Statistica Sinica Preprint No: SS-2016-0069.R2

Title	Nearly Unstable Processes: A Prediction Perspective
Manuscript ID	SS-2016-0069.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0069
Complete List of Authors	Ching-Kang Ing Ngai Hang Chan and Rongmao Zhang
Corresponding Author	Ching-Kang Ing
E-mail	cking@stat.nthu.edu.tw

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Nearly Unstable Processes: A Prediction Perspective

Ngai Hang Chan
The Chinese University of Hong Kong

Ching-Kang Ing
National Tsing Hua University

Rongmao Zhang
Zhejiang University

Abstract

Prediction has long been a vibrant topic in modern probability and statistics. In addition to finding optimal forecasts and for model selection, it is argued in this paper that the prediction principle can also be used to analyze critical phenomena, in particular, in stationary and unstable time series. Although the notion of nearly unstable models has become one of the important concepts in time series econometrics, its role from a prediction perspective is less developed. Based on moment bounds for the extreme-value (EV) and least squares (LS) estimates, asymptotic expressions for the mean squared prediction errors (MSPE) of the EV and LS predictors are obtained for a nearly unstable first-order autoregressive (AR(1)) model with positive error. These asymptotic expressions are further extended to a general class of nearly unstable models, thereby allowing one to understand to what degree such general models can be used to establish a link between stationary and unstable models from a prediction perspective. As applications, we illustrate the usefulness of these results in conducting finite sample approximations of the MSPE for near unit-root time series.

Keywords: Extreme-value predictor, least squares predictor, mean squared prediction error, nearly unstable process, positive error, quantum leap.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

1 Introduction

Prediction has long been a vibrant topic in probability and statistics. The seminal monograph of Whittle (1963) illustrates the importance of linear prediction. There are several objectives in prediction studies. The first one is computing an optimal forecast, based on either finite or infinite samples. The second is to use prediction methods for model selection. The third, less well known but of no less importance, is to use the prediction principle to understand critical phenomena, in particular, in stationary and unstable processes, see for example Wei (1992). This goal is the main focus of the present paper.

To achieve this goal, moment bounds are indispensable tools. For example, based on maximal moment inequalities for martingales, Wei (1987, 1992) provided an asymptotic expression for the accumulated prediction error (APE) of a linear stochastic regression model, which in turn leads to the Fisher information criterion (FIC) for model selection. Findley and Wei (2002) and Chan and Ing (2011) established inverse moment bounds for the Fisher information matrices of time series models. These bounds enable one to calculate the mean squared prediction errors (MSPE) of least squares predictors, and to derive the Akaike information criterion (AIC; Akaike (1974)) and the final prediction error criterion (FPE; Akaike (1969)) in a rigorous manner.

Studies of moment bounds and MSPE have mostly been focused on least squares procedures, while less attention has been given to the so-called extreme-value estimates (EVE), mainly used for heavy-tailed dependent data. Due to the emergence of big data, dependent heavy-tailed phenomena have been reported in various disciplines, see Finkenstädt and Rootzén (2004) and the examples therein. To appreciate the significance of such types of estimates, suppose that the data are generated from the first-order autoregressive (AR(1)) model

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where $0 \le \rho < 1$ and ε_t's are i.i.d. positive noise with regularly varying density $f_\varepsilon(x)$ at

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

zero,

$$\lim_{x\to 0}\frac{f_\varepsilon(x)}{cx^{\alpha-1}}=1,\ \text{for some unknown } \alpha>0 \text{ and } c>0. \tag{1.2}$$

A popular method for estimating ρ in (1.1) is the least squares estimator (LSE),

$$\tilde{\rho}_n=\sum_{i=2}^{n}(y_{i-1}-\bar{y})(y_i-\bar{y})/\sum_{i=2}^{n}(y_{i-1}-\bar{y})^2, \tag{1.3}$$

where $\bar{y}=\bar{y}_{n-1}=\frac{1}{n-1}\sum_{i=1}^{n-1}y_i$. When the noise has a density like (1.2), LSE may not be efficient and other estimation procedures are required. When the parametric form of the distribution of ε_t is known, a natural alternative to $\tilde{\rho}_n$ is the maximum likelihood estimator (MLE) yet, as argued in Davis and McCormick (1989) and Ing and Yang (2014), the MLE is in general analytically difficult to work with. A remedy for this difficulty is to use the EVE, $\hat{\rho}_n$, instead, where

$$\hat{\rho}_n=\min_{1\le i\le n-1}y_{i+1}/y_i. \tag{1.4}$$

Here $\hat{\rho}_n$ is the MLE when ε_t has an exponential distribution or is uniformly distributed over $[0,a]$ for some $a>0$; see Bell and Smith (1986). Under an assumption more general than (1.2), Bell and Smith (1986) showed that $\hat{\rho}_n$ is consistent. When (1.2) holds, it is shown in Corollaries 2.4 and 2.5 of Davis and McCormick (1989) that the limit distributions of $\hat{\rho}_n$ satisfies

$$\lim_{n\to\infty}P\{(cM_\alpha(\rho)/\alpha)^{1/\alpha}n^{1/\alpha}(\hat{\rho}_n-\rho)>t\}=\exp\{-t^\alpha\}, \tag{1.5}$$

where $M_\alpha(\rho)=\mathrm{E}(\sum_{j=0}^{\infty}\rho^j\varepsilon_{1-j})^\alpha$. Equation (1.5) reveals that when $\alpha<2$ ($\alpha>2$), the convergence rate of $\hat{\rho}_n$ ($\tilde{\rho}_n$) is faster than that of $\tilde{\rho}_n$ ($\hat{\rho}_n$); see Section 2 of Ing and Yang (2014) for a more comprehensive comparison of $\hat{\rho}_n$ and $\tilde{\rho}_n$.

Model (1.1) with ε_t satisfying (1.2) has found broad applications in hydrology, economics, finance, epidemiology, and quality control; see, among others, Gaver and Lewis (1980), Bell and Smith (1986), Lawrance and Lewis (1985), Davis and McCormick (1989), Smith (1994), Barndorff-Nielsen and Shephard (2001), Nielsen and Shephard (2003), Sarlak (2008), and

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Ing and Yang (2014). Bell and Smith (1986) analyzed two sets of pollution data from the Willamette River, Oregon, using model (1.1) with ε_t following the uniform distribution or exponential distribution; both are special cases of (1.2). Sarlak (2008) adopted model (1.1) with a Weibull error to analyze the annual streamflow data from the Kizilirmak River in Turkey. On the other hand, model (1.1), focusing exclusively on the stationary case $0 \leq \rho < 1$, fails to accommodate data that may fluctuate around an upward trend with variance increasing over time. Ing and Yang (2014) therefore generalized (1.1) to $\rho = 1$, referred to as the unit-root model, and established the limit distribution of (1.4) in this case. In addition, they derived asymptotic expressions for the mean squared prediction errors (MSPE) of the EV predictor ($\hat{y}_{n+1}$) and the LS predictor ($\tilde{y}_{n+1}$), $\mathrm{MSPE}_A = \mathrm{E}(y_{n+1} - \hat{y}_{n+1})^2$ and $\mathrm{MSPE}_B = \mathrm{E}(y_{n+1} - \tilde{y}_{n+1})^2$, as

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A^o(\alpha,\rho), \ \lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = R_B^o(\rho). \tag{1.6}$$

Here $0 \leq \rho \leq 1$, $R_A^o(\alpha,\rho)$ is a positive constant depending on α, ρ, and $f_{\varepsilon_1}(\cdot)$, and $R_B^o(\rho)$ is a positive constant depending on ρ and $\sigma^2 = \mathrm{Var}(\varepsilon_1) > 0$.

While (1.6) suggests that $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$ can be approximated by $R_A^o(\alpha,\rho)$ and $R_B^o(\rho)$, such an approximation may be unsatisfactory when ρ is near one; see Tables 1–6 of Section 3.2. This phenomenon is reminiscent of the nearly unstable autoregressive model discussed in Chan and Wei (1987). By virtue of the order of the observed Fisher's information number, they argued that neither the stationary normal limit nor the unit-root limit distributions would be a good approximation to the finite sample behavior of the LSE for the situation where ρ is close to 1. Putting it differently, a main difficulty in using (1.6) when ρ is close to 1 may be due to the critical behaviors of the limit distributions associated with the EVE and LSE. Such critical behaviors perpetuate in the performance of the corresponding predictors.

