J. Japan Statist. Soc.
Vol. 38 No. 1 2008 65–74

# UNIT ROOT MODEL SELECTION*

Peter C. B. Phillips**

Some limit properties for information based model selection criteria are given in the context of unit root evaluation and various assumptions about initial conditions. Allowing for a nonparametric short memory component, standard information criteria are shown to be weakly consistent for a unit root provided the penalty coefficient $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$. Strong consistency holds when $C_n/(\log\log n)^3 \to \infty$ under conventional assumptions on initial conditions and under a slightly stronger condition when initial conditions are infinitely distant in the unit root model. The limit distribution of the AIC criterion is obtained.

*Key words and phrases*: AIC, consistency, model selection, nonparametric, unit root.

## 1. Introduction

Following Akaike (1969, 1973, 1977), information criteria have been systematically explored for order selection purposes, often in the context of time series models like autoregressions. The methods have been studied in both stationary and nonstationary models (Tsay (1984), Pötscher (1989), Wei (1992), Nielsen (2006)) and are widely used in practical work.

A commonly occuring problem in modern time series, particularly econometrics, is model evaluation that involves testing for a unit root and cointegration. Again, order selection methods have been considered in this context (Phillips and Ploberger (1996), Phillips (1996), Kim (1998)). If the focus is on these particular features of a time series then it is not necessary to build a complete model and it is often desirable to perform the evaluation in a semiparametric context allowing for a general short memory component in the series.

The present note looks at the specific issue of unit root evaluation by information criteria. We seek to distinguish processes with a unit root (UR) from stationary series (SS). The UR model has the autoregressive form

$$X_t = \rho X_{t-1} + u_t, \qquad \rho = 1, \qquad t \in \{1, \dots, n\}, \tag{1.1}$$

where $u_t$ is a weakly dependent stationary time series with zero mean and continuous spectral density $f_u(\lambda)$. The series $X_t$ is initialized at $t = 0$ by some (possibly random) quantity $X_0$. The SS model has the form $X_t = u_t$, so that $\rho = 0$ in (1.1). We aim to treat (1.1) semiparametrically with regard to $u_t$ and in this context $\rho = 0$ is effectively equivalent to $|\rho| < 1$ in (1.1).

Accepted December 8, 2007.

*Thanks go to Xu Cheng and a referee for comments on the original version. Partial support is acknowledge from a Kelly Fellowship and from the NSF under Grant No. SES 06-47086.

**University of Auckland, University of York, Singapore Management University, Yale University; Department of Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, U.S.A. Email: peter.phillips@yale.edu

Standard order selection criteria may be used to evaluate whether $\rho = 1$ or $\rho = 0$ in (1.1). The criteria have the following form

$$IC_k = \log \widehat{\sigma}_k^2 + \frac{kC_n}{n} \tag{1.2}$$

with coefficient $C_n = \log n, \log\log n, 2$ corresponding to the BIC (Schwarz (1978), Akaike (1977), Rissanen (1978)), Hannan and Quinn (1979), and Akaike (1973) penalties, respectively. Sample information-based versions of the coefficient $C_n$ may also be employed, such as those in Wei's (1992) FIC criterion and Phillips and Ploberger's (1996) PIC criterion.

In the unit root (UR) autoregression, $\rho = 1$ and there is no unknown autoregressive parameter to estimate, so in this case we set the parameter count to $k = 0$ in (1.2). In the stationary model (SS), a parametric autoregressive model may still be fitted by least squares regression with

$$\hat{\rho} = \sum_{t=1}^{n} X_t X_{t-1} \Big/ \sum_{t=1}^{n} X_{t-1}^2 = \sum_{t=1}^{n} u_t u_{t-1} \Big/ \sum_{t=1}^{n} u_{t-1}^2,$$

and so in this case the parameter count is set to $k = 1$. The residual variance estimates in (1.2) for the two models are formed in the usual manner, viz.,

$$\widehat{\sigma}_0^2 = n^{-1} \sum_{t=1}^{n} (\Delta X_t)^2, \qquad \widehat{\sigma}_1^2 = n^{-1} \sum_{t=1}^{n} (X_t - \hat{\rho} X_{t-1})^2.$$

Model evaluation based on $IC_k$ then leads to the selection criterion $\hat{k} = \arg\min_{k \in \{0,1\}} IC_k$.

