}$ and $\{P_{n1}\}$ be the sequences of the joint distributions of $(X_1, \dots, X_n)$ under K_0 and $\{K_n\}$, respectively. They are contiguous in the sense of Le Cam [see, for the definition, Roussas (1972)].

Since the true errors are not observable, we fit the process by a long AR(q) model, where $q = q_n$ is a sequence of positive integers such that $q \to \infty$ and $q/n \to 0$ as $n \to \infty$. Rewrite the model and allow the error terms to depend on the chosen contiguous alternatives. Then we have

(3.6) $$x_{nt} = \sum_{\nu=1}^{q} \beta_\nu x_{n,\,t-\nu} + r_{nt} + \varepsilon_{nt},$$

where $r_{nt} = \sum_{\nu=q+1}^{\infty} \beta_\nu x_{n,\,t-\nu}$ is the model bias, and $\{x_{nt}\}$ and $\{\varepsilon_{nt}\}$ denote the random variables $\{X_t\}$ and $\{\varepsilon_t\}$ under K_n, respectively. Note that similarly to (3.2), $\{x_{nt}\}$ has the linear representation

(3.7) $$x_{nt} = \sum_{\nu=0}^{\infty} a_\nu \varepsilon_{n,\,t-\nu}.$$

Assume that $x_{n1}, \dots, x_{nn}$ are observed. The least squares estimate is

(3.8) $$\tilde{\boldsymbol{\beta}}_n = \left(\sum_{t=q+1}^{n} \mathbf{x}_{n,\,t-1}\mathbf{x}'_{n,\,t-1} \right)^{-1} \sum_{t=q+1}^{n} \mathbf{x}_{n,\,t-1} x_{nt},$$

where $\mathbf{x}_{nt} = (x_{nt}, \dots, x_{n,\,t-q+1})'$, and the residuals are

(3.9) $$\tilde{\varepsilon}_{nt} = x_{nt} - \tilde{\boldsymbol{\beta}}'_n \mathbf{x}_{n,\,t-1}, \qquad t = q+1, \dots, n.$$

Based on $\tilde{\varepsilon}_{nt}$, the residual empirical process is defined by

(3.10) $$\tilde{Y}_n(u) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \big[I(\Phi(\tilde{\varepsilon}_{nt}/\tilde{\sigma}_n) \le u) - u \big], \qquad u \in [0,1],$$

where

(3.11) $$\tilde{\sigma}_n^2 = (n-q)^{-1} \sum_{t=q+1}^{n} \big(x_{nt} - \tilde{\boldsymbol{\beta}}'_n \mathbf{x}_{n,\,t-1} \big)^2.$$

The following is the main result of Section 3.1.

THEOREM 3.1. *Suppose that q satisfies*

$$n^{-1/2}q^2 \log n \to 0 \quad \textit{and} \quad n^{7/4} q \rho^q \to 0 \textit{ for all } \rho \in (0,1) \textit{ as } n \to \infty, \tag{3.12}$$

and the distribution H in (3.5) *satisfies*

$$\sup_x |H'(x)| < \infty, \sup_x |H''(x)| < \infty \quad \textit{and} \quad \int x^4 dH(x) < \infty. \tag{3.13}$$

Then, under $\{K_n\}$, $\tilde{Y}_n$ converges weakly to a Gaussian process Y, with

$$\text{(i)} \qquad EY(u) = -\gamma(u - H \circ \Phi^{-1}(u)),$$

$$\text{(ii)} \quad \mathrm{Cov}(Y(u), Y(v)) = u \wedge v - uv \\ -2^{-1}\phi(\Phi^{-1}(u))\Phi^{-1}(u)\phi(\Phi^{-1}(v))\Phi^{-1}(v), \tag{3.14}$$

for all $0 \le u, v \le 1$. Here, ϕ denotes the density of Φ.

REMARK 3.1. A typical example of such q satisfying (3.12) is $(\log n)^2$. Kreiss (1988) studied the limiting process of

$$Y_n^*(u) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \left[I(G(\tilde{\varepsilon}_{nt}) \le u) - u \right]$$

under the null H_0: $\varepsilon_t \sim G$. Here, G is a distribution function such that $\sup_x |G''(x)| < \infty$ and $\int x^4 \, dG(x) < \infty$. He showed that Y_n^* converges weakly to a Brownian bridge under the conditions that $n^{-1/2}q^2(\log n)^2 \to 0$ and $n^3 \rho^q \to 0$ as $n \to \infty$ for all $\rho \in (0,1)$. Although in Theorem 3.1, we only considered the Gaussian distribution case, the result can be easily extended to the general case. According to our analysis (cf. Lemma 3.4), Y_n^* converges weakly to a standard Brownian bridge if $\sup_x |G''(x)| < \infty$, $\int x^4 \, dG(x) < \infty$ and q satisfies (3.12). Apparently, conditions (3.12) are weaker than those of Kreiss.