Consider a family of nearly unstable models

$$y_t = \rho_n y_{t-1} + \varepsilon_t, \tag{1.7}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

in which $\rho_n = 1 - b/n$, b is a positive constant, and ε_t is defined as in (1.2). The notion of nearly unstable models is one of the most important concepts in time series econometrics since the papers of Chan and Wei (1987) and Phillips (1987). It has found widespread applications in the analysis of time series data; for more background information, see the survey articles of Chan (2006) and Chan (2009). Using the moment bounds of the EVE and LSE for this class of models, asymptotic expressions for the MSPEs of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ under (1.7),

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A(\alpha, b), \ \lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = R_B(b), \tag{1.8}$$

are established, where $R_A(\alpha, b)$ is a positive constant depending on α, b, and σ^2, and $R_B(b)$ is a positive constant depending on b and σ^2. When data are generated from model (1.1) with ρ fixed but close to 1, $R_A(\alpha, b)$ and $R_B(b)$, with $b = n(1-\rho)$, can be used in place of $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ to approximate $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$. Since $R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ vary with n, they are referred to as the finite sample approximations. It is shown in Section 3.2 that $R_A(\alpha, n(1-\rho))$ and $R_B(n(1-\rho))$ substantially outperform $R_A^o(\alpha, \rho)$ and $R_B^o(\rho)$ in situation where $n(1-\rho)$ is small to moderate.

An intriguing feature of $R_B^o(\rho)$ is that it exhibits a jump behavior at the point $\rho = 1$. According to (47) and (48) of Ing and Yang (2014),

$$R_B^o(\rho) = \begin{cases} 2\sigma^2, \ \text{for } 0 \le \rho < 1, \\ \\ 4\sigma^2, \ \text{for } \rho = 1. \end{cases} \tag{1.9}$$

This phenomenon is analogous to the "quantum jump" behavior observed in physics, where the state of a system remains unchanged until a critical amount of energy is accumulated. With the MSPE jump in (1.9), it is interesting to explore if a connection between the stationary and the unstable regimes can be established via a smooth transition mechanism

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

such as (1.7). Thus, would the relationship

$$R_B(b) \to \begin{cases} 2\sigma^2, \text{ as } b \to \infty\,, \\ \\ 4\sigma^2, \text{ as } b \to 0\,, \end{cases} \tag{1.10}$$

remain valid? The lower half of (1.10) does remain valid for $b \to 0$, the upper half fails to hold and $R_B(b)$ converges to σ^2 as $b \to \infty$; see (2.15) and (2.17) of Section 2.

This discrepancy in the upper half of (1.10) suggests that $1-(b/n)$ may be converging to unity too rapidly and as a result, $R_B(b)$ does not attain the limiting value $R_B^o(\rho)$ of the stationary case. In Section 3.1, we derive the limiting value, $\Lambda_1(\beta, b)$, of $n(\text{MSPE}_B - \sigma^2)$ for the general near unit-root model, (1.7) with $\rho_n = 1 - b/n^\beta, 0 < \beta \le 1$ and $b > 0$. No single β directly connects $\Lambda_1(\beta, b)$ from $2\sigma^2$ to $4\sigma^2$, but our result reveals that a connection can be established through two critical values of β, $\beta = 1/2$ and $\beta = 1$, that connect $\Lambda_1(\beta, b)$ for the stationary and intermediate states, and for the intermediate and unit-root states, respectively: $\lim_{b\to\infty} \Lambda_1(1/2, b) = 2\sigma^2$, $\lim_{b\to 0} \Lambda_1(1/2, b) = \sigma^2$, $\lim_{b\to\infty} \Lambda_1(1, b) = \lim_{b\to\infty} R_B(b) = \sigma^2$, and $\lim_{b\to 0} \Lambda_1(1, b) = \lim_{b\to 0} R_B(b) = 4\sigma^2$. This provides an alternative finite sample approximation, $\Lambda_1(1/2, n^{1/2}(1-\rho))$, for $n(\text{MSPE}_B - \sigma^2)$ when $n(1-\rho)$ stays far away from 0.

Although the EV predictor also encounters the MSPE jump at $\rho = 1$ in the sense that

$$\lim_{\rho\to 1} R_A^o(\alpha, \rho) \neq R_A^o(\alpha, 1), \tag{1.11}$$

it can be eliminated by $R_A(\alpha, b)$ which satisfies, for any $0 < \alpha < \infty$,

$$\lim_{\rho\to 1} R_A^o(\alpha, \rho) = \lim_{b\to\infty} R_A(\alpha, b), \quad R_A^o(\alpha, 1) = \lim_{b\to 0} R_A(\alpha, b). \tag{1.12}$$

To deepen our understanding of the EV predictor in the near unit-root region, we obtain an asymptotic expression for MSPE_A in the general near unit-root model. This result leads to an alternative finite sample approximation of $n^{\min\{1,2/\alpha\}}(\text{MSPE}_A - \sigma^2)$, which notably improves upon $R_A(\alpha, n(1-\rho))$ when $n(1-\rho)$ is relatively large.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Thus we focus on the analysis of near unit-root processes from a prediction perspective, on the analysis of general near unit-root processes, and we illustrate the importance of finite sample approximations of the MSPE derived from near unit-root and general near unit-root processes.

The rest of the paper is organized as follows. In Section 2, asymptotic properties of the EVE and LSE for the near unit-root case with $\rho_n = 1 - (b/n)$ are developed, which include the limit distributions and the moment bounds, and on asymptotic expressions for MSPE_A and MSPE_B. In Section 3.1, we extend the results to the case of $\rho_n = 1 - (b/n^{\beta}), 0 < \beta < 1$. In Section 3.2, we offer finite sample approximations of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$ based on the asymptotic expressions obtained in Sections 2 and 3.1. The resultant performance is illustrated using AR(1) models with Beta$(\alpha, 1)$ errors, in which the AR coefficient lies between 0.86 and 0.99 and $1.5 \le \alpha \le 4$ (see also Section 3.2). We conclude in Section 4. With the help of Section 3.2, we further provide a simple rule for choosing a finite sample approximation from those derived in the near unit-root and the general near unit-root models. This result, together with the proofs of the theorems in Sections 2 and 3.1, is deferred to the supplementary document.

2 Near Unit-Root Models

In this section, we provide asymptotic expressions for the MSPEs of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ under (1.7), where

$$\hat{y}_{n+1} = \hat{\mu}_n + \hat{\rho}_n y_n, \tag{2.1}$$

with $\hat{\mu}_n = \frac{1}{n-1}\sum_{t=1}^{n-1}(y_{t+1} - \hat{\rho}_n y_t)$, and

$$\tilde{y}_{n+1} = \tilde{\mu}_n + \tilde{\rho}_n y_n. \tag{2.2}$$

Here with $\boldsymbol{x}_j = (1, y_j)^T$,

$$(\tilde{\mu}_n, \tilde{\rho}_n) = (\sum_{j=1}^{n-1} \boldsymbol{x}_j \boldsymbol{x}_j^T)^{-1} \sum_{j=1}^{n-1} \boldsymbol{x}_j y_{j+1}. \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Let $z_n = (y_n - y_1)/(n-1)$. Then the MSPE of $\hat{y}_{n+1}$ can be expressed as

$$\mathrm{MSPE}_A = \mathrm{E}(y_{n+1} - \hat{y}_{n+1})^2 = \sigma^2 + \mathrm{E}\big\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y}) + [(1-\rho_n)\bar{y} - \mu + z_n]\big\}^2, \quad (2.4)$$

where $\mu = \mathrm{E}(\varepsilon_1)$. In addition, the MSPE of $\tilde{y}_{n+1}$ satisfies

$$\mathrm{MSPE}_B = \mathrm{E}(y_{n+1} - \tilde{y}_{n+1})^2 = \sigma^2 + \mathrm{E}\big\{n^{-1}\sum_{i=1}^{n}\eta_i + (y_n - \bar{y})(\tilde{\rho}_n - \rho_n)\big\}^2, \quad (2.5)$$

where $\eta_i = \varepsilon_i - \mu$. To simplify the exposition, we assume that $y_0 = 0$ for the rest of this paper. We begin by deriving the limit distributions of $\hat{\rho}_n$ and $\tilde{\rho}_n$ and providing asymptotic expressions for their mean squared errors (MSEs); see Theorems 2.1 and 2.2.