As shown below, the information criterion $IC_k$ is weakly consistent for testing a unit root provided the penalty term in (1.2) satisfies the weak requirements that $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$. No specific expansion rate $C_n$ is required. Strong consistency also holds provided $C_n \to \infty$ faster than $(\log\log n)^3$ under commonly used assumptions about initial conditions.

## 2. Results

The following assumptions make specific the semiparametric and initialization components of (1.1), the second being important when $\rho = 1$. Assumption **LP** is a standard linear process condition of the type that is convenient in developing partial sum limit theory (c.f., Phillips and Solo (1992)). Assumption **IN** gives, for the unit root case, a partial sum structure to the initial observation $X_0$ in terms of past innovations, making $X_0$ analogous to later observations $X_t$ of the series which take the form of partial sums measured from $X_0$. The sequence $\kappa_n$ in (2.2) determines how many past innovations are included in the initialization, with larger values of $\kappa_n$ associated with the more distant past. This type of initial condition has been used in other recent limit theory in econometrics (Phillips and Magdalinos (2007)).

ASSUMPTION **LP**. Let $d(L) = \sum_{j=0}^{\infty} d_j L^j$, with $d_0 = 1$ and $d(1) \neq 0$, and let $u_s$ have Wold representation

$$u_s = d(L)\varepsilon_s = \sum_{j=0}^{\infty} d_j \varepsilon_{s-j}, \qquad \text{with} \quad \sum_{j=0}^{\infty} j^{1/2}|d_j| < \infty, \tag{2.1}$$

where $\varepsilon_t$ is iid$(0, \sigma_\varepsilon^2)$. Define $\lambda = \sum_{h=1}^{\infty} E(u_t u_{t-h})$, $\omega^2 = \sum_{h=-\infty}^{\infty} E(u_t u_{t-h})$, and $\sigma^2 = E\{u_t^2\}$.

ASSUMPTION **IN**. The initialization of (1.1) when $\rho = 1$ has the general form

$$X_0(n) = \sum_{j=0}^{\kappa_n} u_{-j} \tag{2.2}$$

with $u_{-j}$ satisfying Assumption **LP** and $\kappa_n$ an integer valued sequence satisfying $\kappa_n \to \infty$ and

$$\frac{\kappa_n}{n} \to \tau \in [0, \infty] \qquad \text{as} \quad n \to \infty. \tag{2.3}$$

The following cases are distinguished:

(i) If $\tau = 0$, $X_0(n)$ is said to be a recent past initialization.
(ii) If $\tau \in (0, \infty)$, $X_0(n)$ is said to be a distant past initialization.
(iii) If $\tau = \infty$, $X_0(n)$ is said to be an infinite past initialization.

THEOREM 1.

(a) *Under Assumptions* **LP** *and* **IN**, *the criterion* $IC_k$ *is weakly consistent for distinguishing unit root and stationary time series provided* $C_n \to \infty$ *and* $C_n/n \to 0$ *as* $n \to \infty$.