In order to establish Theorem 3.1, consider the process

$$\tilde{E}_n(x) = (n-q)^{-1/2} \sum_{t=1}^{n} \left[I(\tilde{\varepsilon}_{nt}/\tilde{\sigma}_n \le x) - \Phi(x) \right], \qquad x \in R, \tag{3.15}$$

and split it into $A_n(x) + B_n(x) + C_n(x) + D_n(x)$, where

$$A_n(x) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \left[I(\varepsilon_{nt}/\sigma \le x) - H_n(x) \right],$$

$$B_n(x) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \left[H_n(\tilde{\sigma}_n x/\sigma) - \Phi(x) \right],$$

$$C_n(x) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \left[I(\varepsilon_{nt}/\sigma \le \tilde{\sigma}_n x/\sigma) - H_n(\tilde{\sigma}_n x/\sigma) \right. \\ \left. + H_n(x) - I(\varepsilon_{nt}/\sigma \le x) \right],$$

$$D_n(x) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \left[I(\tilde{\varepsilon}_{nt} \le \tilde{\sigma}_n x) - I(\varepsilon_{nt} \le \tilde{\sigma}_n x) \right].$$

Throughout the sequel of this section, we denote $Z_n = O_P(a_n)$ if for any $\eta > 0$, there exists $M > 0$ such that $P_n(|Z_n| > Ma_n) < \eta$ for all n, where P_n denotes the probability measure under K_n. Also, we denote $Z_n = o_P(a_n)$ if for any $\eta > 0$, $P_n(|Z_n/a_n| > \eta) \to 0$ as $n \to \infty$. For $q \times q$ matrix A, $\|A\| = \sup_x\{\|A\mathbf{x}\|; \|\mathbf{x}\| = 1\}$. To prove Theorem 3.1, we need to verify $\sup_x |D_n(x)| = o_P(1)$. The following series of lemmas is useful.

LEMMA 3.1. *Let Σ_n be the $q \times q$ matrix whose (μ, ν)th entry is $E^{(n)} x_{n\mu} x_{n\nu}$. If $\int x^4\, dH(x) < \infty$ and if $q^3/n \to 0$ as $n \to \infty$, then $\|\hat{\Sigma}_n^{-1}\| = O_P(1)$, where $\hat{\Sigma}_n = (n-q)^{-1} \sum_{t=q+1}^{n} \mathbf{x}_{n,t-1}\mathbf{x}'_{n,t-1}$.*

The lemma follows from Lemma 4 of Berk (1974) and Grenander and Szegö (1958), page 14.

LEMMA 3.2. *If $R_n = \max_{q+1 \le t \le n} |r_{nt}|$, then $R_n = O_P(n\rho^q)$, where ρ is the number in* (3.3). *Moreover, if q satisfies* (3.12), *then $R_n = o_P(n^{-1/2})$ and*

$$(nq)^{1/2} R_n \max_{q+1 \le t \le n} \left\{ \mathbf{x}'_{n,t-1} \left(\sum_{q+1}^{n} \mathbf{x}_{n,t-1}\mathbf{x}'_{n,t-1} \right)^{-1} \mathbf{x}_{n,t-1} \right\}^{1/2} = o_P(1). \tag{3.16}$$

PROOF. It is easy to see that $R_n = O_P(n\rho^q)$. Thus, if (3.12) holds, $R_n = o_P(n^{-1/2})$. The result in (3.16) is yielded by Lemma 3.1 and (3.12). □

LEMMA 3.3. *Let*

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_n + \left(\sum_{t=q+1}^{n} \mathbf{x}_{n,t-1}\mathbf{x}'_{n,t-1} \right)^{-1} \sum_{t=q+1}^{n} \mathbf{x}_{n,t-1}\varepsilon_{nt}.$$

If (3.12) *holds*, $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|^2 = O_P(n^{-1}q)$.

PROOF. The lemma is a direct result of Lemma 3.1 and the fact $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n\|^2 \le \|\hat{\Sigma}_n^{-1}\|^2 \|(n-q)^{-1} \sum_{t=q+1}^{n} \mathbf{x}_{n,t-1}\varepsilon_{nt}\|^2$. □

LEMMA 3.4. *Under the conditions of Theorem* 3.1, $\sup_x |D_n(x)| = o_P(1)$.