Theorem 2.1. *Assume (1.7) with $0 < b < \infty$ holds and that $\mathrm{E}\varepsilon_1^{\kappa} < \infty$ for some $\kappa > 0$. Then,*

$$\mathrm{E}\{n^{1+1/\alpha}(\hat{\rho}_n - \rho_n)\}^q < \infty, \ \textit{for any } q > 0. \quad (2.6)$$

Further, if $\mathrm{E}(\varepsilon_1^{1+\tau}) < \infty$ for some $\tau > 0$, then

$$\lim_{n\to\infty} P\{(cM_{\alpha,b}/\alpha)^{1/\alpha} n^{1+1/\alpha}(\hat{\rho}_n - \rho_n) > t\} = \exp\{-t^{\alpha}\},\ t > 0, \quad (2.7)$$

$$\lim_{n\to\infty} \mathrm{E}[(cM_{\alpha,b}/\alpha)^{1/\alpha} n^{1+1/\alpha}(\hat{\rho}_n - \rho_n)]^2 = \Gamma((\alpha+2)/\alpha), \quad (2.8)$$

where

$$M_{\alpha,b} \;=\; \mu^{\alpha}\int_0^1 [(1-\exp(-bx))/b]^{\alpha}\,dx. \quad (2.9)$$

Theorem 2.2. *Assume (1.7) with $0 < b < \infty$ holds, and that $\mathrm{E}\varepsilon_1^{2+\tau} < \infty$ for some $\tau > 0$. Then,*

$$\frac{n^{3/2}}{\sigma_{1,b}}(\tilde{\rho}_n - \rho_n) \xrightarrow{d} N(0,1), \quad (2.10)$$

where $\sigma_{1,b}^2 = \sigma^2/[\mu^2(I_1(b) - I_2(b))]$, with $I_1(b) = 2^{-1}b^{-3}[2b + 4\exp(-b) - \exp(-2b) - 3]$ and $I_2(b) = b^{-4}[b - 1 + \exp(-b)]^2$. If $\mathrm{E}\varepsilon_1^s < \infty$ for some $s > 10$ and there exist positive constants $K, a,$ and δ such that for all $|x-y| \le \delta$ and all large m,

$$|F_m(x) - F_m(y)| \le K|x-y|^a, \quad (2.11)$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

where F_m *is the distribution function of* $m^{-1/2}\sum_{t=1}^{m}(\varepsilon_t - \mathrm{E}\varepsilon_1)$*. Then*

$$\lim_{n\to\infty} \mathrm{E}[n^3(\tilde{\rho}_n - \rho_n)^2] = \sigma_{1,b}^2. \tag{2.12}$$

Here $\lim_{b\to 0}\sigma_{1,b}^2 = 12\sigma^2/\mu^2$ is exactly the limiting variance of $n^{3/2}(\tilde{\rho}_n - 1)$ derived in the unit-root model; see Chan (1989) and Ing and Yang (2014). Let $L_3(b) = b^{-4}[1 - \exp(-b) - b\exp(-b)]^2$ and $M'_{\alpha,b} = M_{\alpha,b}/\mu^\alpha$.

Theorem 2.3. *Assume (1.7) with* $0 < b < \infty$ *holds, and that* $\mathrm{E}\varepsilon_1^\kappa < \infty$ *for some* $\kappa > 2$*. Then*

$$\mathrm{MSPE}_A - \sigma^2 = n^{-2/\alpha}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-2/\alpha}\}), \tag{2.13}$$

yielding

$$\lim_{n\to\infty} n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2) = R_A(\alpha, b), \tag{2.14}$$

where

$$R_A(\alpha,b) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) I(\alpha \geq 2) + \sigma^2 I(\alpha \leq 2),$$

the dependence of $R_A(\alpha, b)$ *on* c *is being suppressed.*

In view of the proof of Theorem 2.3, the cross-product term $2\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})[(1-\rho_n)\bar{y} - \mu + z_n]\}$, in the expectation on the right-hand side of (2.4), is asymptotically negligible compared to the corresponding squared terms $\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2$ and $\mathrm{E}\{(1-\rho_n)\bar{y} - \mu + z_n\}^2$. Therefore, the asymptotic behavior of $\mathrm{MSPE}_A - \sigma^2$ is mainly determined by the last two expectations. Since $(1-\rho_n)\bar{y} - \mu + z_n = (n-1)^{-1}\sum_{j=1}^{n-1}\eta_{j+1}$, we have $\mathrm{E}\{(1-\rho_n)\bar{y} - \mu + z_n\}^2 = \sigma^2/(n-1)$. In addition, $(y_n - \bar{y})^2/n^2$ converges in probability to $\mu^2 L_3(b)$; see (S1.16) in the supplementary document. From this and (2.8), it is expected that $\mathrm{E}\{(\hat{\rho}_n - \rho_n)(y_n - \bar{y})\}^2$ is of order $n^{-2/\alpha}$. The details are presented in the proof of Theorem 2.3.

Equipped with (2.5) and Theorem 2.2, we have an asymptotic expression for MSPE_B.

Theorem 2.4. *Assume (1.7) with $0 < b < \infty$ holds, that ε_1 satisfies (2.11), and* $\mathrm{E}\varepsilon_1^s < \infty$ *for some $s > 12$. Then*

$$\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = R_B(b) = \left\{1 + \frac{L_3(b)}{I_1(b) - I_2(b)}\right\}\sigma^2. \tag{2.15}$$

Remark 1. By (37) and (42) of Ing and Yang (2014), for $0 \le \rho < 1$, (1.11) follows from

$$\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha}{cM_\alpha(\rho)}\Big\}^{\frac{2}{\alpha}} \frac{\sigma^2}{1-\rho^2} \to 0 \text{ as } \rho \to 1.$$

As

$$\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{cM'_{\alpha,b}}\Big)^{\frac{2}{\alpha}} L_3(b) \to \begin{cases} \frac{1}{4}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big\{\frac{\alpha(\alpha+1)}{c}\Big\}^{\frac{2}{\alpha}}, \text{ as } b \to 0\,, \\ \\ 0, \text{ as } b \to \infty\,, \end{cases}$$

the discrepancy between $R_A^o(\alpha, 1)$ and $R_A^o(\alpha, \rho)$ in (1.11) can be connected by $R_A(\alpha, b)$ in the sense of (1.12).

Remark 2. While

$$\lim_{b\to 0} R_B(b) = R_B^o(1) = 4\sigma^2, \tag{2.16}$$

$\lim_{b\to\infty} R_B(b)$ is not equivalent to $R_B^o(\rho)$ when ρ increases to 1. More specifically,

$$\sigma^2 = \lim_{b\to\infty} R_B(b) \neq \lim_{\rho\to 1} R_B^o(\rho) = 2\sigma^2, \tag{2.17}$$

recalling that $R_B^o(\rho) = 2\sigma^2$ for all $0 \le \rho < 1$.

Remark 3. Theorems 2.3 and 2.4 imply that when $\alpha < 2$ ($\alpha > 2$), the EV (LS) predictor is better (worse) than the LS (EV) predictor in the sense that the convergence rate of $\mathrm{MSPE}_A - \sigma^2$ ($\mathrm{MSPE}_B - \sigma^2$) is faster than $\mathrm{MSPE}_B - \sigma^2$ ($\mathrm{MSPE}_A - \sigma^2$). When $\alpha = 2$, $\mathrm{MSPE}_A - \sigma^2$ and $\mathrm{MSPE}_B - \sigma^2$ share the same rate of convergence, and EV (LS) predictor is more efficient than the LS (EV) predictor if and only if $R_A(\alpha, b) < R_B(b)$ ($R_A(\alpha, b) >$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

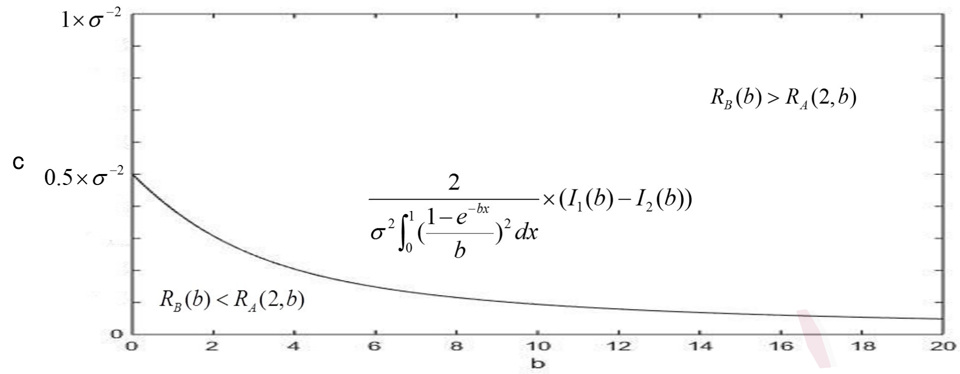

Figure 1: The graph of $c = h_\sigma(b)$.