(b) *The asymptotic distribution of the AIC criterion* ($IC_k$ *with coefficient* $C_n = 2$) *is given by*

$$\lim_{n\to\infty} P\{\hat{k}_{AIC} = 0 \mid k = 1\} = 0, \qquad \lim_{n\to\infty} P\{\hat{k}_{AIC} = 1 \mid k = 1\} = 1,$$
$$\lim_{n\to\infty} P\{\hat{k}_{AIC} = 0 \mid k = 0\} = P\{\xi^2 < 2\}$$
$$\lim_{n\to\infty} P\{\hat{k}_{AIC} = 1 \mid k = 0\} = 1 - P\{\xi^2 < 2\},$$

*where*

$$\xi^2 = \begin{cases} \left(\int_0^1 dBB + \lambda\right)^2 \Big/ \left(\sigma^2 \int_0^1 B^2\right) & \textit{under } \text{IN(i)} \\ \left(\int_0^1 dBB_\tau + \lambda\right)^2 \Big/ \left(\sigma^2 \int_0^1 B_\tau^2\right) & \textit{under } \text{IN(ii)} \\ B(1)^2/\sigma^2 & \textit{under } \text{IN(iii)} \end{cases},$$

*B is Brownian motion with variance* $\omega^2$,

$$B_\tau(s) = B(s) + \sqrt{\tau} B_0(1), \tag{2.4}$$

*and* $B_0$ *is an independent Brownian motion with variance* $\omega^2$.

*Remarks.*

**(1)** The weak consistency results in part (a) of Theorem 1 show that information criteria can be used, essentially in their present form, for distinguishing unit root and stationary time series. This approach allows for a nonparametric treatment of the short memory component in both the stationary and nonstationary models. The simple conditions on the penalty coefficient $C_n$ that $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$ are minimal. Evidently, BIC and the Hannan-Quinn criterion are both consistent. Similar arguments show that the FIC (Wei (1992)) and PIC (Phillips and Ploberger (1996)) criteria are also consistent.

**(2)** The AIC criterion (with fixed $C_n = 2$) is inconsistent and the limit distribution of $\hat{k}$ is seen in part (b) of the theorem to depend on the asymptotic distribution of the squared unit root $t$ statistic $\xi^2$. This distribution involves nuisance parameters. The limit variate $\xi^2$ has a unit root limit distribution under IN(i) and IN(ii) that depends on $\omega^2$, $\sigma^2$ and $\lambda$, and a scaled chi-squared distribution under IN(iii) that depends on $\omega^2$ and $\sigma^2$.

Theorem 2. *Under Assumptions* **LP** *and* **IN**, *the criterion $IC_k$ is strongly consistent for distinguishing unit root and stationary time series provided*

$$\frac{C_n}{(\log\log n)^3} \to \infty \qquad \text{under IN(i) and IN(ii)}$$
$$\frac{C_n}{\frac{\kappa_n}{n}(\log\log n)^2 \log\log \kappa_n} \to \infty \qquad \text{under IN(iii)} \tag{2.5}$$

*and $C_n/n \to 0$ as $n \to \infty$.*

*Remarks.*

**(3)** The rate condition (2.5) implies that BIC is strongly consistent in distinguishing unit root and stationary models under IN(i) and IN(ii) and strongly consistent under IN(iii) provided $\kappa_n$ does not increase too fast relative to $n$. The results complement those of Wei (1992), who proved strong consistency of the BIC and FIC criteria in order selection for parametric autoregressions allowing for nonstationarity.

**(4)** The proof of Theorem 2 depends on the asymptotic behavior of the quadratic form $u'P_{-1}u$, which involves the projection matrix $P_{-1} = X_{-1}(X_{-1}'X_{-1})^{-1}X_{-1}'$, where $X_{-1} = (X_0, X_1, \ldots, X_{n-1})'$, and $u = (u_1, \ldots, u_n)'$. The asymptotic properties of projections of this type arising in stochastic regression models were studied by Lai and Wei (1982a) for martingale differences $u_t$ in a general regression setting. Their Lemma 2(ii) and Theorem 3 (see, in particular, equation (2.18)) give the following order for $u'P_{-1}u$ when $X_t$ is generated by the unit root model UR and $u_t$ is a martingale difference with uniform $2+\eta$ moments with $\eta > 0$:

$$u'P_{-1}u = O_{\text{a.s.}}\left(\log\left(\sum_1^n X_{t-1}^2\right)\right) = O_{\text{a.s.}}(\log n). \tag{2.6}$$