PROOF. First, observe that H_n satisfies (2.11) [cf. Remark 2.2 (a)] and $R_n = o_P(n^{-1/2})$ by Lemma 3.2. Let $\hat{\varepsilon}_{nt} = \varepsilon_{nt} - (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{n,t-1} + r_{nt}$, where $\hat{\boldsymbol{\beta}}_n$ is the random variable defined in Lemma 3.3. Define

$$\begin{aligned} D_n^*(x) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \big[& I(\hat{\varepsilon}_{nt} \le x) - H_n(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{n,t-1}) \\ & + H_n(x) - I(\varepsilon_{nt} \le x) \big]. \end{aligned}$$

Put $\Lambda_n = n^{1/2} q^{-1/2} I_{q \times q}$, $b_n = n^{1/2} q$, $c_n = 1$ and $\zeta_n = n$. By Lemma 3.1, we obtain (i) of Corollary 2.2. Applying Corollary 2.2 with the above b_n, c_n and

ζ_n, we have that $\sup_x |D_n^*(x)| = o_P(1)$. Hence, in view of Lemma 3.2 and Theorem 2.1, it suffices to show that $\sup_x |\hat{D}_n(x)| = o_P(1)$, where

$$\hat{D}_n(x) = D_n^*(x) + (n-q)^{-1/2} \sum_{t=q+1}^{n} \big[H_n(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n)'\mathbf{x}_{n,\,t-1}) - H_n(x)\big].$$

It can be readily seen from Lemmas 3.2 and 3.3 and Taylor's series expansion that the second term in the right-hand side of the above equality is $o_P(1)$. This completes the proof. □

Proof of Theorem 3.1. Recall that $\tilde{E}_n(x)$ in (3.15) is equal to $A_n(x) + B_n(x) + C_n(x) + D_n(x)$. Let $\tilde{\sigma}_n^2$ be the random variable defined in (3.11). By using Lemmas 3.2 and 3.3, we have that $\tilde{\sigma}_n^2 = (n-q)^{-1}\sum_{t=q+1}^{n} \varepsilon_{nt}^2 + \Delta_n$ with $\Delta_n = o_P(n^{-1/2})$. Therefore, $\sup_x |H_n(\tilde{\sigma}_n x/\sigma) - H_n(x)| = o_P(1)$ and thus $\sup_x |C_n(x)| = o_P(1)$ [cf. (2.8) and Stute (1982)]. Meanwhile, Taylor's series expansion yields that

$$B_n(x) = -\gamma(\Phi(x) - H(x)) + 2^{-1}x\phi(x)(n-q)^{-1/2} \sum_{t=q+1}^{n} (\varepsilon_{nt}^2/\sigma^2 - 1) + \eta_n(x),$$

where $\sup_x |\eta_n(x)| = o_P(1)$. Since $\sup_x |D_n(x)| = o_P(1)$ by Lemma 3.4, and $\tilde{Y}_n(u)$ in (3.10) is equal to $\tilde{E}(\Phi^{-1}(u))$, we can write that

$$\tilde{Y}_n(u) = -\gamma(u - G\circ\Phi^{-1}(u)) + Z_n(u) + \delta_n(u),$$

where $\sup_u |\delta_n(u)| = o_P(1)$ and

$$Z_n(u) = (n-q)^{-1/2} \sum_{t=q+1}^{n} \big[I(H_n(\varepsilon_{nt}) \le u) - u + 2^{-1}\Phi^{-1}(u)\phi(\Phi^{-1}(u))(\varepsilon_{nt}^2/\sigma^2 - 1)\big].$$

Here, one can readily check that Z_n converges weakly to a Gaussian process Z whose covariance structure is the same as of Y in (3.14). This establishes the theorem. □

3.2. *Unstable AR(q) processes.* In this subsection, we consider the time series model

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \varepsilon_t, \tag{3.17}$$

where ε_t are iid random variables with zero mean, finite variance σ^2 and common distribution H, such that

$$\sup_x |H'(x)| < \infty \quad \text{and} \quad \sup_x |H''(x)| < \infty. \tag{3.18}$$

We assume that the corresponding characteristic polynomial φ has a decomposition

$$\begin{aligned}\varphi(z) &= 1 - \beta_1 z - \cdots - \beta_q z^q \\ &= (1-z)^a(1+z)^b \prod_{k=1}^{l} \big(1 - 2\cos\theta_k z + z^2\big)^{d_k} \psi(z),\end{aligned}$$

where a, b, l, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is the polynomial of order $r = q - (a + b + 2d_1 + \cdots + 2d_l)$ that has no zero on the unit disk in the complex plane. When $a = b = d_k = 0$, $\{X_t\}$ is stationary. When one of a, b and d_k is nonzero, the process is said to be unstable. The commonly used IAR model is the case where $a \neq 0 = b = d_k$, for all k.