$R_B(b)$). In fact, $R_A(\alpha, b) < R_B(b)$ or $R_A(\alpha, b) > R_B(b)$ depends on whether c is greater or smaller than the threshold function $h_\sigma(b) = 2(I_1(b) - I_2(b))[(\sigma^2/b^2)\int_0^1(1-\exp(-bx))^2dx]^{-1}$. Moreover, $\lim_{b\to 0} h_\sigma(b) = 1/(2\sigma^2)$ is exactly the threshold value in the case of $\rho = 1$ that partitions c into $c > 1/(2\sigma^2)$, in which the EV predictor dominates the LS predictor; and $0 < c < 1/(2\sigma^2)$, in which the latter predictor becomes more appealing, see Section 3 of Ing and Yang (2014). Additionally, $h_\sigma(b)$ is a decreasing function of b, suggesting that the advantage of the EV predictor over the LS one is more evident as the underlying process becomes "less non-stationary". It is also interesting to point out that in the case $0 \le \rho < 1$, $h_{\mu,\sigma}(\rho) = 2(1-\rho)\{(1+\rho)\mu^2 + (1-\rho)\sigma^2\}^{-1}$ is the threshold function playing the same role as $h_\sigma(b)$ in the near unit-root case (see Section 3 of Ing and Yang (2014)), and $h_{\mu,\sigma}(\rho)$ and $h_\sigma(b)$ coincide in the limit in the sense that $\lim_{\rho\to 1} h_{\mu,\sigma}(\rho) = \lim_{b\to\infty} h_\sigma(b) = 0$. This discussion is illustrated in Figure 1. Finally, in view of Theorems 2.1 and 2.2, the rankings of $\hat{\rho}_n$ and $\tilde{\rho}_n$ in terms of MSE are exactly the same as those of $\hat{y}_{n+1}$ and $\tilde{y}_{n+1}$ in terms of MSPE.

3 General Near Unit-Root Models

In this section, we extend the results in Section 2 to the general near unit-root case $\rho_n = 1 - b/n^\beta$, where $0 < \beta < 1$ and $0 < b < \infty$. This extension is to alleviate the difficulty of

$R_A(\alpha, n(1-\rho))$ $(R_B(n(1-\rho)))$ and $R_A^o(\alpha, \rho)$ $(R_B^o(\rho))$ to approximate $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ $(n(\mathrm{MSPE}_B - \sigma^2))$ when $1-\rho$ is small but $n(1-\rho)$ is relatively large, see Tables 1–6 for details.

3.1 Asymptotic Theories

We start by exploring the asymptotic distribution and the MSE of $\hat{\rho}_n$.

Theorem 3.1. *Assume* (1.7) *holds with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$, *and that* $\mathrm{E}\varepsilon_1^\kappa < \infty$ *for some* $\kappa > 0$. *Then*

$$\mathrm{E}\{n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)\}^q < \infty, \text{ for any } q > 0. \tag{3.1}$$

Further, if $\mathrm{E}(\varepsilon_1^{q_1}) < \infty$, *for some* $q_1 > 2/\beta$, *then*

$$\lim_{n\to\infty} P\{(c/\alpha)^{1/\alpha}(\mu/b)n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n) > t\} = \exp\{-t^\alpha\},\ t > 0\,, \tag{3.2}$$

$$\lim_{n\to\infty} \mathrm{E}[(c/\alpha)^{1/\alpha}(\mu/b)n^{\beta+1/\alpha}(\hat{\rho}_n - \rho_n)]^2 = \Gamma((\alpha+2)/\alpha). \tag{3.3}$$

Theorem 3.2. *Assume* (1.7) *holds with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$, *and that* $\mathrm{E}(\varepsilon_1^{2+\tau}) < \infty$ *for some* $\tau > 0$. *Then,*

$$\sqrt{k_n}(\tilde{\rho}_n - \rho_n) \xrightarrow{d} N(0, \sigma^2_{\beta,b}), \tag{3.4}$$

where $k_n = n^{1+\beta}$ *if* $0 < \beta \le 1/2$ *and* $n^{3\beta}$ *if* $\beta > 1/2$, *and*

$$\sigma^2_{\beta,b} = \begin{cases} \{(2b^3\sigma^2)^{-1}\mu^2 I(\beta = 1/2) + (2b)^{-1}\}^{-1}\,, & \text{if } 0 < \beta \le 1/2\,, \\ 2b^3\sigma^2/\mu^2\,, & \text{if } 1/2 < \beta < 1. \end{cases}$$

Moreover, if $\mathrm{E}\varepsilon_1^s < \infty$ *for some* $s > 10$ *and* (2.11) *holds, then*

$$\lim_{n\to\infty} \mathrm{E}[k_n(\tilde{\rho}_n - \rho_n)^2] = \sigma^2_{\beta,b}. \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

It is shown in the proof of Theorem 3.2 that $\tilde{\rho}_n - \rho_n = \sum_{i=2}^n (y_{i-1} - \bar{y})\eta_i / \sum_{i=2}^n (y_{i-1} - \bar{y})^2$ $= \sum_{i=2}^n (a_{n,i-1} + \xi_{n,i-1})\eta_i / \sum_{i=2}^n (a_{n,i-1} + \xi_{n,i-1})^2$, where $a_{n,i}$ and $\xi_{n,i}$, respectively, denote the deterministic and random components of $y_i - \bar{y}$. The order of magnitude of $(\tilde{\rho}_n - \rho_n)^2$ is determined by $(\sum_{i=2}^n a_{n,i-1}^2)^{-1}$ for $1/2 < \beta < 1$, $(\sum_{i=2}^n \xi_{n,i-1}^2)^{-1}$ for $0 < \beta < 1/2$, and $(\sum_{i=2}^n a_{n,i-1}^2 + \xi_{n,i-1}^2)^{-1}$ for $\beta = 1/2$, with the growth rates of $\sum_{i=2}^n a_{n,i-1}^2$ and $\sum_{i=2}^n \xi_{n,i-1}^2$ being $n^{3\beta}$ and $n^{1+\beta}$, respectively.

We have asymptotic expressions for MSPE_A and MSPE_B in the general near unit-root model.

Theorem 3.3. *Assume* (1.7) *holds with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$, *and that* $\mathrm{E}(\varepsilon_1^{q_1}) < \infty$, *for some* $q_1 > 2/\beta$. *Then*

(a) For $0 < \beta < 2/3$,

$$\text{MSPE}_A - \sigma^2 = n^{-\beta - \frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \frac{\sigma^2 b}{2\mu^2} + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\beta-\frac{2}{\alpha}}\}). \tag{3.6}$$

(b) For $2/3 < \beta < 1$,

$$\text{MSPE}_A - \sigma^2 = n^{2\beta - 2 - \frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \frac{1}{b^2} + \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{2\beta-2-\frac{2}{\alpha}}\}). \tag{3.7}$$

(b) For $\beta = 2/3$,

$$\begin{aligned} \text{MSPE}_A - \sigma^2 &= n^{-\frac{2}{3} - \frac{2}{\alpha}} \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \left\{\frac{\sigma^2 b}{2\mu^2} + \frac{1}{b^2}\right\} \\ &+ \frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\frac{2}{3}-\frac{2}{\alpha}}\}). \end{aligned} \tag{3.8}$$

Theorem 3.4. *Assume* (1.7) *holds with* $\rho_n = 1 - b/n^\beta$, *where* $0 < \beta < 1$ *and* $0 < b < \infty$,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

and that $\mathrm{E}\varepsilon_1^s < \infty$ *for some* $s > 12$. *Then*

$$\Lambda_1(\beta, b) \equiv \lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2) = \begin{cases} 2\sigma^2\,, & 0 < \beta < 1/2\,, \\ \sigma^2\{1 + \left(\frac{b^2\sigma^2}{\mu^2+b^2\sigma^2}\right)\}\,, & \beta = 1/2\,, \\ \sigma^2\,, & 1/2 < \beta < 1\,. \end{cases} \tag{3.9}$$

In view of (2.15), (2.16) and (2.17), Theorem 3.4 can be succinctly summarized to include the case $\beta = 1$ as

$$\Lambda_1(\beta, b) = \begin{cases} 2\sigma^2\,, & 0 < \beta < 1/2 \text{ and } 0 < b < \infty\,, \\ \sigma^2\{1 + \left(\frac{b^2\sigma^2}{\mu^2+b^2\sigma^2}\right)\}\,, & \beta = 1/2 \text{ and } 0 < b < \infty\,, \\ \sigma^2\,, & 1/2 < \beta < 1 \text{ and } 0 < b < \infty\,, \\ R_B(b)\,, & \beta = 1 \text{ and } 0 < b < \infty\,, \\ 4\sigma^2\,, & \beta = 1 \text{ and } b = 0. \end{cases} \tag{3.10}$$

Here $\Lambda_1(\beta, b) = 2\sigma^2, 0 < \beta < 1/2$ is designated as the "stationary state", $\Lambda_1(1, 0) = 4\sigma^2$ is designated as the "unit-root state", and $\Lambda_1(\beta, b)$ with $1/2 < \beta < 1$ is designated as the "intermediate state" because its value, σ^2, is different from the values of the unit-root and the stationary states.