(See also Lemma 2(iii) and equation (3.25) in Lai and Wei (1982b).) In parametric stochastic regression models that include nonstationary autoregressions, Pötscher (1989) used (2.6) to establish strong consistency of information criteria of the form $IC_k$ under an expansion rate for the penalty coefficient $C_n$ that requires $C_n/\log n \to \infty$, thereby excluding BIC. In the proof of Theorem 2, we make explicit use of the fact that $X_{t-1} = S_{t-1} + X_0$, where $S_t$ is a partial sum of the $u_j$, to establish that in this case

$$u'P_{-1}u = \begin{cases} O_{\text{a.s.}}((\log\log n)^3) & \text{under IN(i) and IN(ii)} \\ O_{\text{a.s.}}\left(\dfrac{\kappa_n}{n}(\log\log n)^2 \log\log \kappa_n\right) & \\ & \text{under IN(iii)} \end{cases}, \tag{2.7}$$

giving a sharper result than (2.6). In proving (2.6), Lai and Wei (1982a, Lemma 2) consider the sample covariance of a martingale difference with a general random sequence that is not necessarily a partial sum process of the innovations. In view of the unit root structure, we can make use of the following explicit decomposition of the sample covariance

$$\sum_{t=1}^{n} S_{t-1}u_t = \frac{1}{2}\left\{S_n^2 - \sum_{t=1}^{n} u_t^2\right\} = O_{\text{a.s.}}(n\log\log n),$$

which, in conjunction with a lower bound result for $\sum_{t=1}^{n} S_{t-1}^2$, leads directly to (2.7) under the commonly used initial conditions IN(i) and IN(ii).

## 3. Proofs

Proof of Theorem 1.

**Part (a)** Suppose the true model is a UR model with $\rho = 1$ in (1.1). Then

$$IC_0 = \log \widehat{\sigma}_0^2 = \log\left\{\frac{u'u}{n}\right\}.$$

Define $P_{-1} = X_{-1}(X_{-1}'X_{-1})^{-1}X_{-1}'$, $X_{-1} = (X_0, X_1, \ldots, X_{n-1})'$, and $u = (u_1, \ldots, u_n)'$. The behavior of $IC_1$ depends on $u'P_{-1}u$, which we now investigate under the various initializations.

Under IN(i), we have (Phillips (1987))

$$u'P_{-1}u \Rightarrow \left(\int_0^1 dBB + \lambda\right)^2 \Big/ \int_0^1 B^2, \qquad \text{as} \quad n \to \infty, \tag{3.1}$$

where $\lambda = \sum_{h=1}^{\infty} E(u_t u_{t-h})$, and $B$ is Brownian motion with variance $\omega^2$. Under IN(ii) we have (e.g., Phillips and Magdalinos (2007))

$$u'P_{-1}u \Rightarrow \left(\int_0^1 dBB_\tau + \lambda\right)^2 \Big/ \int_0^1 B_\tau^2, \tag{3.2}$$

where $B_\tau(s) = B(s) + \sqrt{\tau} B_0(1)$, $\tau \in (0, \infty)$, and $B_0$ is a Brownian motion with variance $\omega^2$ that is independent of $B$. Under IC(iii) Phillips and Magdalinos (2007, Theorem 2) show that

$$u' P_{-1} u = \frac{\left( \frac{1}{\sqrt{\kappa_n n}} \sum_{t=1}^n X_{t-1} u_t \right)^2}{\frac{1}{\kappa_n n} \sum_{t=1}^n X_{t-1}^2} \Rightarrow B(1)^2. \tag{3.3}$$

Thus, whether the initialization is recent, distant or infinitely distant, we have $u' P_{-1} u = O_p(1)$.