Denote $\mathbf{X}_t = (X_t, \ldots, X_{t-q+1})'$ and $\mathbf{X}_0 = \mathbf{0}$. Let

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{t=1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}' \right)^{-1} \sum_{t=1}^{n} \mathbf{X}_{t-1} X_t, \qquad n > q, \tag{3.19}$$

be the least squares estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)'$ based on $X_1, \ldots, X_n$. Then the residuals are

$$\hat{\varepsilon}_t = X_t - \hat{\boldsymbol{\beta}}_n' \mathbf{X}_{t-1}, \qquad t = 1, \ldots, n \tag{3.20}$$

and the process under consideration is

$$\hat{E}_n(x) = n^{-1/2} \sum_{t=1}^{n} \big[I(\hat{\varepsilon}_n \leq x) - H(x) \big], \qquad x \in R. \tag{3.21}$$

In the following, we only consider the limiting distribution under the assumption that the true distribution of ε is H. The limiting result, unlike the stationary AR(∞) case, is complicated by the nonstationary feature of $\{X_t\}$. In particular, the locations of the unit roots would affect the Gaussianity of the limiting process.

Note that $\hat{E}_n(x) = \sum_{i=1}^{3} E_{ni}(x)$, where

$$E_{n1}(x) = n^{-1/2} \sum_{t=1}^{n} \big[I(\varepsilon_t \leq x) - H(x) \big],$$

$$E_{n2}(x) = n^{-1/2} \sum_{t=1}^{n} \big[H(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})'\mathbf{X}_{t-1}) - H(x) \big],$$

$$\begin{aligned} E_{n3}(x) = n^{-1/2} \sum_{t=1}^{n} \big[& I(\varepsilon_t \leq x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})'\mathbf{X}_{t-1}) - H(x + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})'\mathbf{X}_{t-1}) \\ & + H(x) - I(\varepsilon_t \leq x) \big]. \end{aligned}$$

We are going to establish the limiting process by three steps. First, we show that E_{n3} is negligible. Second, we obtain the limiting process of E_{n2}. Finally, since E_{n1} is the usual empirical process which has a known limiting process, we will use the continuous mapping theorem to combine all results together.

Next, we follow the idea of Chan and Wei (1988) and decompose the time series into several components so that each component has its own distinct characteristic roots.

Let

$$u_t = \varphi(B)(1 - B)^{-a} X_t,$$
$$v_t = \varphi(B)(1 + B)^{-b} X_t,$$

$$x_t(k) = \varphi(B)\big(1 - 2\cos\theta_k B + B^2\big)^{-d_k} X_t,\;\; k = 1, \dots, l,$$
$$z_t = \varphi(B)\psi^{-1}(B) X_t,$$

where B denotes the back shift operator. Note that

$$(1-B)^a u_t = \varepsilon_t, \qquad (1+B)^b v_t = \varepsilon_t,$$
$$(1 - 2\cos\theta_k B + B^2)^{d_k} x_t(k) = \varepsilon_t \quad \text{and} \quad \varphi(B) z_t = \varepsilon_t.$$

For convenience, set

$$\mathbf{u}_t = \big(u_t, \dots, u_{t-a+1}\big)', \;\; \mathbf{v}_t = (v_t, \dots, v_{t-b+1})',$$
$$\mathbf{x}_t(k) = \big(x_t(k), \dots, x_{t-2d_k+1}(k)\big)'.$$

Since $\mathbf{x}_0 = \mathbf{0}$, we have $\mathbf{u}_0 = \mathbf{v}_0 = \mathbf{x}_0(1) = \cdots = \mathbf{x}_0(l) = \mathbf{z}_0 = \mathbf{0}$.

According to Chan and Wei (1988), there exists a $q \times q$ nonsingular matrix Q such that

$$Q\mathbf{X}_t = \big(\mathbf{u}_t', \mathbf{v}_t', \mathbf{x}_t'(1), \dots, \mathbf{x}_t'(l), \mathbf{z}_t'\big)'$$

and there exist block diagonal matrices $T_n = \operatorname{diag}(J_n, K_n, L_n(1), \dots, L_n(l), M_n)$ such that

$$\begin{aligned} &T_n Q \sum_{t=1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}' Q' T_n' \\ &\sim_P \operatorname{diag}\bigg(J_n \sum_{t=1}^{n} \mathbf{u}_{t-1}\mathbf{u}_{t-1}' J_n', \dots, M_n' \sum_{t=1}^{n} \mathbf{z}_{t-1}\mathbf{z}_{t-1}' M_n'\bigg) \\ &= O_P(1), \end{aligned} \tag{3.22}$$