At the critical point $\beta = 1$ that separates the unit-root and intermediate states, we have

(i) $\lim_{b\to 0} \Lambda_1(1, b) = 4\sigma^2$,

(ii) $\lim_{b\to\infty} \Lambda_1(1, b) = \sigma^2$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

At the critical point $\beta = 1/2$ that separates the stationary and intermediate states, we have

(i) $\lim_{b\to 0} \Lambda_1(\frac{1}{2}, b) = \sigma^2$,

(ii) $\lim_{b\to\infty} \Lambda_1(\frac{1}{2}, b) = 2\sigma^2$.

These critical phenomena are depicted in Figure 2.

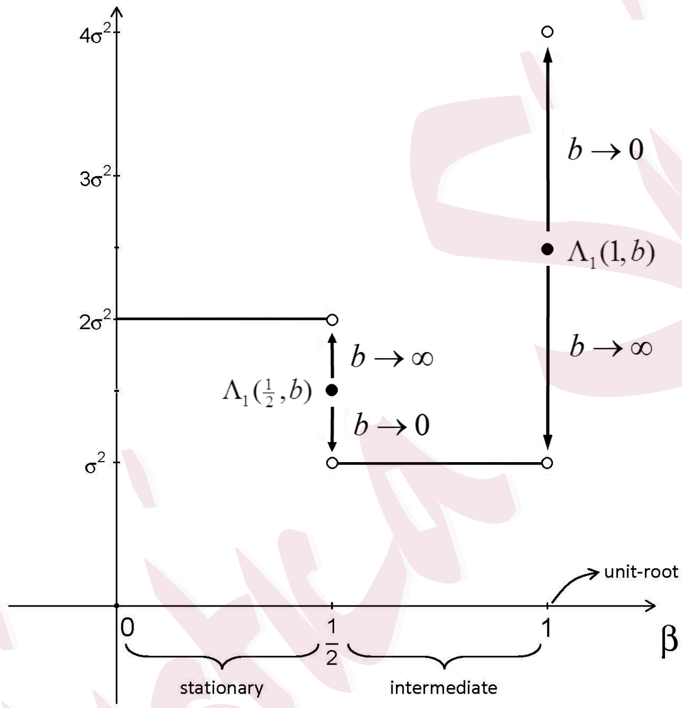

Figure 2: Theorem 3.4: the vertical axis denotes the value of $\Lambda_1(\beta, b)$. The points associated with $\Lambda_1(1/2, b)$ and $\Lambda_1(1, b)$ are not necessarily in the middle. They are only used to illustrate $\Lambda_1(1/2, b)$ ($\Lambda_1(1, b)$) decreases (increases) to σ^2 ($4\sigma^2$) as $b \to 0$, and increases (decreases) to $2\sigma^2$ (σ^2) as $b \to \infty$.

We infer the above discussion that, due to the existence of the intermediate state, there exists no $0 < \beta \leq 1$ such that $\Lambda_1(\beta, b)$ simultaneously satisfies $\lim_{b\to 0} \Lambda_1(\beta, b) = 4\sigma^2$ and $\lim_{b\to\infty} \Lambda_1(\beta, b) = 2\sigma^2$, implying that the discontinuity between the unit-root and stationary states of the MSPE_B cannot be connected by a general near unit-root model. In contrast, this discontinuity in MSPE_B can be connected through "two" general near unit-root models whose β values correspond to the critical points 1 and $1/2$, as illustrated in Figure 2.

3.2 Simulations

In this section, we propose finite sample approximations of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$ and report on their performance via data generated from model (1.1), with $\rho \in \{0.86, 0.9, 0.95, 0.975, 0.99\}$, $n \in \{100, 200, 500, 1000, 3000, 6000, 10000\}$, and ε_t a Beta$(\alpha, 1)$ with $\alpha \in \{1.5, 2, 4\}$. The performance of our finite sample corrections in the case of $0 < \alpha < 1$ is largely similar to that in the case of $\alpha = 1.5$. The details are skipped here. For a given triple (n, ρ, α), let $y_1^{(j)}, \ldots, y_n^{(j)}$ denote the data generated in the j-th simulation, $1 \leq j \leq 5000$. We also generated $y_{n+1}^{(j)}$ and computed the empirical estimate of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$,

$$R_{A_n} = n^{\min\{1,\frac{2}{\alpha}\}} \left\{ \frac{1}{5000} \sum_{j=1}^{5000} (y_{n+1}^{(j)} - \hat{y}_{n+1}^{(j)})^2 - \sigma^2 \right\}, \tag{3.11}$$

and that of $n(\mathrm{MSPE}_B - \sigma^2)$,

$$R_{B_n} = n \left\{ \frac{1}{5000} \sum_{j=1}^{5000} (y_{n+1}^{(j)} - \tilde{y}_{n+1}^{(j)})^2 - \sigma^2 \right\}, \tag{3.12}$$

where $\hat{y}_{n+1}^{(j)}$ and $\tilde{y}_{n+1}^{(j)}$, respectively, denote the EV and LS predictors calculated in the j-th simulation based on $y_1^{(j)}, \ldots, y_n^{(j)}$.

In addition to $R_A^o(\alpha, \rho)$ and $R_A(\alpha, n(1-\rho))$, we suggest an alternative approximation of $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ based on (3.6) and (3.8), derived from the general near unit-root model. Ignoring smaller order terms, we have that $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ is approximately equal to

$$\left\{ \Gamma\Big(\frac{\alpha+2}{\alpha}\Big) \Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2 (b/n^\beta)^2} \Big] \right\} + \frac{\sigma^2}{n^{1-2/\alpha}} \tag{3.13}$$

for $\alpha \geq 2$, and

$$\sigma^2 + n^{1-2/\alpha} \left\{ \Gamma\Big(\frac{\alpha+2}{\alpha}\Big) \Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}} \Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2 (b/n^\beta)^2} \Big] \right\} \tag{3.14}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

for $0 < \alpha < 2$. Replacing b/n^β in (3.13) and (3.14) by $1-\rho$, one gets:

$$R_A^*(\alpha, n, 1-\rho) = \begin{cases} R_1^*(\alpha,\rho) + R_2^*(\alpha, n(1-\rho)) + \frac{\sigma^2}{n^{1-2/\alpha}}\,, & \alpha \geq 2\,, \\ \\ \sigma^2 + n^{1-2/\alpha}R_1^*(\alpha,\rho) + n^{1-2/\alpha}R_2^*(\alpha, n(1-\rho))\,, & 0<\alpha<2, \end{cases} \tag{3.15}$$

where

$$R_1^*(\alpha,\rho) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha}\frac{\sigma^2(1-\rho)}{2\mu^2},$$

$$R_2^*(\alpha, n(1-\rho)) = \Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{2/\alpha}\frac{1}{n^2(1-\rho)^2}.$$

Here (3.15) depends on the parameters, β and b, in the general near unit-root model only through $1-\rho$, and hence can be implemented without facing the identifiability issue associated with β and b.

On the other hand, we consider an alternative

$$\Lambda_1(1/2, n^{1/2}(1-\rho)) = \sigma^2\left\{1 + \Big(\frac{n(1-\rho)^2\sigma^2}{\mu^2 + n(1-\rho)^2\sigma^2}\Big)\right\} \tag{3.16}$$

to $R_B^o(\alpha,\rho)$ and $R_B(n(1-\rho))$. It is expected that $\Lambda_1(1/2, n^{1/2}(1-\rho))$ will provide a more satisfactory approximation of $n(\mathrm{MSPE}_B - \sigma^2)$ when ρ is too small (large) for $R_B(n(1-\rho))$ ($R_B^o(\alpha,\rho)$) to do a good job.