It follows that

$$\begin{aligned} IC_1 &= \log \widehat{\sigma}_1^2 + \frac{C_n}{n} = \log \left\{ n^{-1} \sum_{t=1}^n (X_t - \hat{\rho} X_{t-1})^2 \right\} + \frac{C_n}{n} \\ &= \log \left\{ \frac{1}{n} (u'u - u' P_{-1} u) \right\} + \frac{C_n}{n} \\ &= \log \frac{u'u}{n} + \log \left[ 1 - \frac{u' P_{-1} u}{u'u} \right] + \frac{C_n}{n} \\ &= \log \frac{u'u}{n} + \log \left[ 1 - \frac{u' P_{-1} u}{n\sigma^2 \{1 + o_{\text{a.s.}}(1)\}} \right] + \frac{C_n}{n} \\ &= \log \frac{u'u}{n} + \frac{C_n}{n} - \frac{u' P_{-1} u}{n\sigma^2} \{1 + o_{\text{a.s.}}(1)\}. \end{aligned}$$

Hence,

$$\begin{aligned} IC_0 - IC_1 &= -\frac{C_n}{n} + \frac{u' P_{-1} u}{n\sigma^2} \{1 + o_{\text{a.s.}}(1)\} \\ &= -\frac{C_n}{n} \left\{ 1 - \frac{u' P_{-1} u}{C_n \sigma^2} + o_{\text{a.s.}} \left( \frac{1}{C_n} \right) \right\} \\ &< 0 \end{aligned} \tag{3.4}$$

when $C_n \to \infty$ as $n \to \infty$ because $\frac{u' P_{-1} u}{C_n \sigma^2} = O_p(C_n^{-1})$. Thus, criterion $IC_k$ correctly selects the unit root model in favor of the stationary model when $\rho = 1$. This is the case irrespective of the initial condition and holds provided $C_n \to \infty$.

Next suppose the true model is stationary and $X_t = u_t$. Then we have

$$\hat{\rho} = \sum_{t=1}^n X_t X_{t-1} \Big/ \sum_{t=1}^n X_{t-1}^2 = \sum_{t=1}^n u_t u_{t-1} \Big/ \sum_{t=1}^n u_{t-1}^2 \underset{\text{a.s.}}{\to} \rho = \frac{E(u_t u_{t-1})}{E(u_{t-1}^2)} := \frac{\gamma_1}{\gamma_0},$$

where $\gamma_h = E\{u_t u_{t-h}\}$. Observe that in this case by the strong law of large numbers

$$\begin{aligned} IC_0 &= \log \widehat{\sigma}_0^2 = \log \left\{ \frac{1}{n} (u - u_{-1})' (u - u_{-1}) \right\} \\ &= \log \left\{ \frac{u'u}{n} - 2 \frac{u' u_{-1}}{n} + \frac{u_{-1}' u_{-1}}{n} \right\} = \log\{2\gamma_0 - 2\gamma_1 + o_{\text{a.s.}}(1)\} \\ &= \log\{2\gamma_0 (1 - \rho)\} + o_{\text{a.s.}}(1). \end{aligned}$$

Also, $n^{-1}u'P_{-1}u = (n^{-1}u'u_{-1})^2/(n^{-1}u'_{-1}u_{-1}) \underset{\text{a.s.}}{\to} \gamma_1^2/\gamma_0$, and then

$$\begin{aligned} IC_1 &= \log\left\{\frac{1}{n}(u'u - u'P_{-1}u)\right\} + \frac{C_n}{n} \\ &= \log\left\{\left(\frac{u'u}{n}\right)\left(1 - \frac{n^{-1}u'P_{-1}u}{n^{-1}u'u}\right)\right\} + \frac{C_n}{n} \\ &= \log\left\{\gamma_0\left(1 - \frac{\gamma_1^2}{\gamma_0^2}\right) + o_{\text{a.s.}}(1)\right\} + \frac{C_n}{n} \\ &= \log\{\gamma_0(1-\rho^2)\} + \frac{C_n}{n} + o_{\text{a.s.}}(1). \end{aligned}$$