where J_n, K_n, $L_n(1), \dots, L_n(l)$, M_n are $a \times a$, $b \times b$, $2d_1 \times 2d_1, \dots, 2d_l \times 2d_l$ and $r \times r$ matrices. Moreover, it holds that

$$\begin{aligned} (Q'T_n')^{-1}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta) &\sim_P \begin{pmatrix} (J_n')^{-1}(\sum_{t=1}^n \mathbf{u}_{t-1}\mathbf{u}_{t-1}')^{-1} \sum_{t=1}^n \mathbf{u}_{t-1}\varepsilon_t \\ \vdots \\ (M_n')^{-1}(\sum_{t=1}^n \mathbf{z}_{t-1}\mathbf{z}_{t-1}')^{-1} \sum_{t=1}^n \mathbf{z}_{t-1}\varepsilon_t \end{pmatrix} \\ &= O_P(1). \end{aligned} \tag{3.23}$$

For details, see (3.2)–(3.5) of Chan and Wei (1988). Now, from (3.18), (3.22) and (3.23), we obtain $\sup_x |E_{n3}(x)| = o_P(1)$ in view of Corollary 2.3 and the proposition in the Appendix. This completes the first step.

On the other hand, by Taylor's series expansion,

$$\begin{aligned} E_{n2}(x) &= n^{-1/2} \sum_{t=1}^{n} \big(\hat{\boldsymbol\beta}_n - \boldsymbol\beta\big)'\mathbf{X}_{t-1} H'(x) \\ &\quad + (4n)^{-1/2} \sum_{t=1}^{n} \big\{(\hat{\boldsymbol\beta}_n - \boldsymbol\beta)'\mathbf{X}_{t-1}\big\}^2 H''(\zeta_{nt}) \end{aligned} \tag{3.24}$$

for some random variables ζ_{nt}. In view of (3.18), (3.22) and (3.23), one can see that the second term of the right-hand side of (3.24) is $o_P(1)$. Thus, to complete the analysis of the second step, it remains to deal with the first term. Note that in terms of (3.22) and (3.23),

$$\begin{aligned} & n^{-1/2}\sum_{t=1}^{n}(\hat{\boldsymbol\beta}_n-\boldsymbol\beta)'\mathbf{X}_{t-1} \\ (3.25)\qquad & \sim_P \left[(J_n')^{-1}\left(\sum_{t=1}^{n}\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\right)^{-1}\sum_{t=1}^{n}\mathbf{u}_{t-1}\varepsilon_t\right]^{-1} n^{-1/2}\sum_{t=1}^{n}J_n\mathbf{u}_{t-1} \\ & \quad + \cdots + \left[(M_n')^{-1}\left(\sum_{t=1}^{n}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\sum_{t=1}^{n}\mathbf{z}_{t-1}\varepsilon_t\right]^{-1} n^{-1/2}\sum_{t=1}^{n}M_n\mathbf{z}_{t-1}. \end{aligned}$$

Since the limiting behavior of the first term in each summand in (3.25) is already established by Chan and Wei (1988), we only have to deal with the second terms. For this, we separate the terms into four cases according to the locations of their characteristic roots. The following four lemmas are concerned with this task. The proofs can be found in Lee [(1991), pages 84–88] and are omitted for brevity.

LEMMA 3.5. *Let W denote a standard Brownian motion on* [0,1] *and define $F_0(u)=\sigma W(u)$ and $F_j(u)=\int_0^u F_{j-1}(s)\,ds$. Then it holds that $n^{-1/2}\sum_{t=1}^{n} J_n\mathbf{u}_{t-1}\to_{\mathscr{D}}(F_1(1),\dots,F_a(1))$ as $n\to\infty$.*

LEMMA 3.6. $n^{-1/2}\sum_{t=1}^{n}K_n\mathbf{v}_{t-1}\to\mathbf{0}$ *a.s. as* $n\to\infty$.

LEMMA 3.7. *For* $k=1,\dots,l$, $n^{-1/2}\sum_{t=1}^{n}L_n(k)\mathbf{x}_{t-1}(k)\to\mathbf{0}$ *as* $n\to\infty$.

LEMMA 3.8. $n^{-1/2}\sum_{t=1}^{n}M_n\mathbf{z}_{t-1}\to\mathbf{0}$ *a.s. as* $n\to\infty$.