For notational simplicity, let

$$\begin{aligned} &R_A^{(1)} = R_A^o(\alpha,\rho),\ R_A^{(2)} = R_A(\alpha, n(1-\rho)),\ R_A^{(3)} = R_A^*(\alpha, n, 1-\rho), \\ &R_B^{(1)} = R_B^o(\rho), \quad R_B^{(2)} = R_B(n(1-\rho)), \quad R_B^{(3)} = \Lambda_1(1/2, n^{1/2}(1-\rho)). \end{aligned}$$

The degree of closeness between $R_A^{(i)}$ and R_{A_n}, and that between $R_B^{(i)}$ and R_{B_n} is assessed by

$$P_A^{(i)} = \frac{\min\{R_A^{(i)}, R_{A_n}\}}{\max\{R_A^{(i)}, R_{A_n}\}} \text{ and } P_B^{(i)} = \frac{\min\{R_B^{(i)}, R_{B_n}\}}{\max\{R_B^{(i)}, R_{B_n}\}},\ i = 1,2,3.$$

Clearly, $0 \leq P_A^{(i)}, P_B^{(i)} \leq 1$ and a larger value represents a better performance. The values of R_{A_n} (R_{B_n}) and $P_A^{(i)}(P_B^{(i)}), i = 1,2,3$, are summarized in Tables 1–3 (Tables 4–6) for $\alpha = 1.5, 2$, and 4, respectively. In them, the value of $n(1-\rho)$, denoted by b, is included.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 1 has $P_A^{(3)}$ notably larger than $P_A^{(1)} = P_A^{(2)}$ for $n = 100$ or $1 \le n(1-\rho) \le 5$ although $P_A^{(i)}, i = 1, 2, 3$, have similar values otherwise. For any fixed ρ, all $P_A^{(i)}, i = 1, \ldots, 3$, gradually approach 1 as n increases, which is in line with the first relation of (1.6). Table 2 shows that when $\alpha = 2$ and $1 \le n(1-\rho) \le 5$, $P_A^{(2)}$ usually has the highest value among $P_A^{(i)}, i = 1, \ldots, 3$, except for the case of $(n, \rho) = (200, 0.99)$. For $5 < n(1-\rho) \le 28$, $P_A^{(3)}$ appears to dominate its competitors with the exception of $(n, \rho) = (1000, 0.99)$, in which $P_A^{(2)}$ is slightly larger than $P_A^{(3)}$. For $28 < n(1-\rho) \le 1400$, $P_A^{(1)}$ and $P_A^{(3)}$ behave quite similarly and are usually significantly larger than $P_A^{(2)}$, except for $\rho \ge 0.975$. For any fixed ρ, $P_A^{(1)}$ has an obvious tendency to increase to 1 as n grows from 100 to 10000, which is in agreement with the first relation of (1.6). Like Table 2, Table 3 ($\alpha = 4$) also shows that $P_A^{(2)}$ usually dominates $P_A^{(i)}, i = 1$ and 3, when $1 \le n(1-\rho) \le 5$. However, an exception happens in the case of $(n, \rho) = (100, 0.975)$, where $P_A^{(3)}$ is ranked first. The advantage of $R_A^{(3)}$ is more evident in Table 3 since $P_A^{(3)}$ is noticeably larger than $P_A^{(1)}$ and $P_A^{(2)}$ for almost all $10 \le n(1-\rho) \le 1400$. When $70 \le n(1-\rho) \le 1400$, $P_A^{(2)}$ in Table 3 is less than 0.05 and distinctively smaller than $P_A^{(2)}$ in Table 2. For any fixed ρ, Table 3 shows that $P_A^{(1)}$ still possesses a clear upward trend. Because the convergence rate of $\mathrm{MSPE}_A - \sigma^2$ is much slower in $\alpha = 4$ than in $\alpha \le 2$, $P_A^{(1)}$ may not be very close to 1 even when $n = 10000$.

It is shown in Table 4 that $P_B^{(2)} = \max_{1 \le i \le 3} P_B^{(i)}$ for $1 \le n(1-\rho) \le 12.5$, and $P_B^{(3)} = \max_{1 \le i \le 3} P_B^{(i)}$ for $12.5 < n(1-\rho) \le 300$, with the exception of $n(1-\rho) = 30$. For $300 < n(1-\rho) \le 1400$, $P_B^{(1)}$ and $P_B^{(3)}$ have similar values and are much larger than $P_B^{(2)}$. Tables 5 and 6 share similar features as Table 4. Moreover, as α grows, the areas of $n(1-\rho)$ for which $R_B^{(2)}$ works best and $R_B^{(3)}$ outperforms $R_B^{(1)}$ tend to expand. In Figure 3, we give the plots of time series realizations generated from model (1.1), with $\rho = 0.95$ and ε_t obeying Beta$(1.5, 1)$, Beta$(2, 1)$, and Beta$(4, 1)$ distributions. This figure shows that the nonstationary feature of these series becomes more evident as α increases, which may partly explain why $R_B^{(2)}$ and $R_B^{(3)}$ play increasingly essential roles in approximating R_{B_n} when α becomes larger.

As a final remark, Tables 1–3 (Tables 4–6) together portray situations where $R_A^{(i)}, i = 2, 3$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

$(R_B^{(i)}, i = 2, 3)$ can approximate R_{A_n} (R_{B_n}) better than $R_A^{(1)}$ $(R_B^{(1)})$. This information, in conjunction with suitable estimators of b, c, and α, enables one to construct a data-driven procedure for estimating $n^{\min\{1,2/\alpha\}}(\mathrm{MSPE}_A - \sigma^2)$ and $n(\mathrm{MSPE}_B - \sigma^2)$ in the near unit-root region. The details are deferred to the supplementary document.

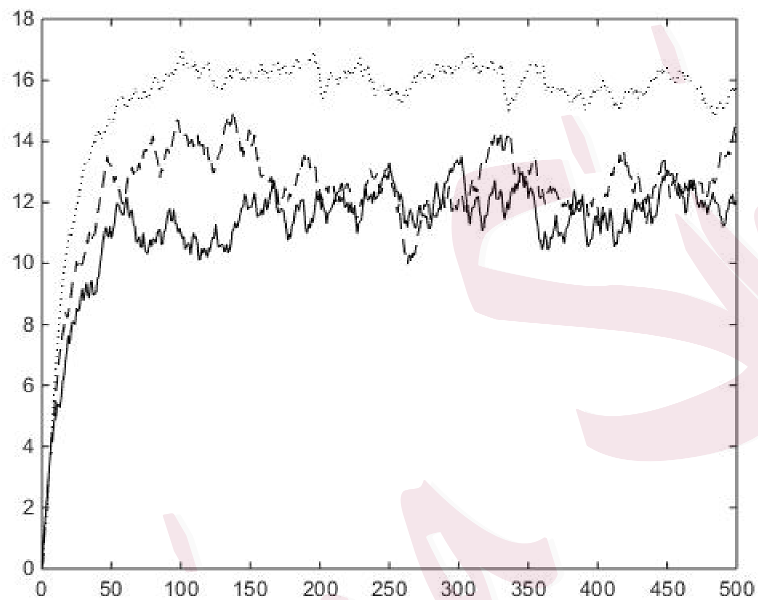

Figure 3: Plots of time series realizations generated from model (1.1), where $\rho = 0.95$ and ε_t has a Beta(α, 1) distribution, with $\alpha = 4$ (dotted line), 2 (dashed line) and 1.5 (solid line).