It follows that

$$\begin{aligned} IC_0 - IC_1 &= \log\{2\gamma_0(1-\rho)\} - \log\{\gamma_0(1-\rho^2)\} - \frac{C_n}{n} + o_{\text{a.s.}}(1) \\ &= \log\frac{2}{1+\rho} - \frac{C_n}{n} + o_{\text{a.s.}}(1) > 0, \qquad \text{a.s.} \end{aligned} \tag{3.5}$$

as $n \to \infty$ provided $\frac{C_n}{n} \to 0$. Hence, the criterion $IC_k$ correctly selects the stationary model a.s. as $n \to \infty$ for both fixed $C_n$ and $C_n \to \infty$ at a slower rate than $n$.

**Part (b)** We seek to find the limit distribution of the AIC criterion (i.e., $IC_k$ in (1.2) with $C_k = 2$). Note from the above that AIC makes the correct choice when the model is stationary as (3.5) holds when $C_n$ is fixed as $n \to \infty$. Then

$$\lim_{n\to\infty} P\{\hat{k}_{AIC} = 0 \mid k = 1\} = 0, \qquad \lim_{n\to\infty} P\{\hat{k}_{AIC} = 1 \mid k = 1\} = 1, \tag{3.6}$$

as stated. On the other hand, when the model has a unit root, we have from (3.4) and (3.1)–(3.3)

$$\begin{aligned} n(IC_0 - IC_1) &= -2 + \frac{u'P_{-1}u}{\sigma^2}\{1 + o_{\text{a.s.}}(1)\} \\ &\Rightarrow -2 + \xi^2, \end{aligned}$$

where

$$\xi^2 = \begin{cases} \left(\int_0^1 dBB + \lambda\right)^2 \Big/ \left(\sigma^2 \int_0^1 B^2\right) & \text{under IN(i)} \\ \left(\int_0^1 dBB_\tau + \lambda\right)^2 \Big/ \left(\sigma^2 \int_0^1 B_\tau^2\right) & \text{under IN(ii)} \\ B(1)^2/\sigma^2 & \text{under IN(iii)} \end{cases}.$$

It follows that

$$\begin{aligned} &\lim_{n\to\infty} P\{\hat{k}_{AIC} = 0 \mid k = 0\} = \lim_{n\to\infty} P\{n(IC_0 - IC_1) < 0\} = P\{\xi^2 < 2\} \\ &\lim_{n\to\infty} P\{\hat{k}_{AIC} = 1 \mid k = 0\} = 1 - P\{\xi^2 < 2\}. \end{aligned}$$

Combining this with (3.6) gives the required limit distribution.

PROOF OF THEOREM 2. In view of (3.5), when the stationary model is true the criterion $IC_k$ correctly selects the stationary model a.s. as $n \to \infty$ for fixed $C_n$ and for $C_n \to \infty$ provided $C_n/n \to 0$. To establish strong consistency we therefore need only consider the limit behavior under the unit root model. From (3.4) we have

$$IC_0 - IC_1 = -\frac{C_n}{n}\left\{1 - \frac{u'P_{-1}u}{C_n\sigma^2} + o_{a.s.}\left(\frac{1}{C_n}\right)\right\},$$

and so we need to examine the limit behavior of

$$\frac{u'P_{-1}u}{C_n} = \frac{(u'X_{-1})^2}{C_n(X'_{-1}X_{-1})}.$$

By a result of Donsker and Varadhan (1977, equation (4.6) on p. 751)—see also equation (3.29) of Lai and Wei (1982a)—we have under IN(i)

$$\liminf_{n\to\infty} \frac{\log\log n}{n^2}\sum_{t=1}^{n} X_{t-1}^2 > 0, \qquad \text{a.s.,} \tag{3.7}$$