From Lemmas 3.5–3.8 and the argument in (3.25), we can see that only the root 1 affects the asymptotic behavior of the residual empirical process. Before we state the main theorem of this subsection, we introduce a lemma.

LEMMA 3.9. *Suppose that $\xi_1,\xi_2,\dots,$ are iid random variables with mean* 0, *variance $\sigma^2\in(0,\infty)$ and continuous distribution G. Define*

$$W_{n1}(u)=n^{-1/2}\sum_{j=1}^{n}\big[I(G(\xi_j)\le u)-u\big],\qquad u\in[0,1]$$

and

$$W_{n2}(u)=n^{-1/2}\sum_{j=1}^{[nu]}\xi_j/\sigma,\qquad u\in[0,1].$$

Then (W_{n1}, W_{n2}) *converges weakly to a mean zero Gaussian process* (W_1, W_2) *in* $D^2[0,1]$ *space, such that for all* $s, t \in [0,1]$,

$$\begin{aligned}
&\mathrm{Cov}(W_1(s), W_1(t)) = s \wedge t - st,\\
&\mathrm{Cov}(W_2(s), W_2(t)) = s \wedge t,\\
&\mathrm{Cov}(W_1(s), W_2(t)) = (t/\sigma)\int_{-\infty}^{G^{-1}(s)} x\, dG(x).
\end{aligned}$$

The proof is rather standard and is omitted.

THEOREM 3.2. *Let* $\hat{E}_n(x)$ *be the empirical process defined by* (3.21) *and suppose that the characteristic polynomial* φ *has root* 1 *with multiplicity* $a \geq 1$. *Suppose that* (W_1, W_2) *is the Gaussian process defined in Lemma* 3.9, $F_0 = \sigma W_2$, $F_j = \int_0^1 F_{j-1}(s)\, ds$, $\boldsymbol{\xi} = (\int_0^1 F_{a-1}(s)\, dW_2(s), \ldots, \int_0^1 F_0(s)\, dW_2(s))'$, $\boldsymbol{\eta} = (F_a(1), \ldots, F_1(1))'$, *and* F *is the matrix whose* (j,l)*th entry is* $\sigma_{jl} = \int_0^1 F_{j-1}(s) F_{l-1}(s)\, ds$. *Then, as* $n \to \infty$,

$$\hat{Y}_n(u) := \hat{E}_n(H^{-1}(u)) \xrightarrow{\mathscr{D}} W_1(u) + \sigma(F^{-1}\boldsymbol{\xi})'\boldsymbol{\eta} H'(H^{-1}(u)). \tag{3.26}$$

PROOF. As we have seen earlier, it holds that

$$\begin{aligned}
\sup_x \Bigg| E_{n2}(x) - \Bigg[(J_n')^{-1} \bigg(\sum_{t=1}^n \mathbf{u}_{t-1}\mathbf{u}_{t-1}' \bigg)^{-1} \sum_{t=1}^n \mathbf{u}_{t-1}\varepsilon_t \Bigg]^{-1} & \\
\times n^{1/2} \sum_{t=1}^n J_n \mathbf{u}_{t-1} H'(x) \Bigg| &= o_P(1).
\end{aligned} \tag{3.27}$$

Meanwhile, by Lemma 3.5, Theorem 3.1.2 and the Proposition in Appendix 3 of Chan and Wei (1988), and the continuous mapping theorem, we have that

$$\Bigg[(J_n')^{-1} \bigg(\sum_{t=1}^n \mathbf{u}_{t-1}\mathbf{u}_{t-1}' \bigg)^{-1} \sum_{t=1}^n \mathbf{u}_{t-1}\varepsilon_t \Bigg]^{-1} n^{1/2} \sum_{t=1}^n J_n \mathbf{u}_{t-1} \xrightarrow{D} \sigma(F^{-1}\boldsymbol{\xi})'\boldsymbol{\eta}. \tag{3.28}$$

Recall here that $E_n(x) = \sum_{i=1}^3 E_{ni}(x)$ and $\sup_x |E_{n3}(x)| = o_P(1)$. Then, (3.26) is yielded by (3.27), (3.28), Lemma 3.9, the Proposition of Appendix 3 of Chan and Wei (1988) and the continuous mapping theorem. □

Theorem 3.2 indicates that, unlike the stationary case, the residual empirical process from the unstable process with the root 1 does not behave like a

Brownian bridge asymptotically. Therefore one should perform a unit root test a priori before using the conventional tests, such as the Kolmogorov–Smirnov test and the Cramér–von Mises test.

Finally, we remark that for the explosive case, the proof of the result (3.29) given by Koul and Levental (1989) can be substantially shortened by Lemma 2.2.