$\square$

4 Concluding Remarks

By deriving asymptotic expressions for the MSPEs of the EV and LS predictors under near unit-root and general near unit-root models, this paper provides a look at the performance of the LS and EV predictors in the near unit-root region. Our analysis reveals that the expressions derived for the LS predictor for the critical points, $\beta = 1$ and $\beta = 1/2$, not only jointly connect the discontinuities in $\lim_{n\to\infty} n(\mathrm{MSPE}_B - \sigma^2)$, but also combine their strengths to yield finite sample approximations of $n(\mathrm{MSPE}_B - \sigma^2)$ that perform satisfactorily

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 1: *The values of* R_{An} *and* $P_A^{(i)}, i = 1, 2, 3$, *under the* Beta(1.5, 1) *noise.* $P_A^{(1)}$ *(*$P_A^{(2)}$*,* $P_A^{(3)}$*) is marked in* **bold** *(***bold sans-serif***,* ***bold italics****) when it is equal to* $\max_{1 \le i \le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0770	0.0786	0.0917	0.1349	0.2150
	$P_A^{(1)}$	0.8909	0.8715	0.7470	0.5078	0.3186
	$P_A^{(2)}$	0.8909	0.8715	0.7470	0.5078	0.3186
	$P_A^{(3)}$	***0.9519***	***0.9361***	***0.8729***	***0.8170***	***0.6608***
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0707	0.0741	0.0743	0.0866	0.1299
	$P_A^{(1)}$	0.9703	0.9244	0.9219	0.7910	0.5273
	$P_A^{(2)}$	0.9703	0.9244	0.9219	0.7910	0.5273
	$P_A^{(3)}$	***0.9881***	***0.9584***	***0.9633***	***0.8914***	***0.9212***
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0718	0.0713	0.0682	0.0719	0.0792
	$P_A^{(1)}$	0.9555	0.9607	**0.9956**	0.9527	0.8649
	$P_A^{(2)}$	0.9555	0.9607	**0.9956**	0.9527	0.8649
	$P_A^{(3)}$	***0.9833***	***0.9826***	0.9809	***0.9720***	***0.9433***
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0690	0.0707	0.0699	0.0709	0.0729
	$P_A^{(1)}$	**0.9938**	0.9689	0.9799	0.9661	0.9396
	$P_A^{(2)}$	**0.9938**	0.9689	0.9799	0.9661	0.9396
	$P_A^{(3)}$	0.9833	***0.9861***	***0.9897***	***0.9738***	***0.9585***
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0700	0.0691	0.0670	0.0693	0.0676
	$P_A^{(1)}$	0.9785	0.9913	**0.9781**	0.9884	**0.9868**
	$P_A^{(2)}$	0.9785	0.9913	**0.9781**	0.9884	**0.9868**
	$P_A^{(3)}$	***0.9942***	***0.9971***	0.9724	***0.9913***	0.9839
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0688	0.0661	0.0702	0.0694	0.06853
	$P_A^{(1)}$	**0.9956**	**0.9649**	0.9757	0.9870	**0.9994**
	$P_A^{(2)}$	**0.9956**	**0.9649**	0.9757	0.9870	**0.9994**
	$P_A^{(3)}$	0.9913	0.9565	***0.9800***	***0.9903***	0.9989
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0681	0.0693	0.0702	0.0702	0.0671
	$P_A^{(1)}$	**0.9927**	0.9885	0.9757	0.9757	**0.9795**
	$P_A^{(2)}$	**0.9927**	0.9885	0.9757	0.9757	**0.9795**
	$P_A^{(3)}$	0.9825	***0.9970***	***0.9805***	***0.9786***	0.9777

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 2: *The values of* R_{An} *and* $P_A^{(i)}, i = 1, 2, 3$, *under the* Beta(2, 1) *noise.* $P_A^{(1)}$ ($P_A^{(2)}$, $P_A^{(3)}$) *is marked in* **bold** (**bold sans-serif**, ***bold italics***) *when it is equal to* $\max_{1\le i\le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0756	0.0812	0.1307	0.2644	0.5096
	$P_A^{(1)}$	0.8580	0.7635	0.4491	0.2160	0.1101
	$P_A^{(2)}$	0.8103	0.8288	**0.8256**	**0.8718**	**0.9241**
	$P_A^{(3)}$	***0.9180***	***0.8843***	0.7550	0.8211	0.4824
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0685	0.0690	0.0745	0.1198	0.3141
	$P_A^{(1)}$	0.9465	0.8986	0.7879	0.4766	0.1786
	$P_A^{(2)}$	0.8302	0.8435	0.9034	**0.9007**	0.9144
	$P_A^{(3)}$	***0.9573***	***0.9319***	***0.9218***	0.8106	***0.9747***
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0658	0.0623	0.0600	0.0673	0.1158
	$P_A^{(1)}$	**0.9864**	0.9952	0.9783	0.8484	0.4845
	$P_A^{(2)}$	0.8479	0.8973	0.9533	0.9331	**0.9318**
	$P_A^{(3)}$	0.9803	***0.9984***	***0.9953***	***0.9438***	0.8305
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0641	0.0619	0.0595	0.0588	0.0696
	$P_A^{(1)}$	0.9880	0.9984	0.9866	0.9711	0.8060
	$P_A^{(2)}$	0.8675	0.8982	0.9395	0.9728	**0.9669**
	$P_A^{(3)}$	***0.9960***	***0.9999***	***0.9929***	***0.9986***	0.9508
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0657	0.0602	0.0605	0.0593	0.0566
	$P_A^{(1)}$	**0.9861**	0.9710	**0.9708**	0.9621	0.9914
	$P_A^{(2)}$	0.8445	0.9216	0.9188	0.9386	**0.9989**
	$P_A^{(3)}$	0.9775	***0.9753***	0.9704	***0.9648***	0.9890
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0677	0.0610	0.0594	0.0576	0.0587
	$P_A^{(1)}$	**0.9575**	0.9837	**0.9880**	0.9905	0.9566
	$P_A^{(2)}$	0.8199	0.9095	0.9344	0.9639	0.9508
	$P_A^{(3)}$	0.9491	***0.9882***	0.9870	***0.9910***	***0.9613***
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0652	0.0633	0.0588	0.0562	0.0584
	$P_A^{(1)}$	**0.9938**	**0.9794**	**0.9982**	0.9842	0.9606
	$P_A^{(2)}$	0.8512	0.8767	0.9438	**0.9875**	0.9503
	$P_A^{(3)}$	0.9862	0.9764	0.9980	0.9836	***0.9637***

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 3: *The values of* R_{An} *and* $P_A^{(i)}, i = 1, 2, 3$, *under the* Beta(4, 1) *noise.* $P_A^{(1)}$ ($P_A^{(2)}$, $P_A^{(3)}$) *is marked in* **bold** (**bold sans-serif**, ***bold italics***) *when it is equal to* $\max_{1\le i\le 3} P_A^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{An}	0.0109	0.0158	0.0493	0.1444	0.3148
	$P_A^{(1)}$	0.2522	0.1203	0.0183	0.0035	0.0006
	$P_A^{(2)}$	0.4488	0.6266	**0.8621**	0.9252	**0.9371**
	$P_A^{(3)}$	***0.8966***	***0.8465***	0.7918	***0.9963***	0.3540
$n = 200$	b	28	20	10	5.0	2.0
	R_{An}	0.0062	0.0067	0.0134	0.0474	0.1808
	$P_A^{(1)}$	0.4440	0.2836	0.0672	0.0105	0.0011
	$P_A^{(2)}$	0.1894	0.3433	0.7388	**0.8966**	**0.9591**
	$P_A^{(3)}$	***0.9033***	***0.8876***	***0.8709***	0.7973	0.8084
$n = 500$	b	70	50	25	12.5	5.0
	R_{An}	0.0041	0.0035	0.0037	0.0081	0.0454
	$P_A^{(1)}$	0.6729	0.5429	0.2432	0.0617	0.0044
	$P_A^{(2)}$	0.0486	0.1143	0.4054	0.7654	**1.0000**
	$P_A^{(3)}$	***0.9654***	***0.9695***	***0.9550***	***0.9044***	0.8111
$n = 1000$	b	140	100	50	25	10
	R_{An}	0.0035	0.0029	0.0022	0.0029	0.0112
	$P_A^{(1)}$	0.7818	0.6551	0.4091	0.1724	0.0179
	$P_A^{(2)}$	0.0129	0.0345	0.1364	0.5172	**0.8839**
	$P_A^{(3)}$	***0.9923***	***0.9580***	***0.9640***	***0.9389***	0.8830
$n = 3000$	b	420	300	150	75	30
	R_{An}	0.0032	0.0024	0.0015	0.0011	0.0017
	$P_A^{(1)}$	0.8550	0.7910	0.5997	0.4194	0.1079
	$P_A^{(2)}$	0.0015	0.0040	0.0252	0.1436	0.5942
	$P_A^{(3)}$	***0.9553***	***0.9598***	***0.9209***	***0.9940***	***0.9641***
$n = 6000$	b	840	600	300	150	60
	R_{An}	0.0031	0.0023	0.0013	0.0008	0.0008
	$P_A^{(1)}$	0.8979	0.8313	0.7151	0.5316	0.2298
	$P_A^{(2)}$	0.0004	0.0010	0.0074	0.0451	0.3105
	$P_A^{(3)}$	***0.9555***	***0.9442***	***0.9677***	***0.9628***	***0.9605***
$n = 10000$	b	1400	1000	500	250	100
	R_{An}	0.0030	0.0022	0.0011	0.0007	0.0005
	$P_A^{(1)}$	0.9000	0.8636	0.8545	0.6571	0.3600
	$P_A^{(2)}$	0.0002	0.0040	0.0031	0.0202	0.1800
	$P_A^{(3)}$	***0.9506***	***0.9608***	***0.9217***	***0.9428***	***0.9260***