which gives a lower limit of $O_{\text{a.s.}}(\frac{n^2}{\log\log n})$ to the fluctuations of the sample information $\sum_{t=1}^{n} X_{t-1}^2$. In this case, $n^{-1/2}X_t$ behaves in the limit like a Brownian motion and the lower limit (3.7) is obtained in Donsker and Varadhan (1977, p. 751) by way of the lower limit of the corresponding limiting quantity $\frac{\log\log n}{n^2}\int_0^n B(s)^2 ds$. Result (3.7) may also be shown to hold under IN(ii) and IN(iii). In particular, note that in both these cases we can write $X_t = S_t + X_0(n)$, where $S_t = \sum_1^t u_j$. Then

$$\sum_1^n X_{t-1}^2 \geq \sum_{t=1}^{n}(X_{t-1} - \bar{X}_{-1})^2 = \sum_{t=1}^{n}(S_{t-1} - \bar{S}_{-1})^2,$$

where $\bar{X}_{-1} = n^{-1}\sum_{t=1}^{n} X_{t-1}$ and $\bar{S}_{-1} = n^{-1}\sum_{t=1}^{n} S_{t-1}$, so that

$$\liminf_{n\to\infty} \frac{\log\log n}{n^2}\sum_{t=1}^{n} X_{t-1}^2 \geq \liminf_{n\to\infty} \frac{\log\log n}{n^2}\sum_{t=1}^{n}(S_{t-1} - \bar{S}_{-1})^2. \tag{3.8}$$

The lower bound (3.8) is the same as

$$\begin{aligned}\liminf_{n\to\infty} &\frac{\log\log n}{n^2}\int_0^n \left(B(s) - n^{-1}\int_0^n B(s)ds\right)^2 ds \\ &= \liminf_{n\to\infty} \int_0^1 \left(B_n(p) - \int_0^1 B_n\right)^2 dp,\end{aligned} \tag{3.9}$$

where

$$B_n(p) = \left(\frac{\log\log n}{n}\right)^{1/2} B(pn) \quad \text{for} \quad 0 \leq p \leq 1$$

is Brownian motion over $[0,1]$ with variance $(\log\log n)\omega^2 \to \infty$. The lower limit of the variation (3.9) therefore satisfies

$$\liminf_{n\to\infty} \int_0^1 \left(B_n(p) - \int_0^1 B_n\right)^2 dp > 0 \qquad \text{a.s.}$$

by virtue of the properties of Brownian motion. Otherwise $B_n(p)$ would necessarily be constant with positive probability as $n \to \infty$. It follows that under IN(ii) and IN(iii) the same lower bound of order $O_{\text{a.s.}}(\frac{n^2}{\log\log n})$ applies for $\sum_1^n X_{t-1}^2$.

Next note that

$$\begin{aligned} \sum_{t=1}^n u_t X_{t-1} &= \sum_{t=1}^n u_t S_{t-1} + X_0 \sum_{t=1}^n u_t \\ &= \frac{1}{2}\left\{S_n^2 - \sum_{t=1}^n u_t^2\right\} + X_0 \sum_{t=1}^n u_t \\ &= O_{\text{a.s.}}(n(\log\log n)) + O_{\text{a.s.}}(X_0(n)\sqrt{n\log\log n}), \end{aligned} \tag{3.10}$$

by virtue of the law of the iterated logarithm (e.g., Phillips and Solo (1992)) for $S_n$. It follows that under IN(i) and IN(ii) and in view of a further application of the law of the iterated logarithm to $X_0(n)$ and using (3.7) we find that

$$u'P_{-1}u = O_{\text{a.s.}}\left(\frac{n^2(\log\log n)^2}{n^2/\log\log n}\right) = O_{\text{a.s.}}((\log\log n)^3).$$

We deduce that

$$IC_0 - IC_1 = -\frac{C_n}{n}\left\{1 - \frac{u'P_{-1}u}{C_n\sigma^2} + o_{\text{a.s.}}\left(\frac{1}{C_n}\right)\right\} < 0,$$

whenever

$$\frac{C_n}{(\log\log n)^3} \to \infty, \tag{3.11}$$

as $n \to \infty$ for then $\frac{u'P_{-1}u}{C_n\sigma^2} = o_{\text{a.s.}}(1)$ and $IC_0 < IC_1$ a.s. as $n \to \infty$. This proves strong consistency under the rate condition (3.11) and initial conditions IN(i) and IN(ii).