REMARK 3.1 [Explosive AR(1) processes]. Consider the explosive AR(1) process

$$X_t = \rho X_{t-1} + \varepsilon_t,$$

where ε_1 has the distribution G, $|\rho| > 1$ and $X_0 = 0$. Koul and Levental (1989) showed that

$$\hat{Y}_n(u) = n^{-1/2} \sum_{j=1}^{n} \Big[I(G(\hat{\varepsilon}_j) \le u) - u \Big] \to_{\mathscr{D}} \dot{W}(u), \qquad u \in [0, 1], \tag{3.29}$$

where $\hat{\varepsilon}_j = X_j - \hat{\rho}_n X_{j-1}$, $\hat{\rho}_n$ is an estimate of ρ and $\dot{W}$ is a standard Brownian bridge, under the conditions $E \log^+ |\varepsilon_1| < \infty$; G has a uniformly bounded derivative; $\hat{\rho}_n$ satisfies $\rho^n(\hat{\rho}_n - \rho) = o_P(n^{1/2})$.

They obtained (3.29) by giving a lengthy proof for the result

$$U_n = \sup_x \left| n^{-1/2} \sum_{j=1}^{n} u_{nj}(x) \right| = o_P(1), \tag{3.30}$$

where

$$u_{nj}(x) = I\big(\varepsilon_j \le x + (\hat{\rho}_n - \rho) X_{j-1}\big) - G\big(x + (\hat{\rho}_n - \rho) X_{j-1}\big) + G(x) - I(\varepsilon_j \le x).$$

However, (3.30) can be derived directly from Lemma 2.2 (see also Remark 2.1). Note that

$$U_n \le \sup_x \left| n^{-1/2} \sum_{j=1}^{n-n^{1/3}} u_{nj}(x) \right| + \sup_x \left| n^{-1/2} \sum_{j=n-n^{1/3}+1}^{n} u_{nj}(x) \right|. \tag{3.31}$$

Since $u_{nj}(x)$ is bounded by 2, the second term of the right-hand side of (3.31) is no more than $2n^{-1/6}$, and thus it converges to 0. On the other hand, replacing n in Lemma 2.2 by $n - n^{1/3}$, the first term is $o_P(1)$, since

$$\max_{1 \le j \le n-n^{1/3}} \big| (\hat{\rho}_n - \rho) X_{j-1} \big| = O_P(n^{1/2} \rho^{-n/3}) = o_P(n^{-1/2})$$

[cf. Koul and Levental (1989), Lemma 1].

APPENDIX

PROPOSITION. Let $\{H_n\}$ be a sequence of distribution functions and $\{\eta_n\}$ be a sequence of positive real numbers decaying to 0 as n tends to infinity. If there exists a distribution function G such that $\sup_x |G''(x)| < \infty$ and

$$\sup_x \left|H'_n(x) - G'(x)\right| \to 0 \quad \text{as } n \to \infty, \tag{A.1}$$

then as $n \to \infty$,

$$\sup_{(x,\, y)\in S_n} \left|H'_n(x) - H'_n(y)\right| \to 0, \tag{A.2}$$

where $S_n = \{(x, y); |H_n(x) - H_n(y)| \le \eta_n\}$.

This proposition is proved through the following two lemmas.

LEMMA A.1. *Let $\{H_n\}$ be a sequence of distribution functions and $\{\eta_n\}$ be a sequence of positive real numbers decaying to 0 as n tends infinity. If there exists a continuously differentiable distribution function G satisfying* (A.1) *and*

$$\lim_{|x|\to\infty} G'(x) \to 0, \tag{A.3}$$

then $\{H_n\}$ satisfies (A.2).

PROOF. Note that by (A.1) and Schéffe's theorem,

$$\sup_x \left|H_n(x) - G(x)\right| \to 0 \quad \text{as } n \to \infty. \tag{A.4}$$

If the lemma is not true, there exist $d > 0$, a subsequence $\{m_n\}$ of $\{n\}$ and a sequence $\{(x_n, y_n)\}$, such that

$$\left|H_{m_n}(x_n) - H_{m_n}(y_n)\right| \le \eta_{m_n} \quad \text{and} \quad \left|H'_{m_n}(x_n) - H'_{m_n}(y_n)\right| > d.$$

Then, by (A.1) and (A.4),

$$|G(x_n) - G(y_n)| \to 0 \quad \text{as } n \to \infty, \tag{A.5}$$

$$|G'(x_n) - G'(y_n)| > d \quad \forall \text{ sufficiently large } n. \tag{A.6}$$

We claim that (x_n, y_n) belongs to a compact set. Otherwise, there exists a subsequence $\{(x_{n'}, y_{n'})\}$ such that at least one of $\{x_{n'}\}$ and $\{y_{n'}\}$ diverges to $\pm\infty$. Suppose that $x_{n'} \to \infty$. (The case for $x_{n'} \to -\infty$ can be handled similarly). If $\{y_{n'}\}$ is bounded, there exists a limit point y_0. Then, $G(y_0) = 0$ by (A.5) and so $G'(y_0) = 0$ by the continuity of G'. However, this contradicts (A.6) since $G'(x_{n'}) \to 0$ by (A.3). Thus, $\{y_{n'}\}$ must have a subsequence $\{y_{n''}\}$ that goes to ∞ or $-\infty$. In this case, however, $|G'(x_{n''}) - G'(y_{n''})| \to 0$ due to (A.3). This contradicts (A.6). Hence, we conclude that (x_n, y_n) is contained in a compact set.

Now, assume that (x_0, y_0) is a limit point of (x_n, y_n). By (A.6), $G'(x_0) \ne G'(y_0)$. However, this leads to a contradiction since (A.5) implies $G(x_0) = G(y_0)$ and consequently $G'(x_0) = G'(y_0)$. This completes the proof. □

LEMMA A.2. *If G is a distribution function with uniformly bounded second derivative, then G satisfies condition* (A.3).

PROOF. Assume that $\lim_{x\to\infty} G'(x) = 0$ does not hold. (The case for $x \to -\infty$ can be handled in a similar fashion). Then, there exist a positive real number ε and a sequence of real numbers $\{x_n\}$ diverging to infinity, such that $G'(x_n) > \varepsilon$. Put $\delta = \sup_x |G''(x)|\varepsilon/2$. Since by the mean value theorem, $|G'(x_n + y) - G'(x_n)| \le \sup_x |G''(x)|\delta = \varepsilon/2$ for all $|y| \le \delta$, we have that

$$G'(x_n + y) \ge G'(x_n) - \varepsilon/2 > \varepsilon/2.$$

Integrating both sides of the above inequality over $[0, \delta]$, we obtain

$$G(x_n + \delta) - G(x_n) \ge \varepsilon\delta/2.$$

However, this leads to a contradiction since the left-hand side of the above inequality goes to 0 as n tends to infinity. This completes the proof. □

Acknowledgments. We are grateful to an Associate Editor and two referees for their helpful suggestions.

REFERENCES

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

BOLDIN, M. V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27** 886–871.

BROCKWELL P. J. and DAVIS, R. A. (1990). *Time Series*: *Theory and Methods,* 2nd ed. Springer, New York.

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

CRAMÉR, H. (1974). *Mathematical Methods of Statistics*. Princeton Univ. Press.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

EPPS, T. W. (1987). Testing that a stationary time series is Gaussian. *Ann. Statist.* **15** 1683–1698.

GRENANDER, U. and SZEGÖ, G. (1958). *Toplitz Forms and Their Applications.* Univ. California Press, Berkeley.

HINICH, M. J. (1982). Testing for Gaussianality and linearity of a stationary time series. *J. Time Series Anal.* **3** 169–176.

KOUL, H. L. (1992). *Weighted Empiricals and Linear Models.* IMS, Hayward, CA.

KOUL, H. L. and LEVENTAL, S. (1989). Weak convergence of residual empirical process in explosive autoregression. *Ann. Statist.* **17** 1784–1794.

KREISS, J. P. (1988). Asymptotic for the empirical distribution of residuals in autoregression with infinite order. Unpublished manuscript.

LEE, S. (1991). Testing whether a time series is Gaussian. Ph.D. dissertation, Dept. Mathematics, Univ. Maryland.

LII, K. S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *Ann. Statist.* **10** 1195–1208.

PIERCE, D. A. (1985). Testing normality in autoregressive models. *Biometrika* **72** 293–297.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser. Boston.

ROUSSAS, G. (1972). *Contiguity of Probability Measures*: *Some Applications in Statistics.* Cambridge Univ. Press.

SHIBATA, R. (1980). Asymptotically efficient selection of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.

SHORACK, G. and WELLNER, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

SUBBA RAO, T. and GABR, M. M. (1980). A test for linearity of stationay time series. *J. Time Series Anal.* **1** 145–158.

WEI, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20** 1–42.

DEPARTMENT OF STATISTICS
SEOUL NATIONAL UNIVERSITY
SEOUL, 151-742
KOREA
E-MAIL: sylee@stats.snu.ac.kr

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, 11529
TAIWAN, R.O.C.
E-MAIL: czw@stat.sinica.edu.tw