When IN(iii) applies, (3.10) holds and we have

$$\begin{aligned} \sum_{t=1}^n u_t X_{t-1} &= O_{\text{a.s.}}(n(\log\log n)) + O_{\text{a.s.}}(X_0(n)\sqrt{n\log\log n}) \\ &= O_{\text{a.s.}}(\sqrt{\kappa_n \log\log \kappa_n}\sqrt{n\log\log n}), \end{aligned}$$

by the law of the iterated logarithm for $X_0(n)$. Then, using (3.7), we have

$$u'P_{-1}u = O_{\text{a.s.}}\left(\frac{n\kappa_n \log\log n \log\log \kappa_n}{n^2/\log\log n}\right) = O_{\text{a.s.}}\left(\frac{\kappa_n}{n}(\log\log n)^2 \log\log \kappa_n\right),$$

and deduce that

$$IC_0 - IC_1 = -\frac{C_n}{n}\left\{1 - \frac{u'P_{-1}u}{C_n\sigma^2} + o_{\text{a.s.}}\left(\frac{1}{C_n}\right)\right\} < 0$$

whenever

$$\frac{C_n}{\frac{\kappa_n}{n}(\log\log n)^2\log\log\kappa_n} \to \infty. \tag{3.12}$$

This proves strong consistency under the rate condition (3.12) and the initial condition IN(iii).

## References


Akaike, H. (1969). Fitting autoregressive models for prediction, *Annals of the Intstitute of Statistical Mathematics*, **21**, 243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), Akademiai Kiado, Budapest.

Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics* (ed. P. R. Krishnarah), Amsterdam, North-Holland.

Donsker, M. D. and Varadhan, S. R. S. (1977). On laws of the iterated logarithm for local times, *Communications in Pure and Applied Mathematics*, **30**, 707–753.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society*, *Series B*, **41**, 190–195.

Kim, J.-Y. (1998). Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models, *Econometrica*, **66**, 359–380.

Lai, T. L. and Wei, C. Z. (1982a). Asymptotic properties of projections with applications to stochastic regression problems, *Journal of Multivariate Analysis*, **12**, 346–370.

Lai, T. L. and Wei, C. Z. (1982b). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Annals of Statistics*, **10**, 154–166.

Magdalinos, T. and Phillips, P. C. B. (2006). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors, unpublished paper, Yale University.

Nielsen, B. (2006). Order determination in general vector autoregressions, *IMS Lecture Notes-Monograph Series*, **52**, 93–112.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica*, **55**, 277–302.

Phillips, P. C. B. (1996). Econometric model determination, *Econometrica*, **64**, 763–812.

Phillips, P. C. B. and Magdalinos, T. (2007). Unit root and cointegrating limit theory when initialization is in the infinite past, unpublished paper, Yale University.

Phillips P. C. B. and Ploberger, W. (1996). An asymptotic theory of Bayesian inference for time series, *Econometrica*, **64**, 381–413.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes, *Annals of Statistics*, **20**, 971–1001.

Pötscher, B. M. (1989). Model selection under nonstationarity: Autoregressive models and stochastic linear regression models, *Annals of Statistics*, **17**, 1257–1274.

Rissanen, J. (1978). Modeling by shortest data description, *Automatica*, **14**, 465–471.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

Tsay, R. S. (1984). Order selection in nonstationary autoregressive models, *Annals of Statistics*, **12**, 1425–1433.

Wei, C. Z. (1992). On predictive least squares principles, *Annals of Statistics*, **20**, 1–42.