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 4: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1, 2, 3$, *under the* Beta$(1.5, 1)$ *noise.* $P_B^{(1)}$ $(P_B^{(2)}, P_B^{(3)})$ *is marked in* **bold** (**bold sans-serif**, ***bold italics***) *when it is equal to* $\max_{1 \le i \le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.1086	0.1028	0.1167	0.1584	0.2170
	$P_B^{(1)}$	0.7920	0.7498	0.8512	0.8655	0.6318
	$P_B^{(2)}$	0.7366	**0.8327**	**0.9443**	**0.9836**	**0.9894**
	$P_B^{(3)}$	***0.8030***	0.7737	0.6142	0.4379	0.3165
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.1084	0.1007	0.0919	0.1130	0.1688
	$P_B^{(1)}$	0.7901	0.7345	0.6703	0.8242	0.8122
	$P_B^{(2)}$	0.6815	0.7557	**0.9314**	**0.9752**	**0.9819**
	$P_B^{(3)}$	***0.9029***	***0.8687***	0.8110	0.6209	0.4077
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.1192	0.1096	0.0860	0.0868	0.1112
	$P_B^{(1)}$	0.8692	0.7994	0.6273	0.6331	0.8111
	$P_B^{(2)}$	0.5921	0.6515	0.8663	**0.9401**	**0.9910**
	$P_B^{(3)}$	***0.9498***	***0.9308***	***0.9506***	0.8343	0.6224
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.1291	0.1209	0.0976	0.0835	0.0899
	$P_B^{(1)}$	0.9411	0.8818	0.7119	0.6090	0.6562
	$P_B^{(2)}$	0.5390	0.5782	0.7316	0.8922	**0.9522**
	$P_B^{(3)}$	***0.9500***	***0.9390***	***0.9292***	***0.9085***	0.7770
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.1323	0.1257	0.1133	0.0900	0.0761
	$P_B^{(1)}$	0.9647	0.9166	0.8264	0.6564	0.5550
	$P_B^{(2)}$	0.5201	0.5488	0.6125	0.7822	**0.9645**
	$P_B^{(3)}$	***0.9939***	***0.9904***	***0.9612***	***0.9624***	0.9497
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.1329	0.1306	0.1214	0.0977	0.0766
	$P_B^{(1)}$	0.9691	0.9523	0.8854	0.7126	0.5587
	$P_B^{(2)}$	0.5172	0.5267	0.5683	0.7103	0.9255
	$P_B^{(3)}$	***0.9903***	***0.9932***	***0.9832***	***0.9942***	***0.9870***
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.1340	0.1392	0.1356	0.1096	0.0804
	$P_B^{(1)}$	0.9771	**0.9849**	**0.9890**	0.7873	0.5864
	$P_B^{(2)}$	0.5126	0.4935	0.5073	0.6304	0.8694
	$P_B^{(3)}$	***0.9899***	0.9606	0.9236	***0.9656***	***0.9893***

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 5: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1, 2, 3$, *under the* Beta(2, 1) *noise.* $P_B^{(1)}$ ($P_B^{(2)}$, $P_B^{(3)}$) *is marked in* **bold** (**bold sans-serif**, ***bold italics***) *when it is equal to* $\max_{1\le i\le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.0824	0.0793	0.0956	0.1285	0.1736
	$P_B^{(1)}$	0.7414	0.7138	0.8605	0.8646	0.6399
	$P_B^{(2)}$	0.7868	**0.8752**	**0.9341**	**0.9821**	**0.9983**
	$P_B^{(3)}$	***0.8068***	0.7784	0.5987	0.4356	0.3204
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.0828	0.0757	0.0728	0.0891	0.1370
	$P_B^{(1)}$	0.7454	0.6814	0.6553	0.8020	0.8109
	$P_B^{(2)}$	0.7223	0.8151	**0.9533**	**0.9978**	**0.9835**
	$P_B^{(3)}$	***0.8916***	***0.8806***	0.8080	0.6331	0.4065
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.0920	0.0824	0.0666	0.0694	0.0880
	$P_B^{(1)}$	0.8282	0.7417	0.5995	0.6247	0.7921
	$P_B^{(2)}$	0.6215	0.7015	0.9054	**0.9524**	**0.9854**
	$P_B^{(3)}$	***0.9363***	***0.9335***	***0.9468***	0.8306	0.6352
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.0942	0.0900	0.0719	0.0651	0.0716
	$P_B^{(1)}$	0.8478	0.8101	0.6472	0.5860	0.6445
	$P_B^{(2)}$	0.5983	0.6289	0.8039	**0.9263**	**0.9693**
	$P_B^{(3)}$	***0.9914***	***0.9602***	***0.9566***	0.9166	0.7854
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.1061	0.1027	0.0888	0.0699	0.0600
	$P_B^{(1)}$	0.9550	0.9249	0.7994	0.6291	0.5406
	$P_B^{(2)}$	0.5260	0.5441	0.6338	0.8165	**0.9908**
	$P_B^{(3)}$	***0.9843***	***0.9673***	***0.9280***	***0.9456***	0.9582
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.1164	0.1041	0.0924	0.0739	0.0620
	$P_B^{(1)}$	**0.9538**	0.9375	0.8318	0.6656	0.5588
	$P_B^{(2)}$	0.4780	0.5350	0.6050	0.7612	0.9256
	$P_B^{(3)}$	0.9234	***0.9961***	***0.9930***	***0.9908***	***0.9571***
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.1063	0.1059	0.0994	0.0805	0.0653
	$P_B^{(1)}$	0.9567	0.9531	0.8946	0.7245	0.5877
	$P_B^{(2)}$	0.5189	0.5250	0.5603	0.6956	0.8667
	$P_B^{(3)}$	***0.9758***	***0.9897***	***0.9823***	***0.9928***	***0.9453***

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

Table 6: *The values of* R_{Bn} *and* $P_B^{(i)}, i = 1, 2, 3$, *under the* Beta(4, 1) *noise.* $P_B^{(1)}$ $(P_B^{(2)}, P_B^{(3)})$ *is marked in* **bold** (**bold sans-serif**, ***bold italics***) *when it is equal to* $\max_{1 \le i \le 3} P_B^{(i)}$.

ρ		0.86	0.9	0.95	0.975	0.99
$n = 100$	b	14	10	5.0	2.5	1.0
	R_{Bn}	0.0348	0.0343	0.0444	0.0625	0.0854
	$P_B^{(1)}$	0.6526	0.6435	0.8330	0.8528	0.6241
	$P_B^{(2)}$	**0.8939**	**0.9708**	**0.9640**	**0.9696**	**0.9778**
	$P_B^{(3)}$	0.8241	0.8085	0.6067	0.4277	0.3123
$n = 200$	b	28	20	10	5.0	2.0
	R_{Bn}	0.0340	0.0335	0.0336	0.0431	0.0666
	$P_B^{(1)}$	0.6371	0.6285	0.6304	0.8086	0.8003
	$P_B^{(2)}$	0.8452	**0.8836**	**0.9911**	**0.9930**	**0.9970**
	$P_B^{(3)}$	***0.8944***	0.8572	0.8098	0.6219	0.4007
$n = 500$	b	70	50	25	12.5	5.0
	R_{Bn}	0.0370	0.0336	0.0309	0.0322	0.0425
	$P_B^{(1)}$	0.6929	0.6304	0.5797	0.6041	0.7974
	$P_B^{(2)}$	0.7429	0.8244	**0.9353**	**0.9844**	**1.0000**
	$P_B^{(3)}$	***0.9296***	***0.9304***	0.9057	0.8388	0.6287
$n = 1000$	b	140	100	50	25	10
	R_{Bn}	0.0407	0.0376	0.0303	0.0303	0.0344
	$P_B^{(1)}$	0.7638	0.7054	0.5685	0.5685	0.6454
	$P_B^{(2)}$	0.6641	0.7234	0.9142	**0.9538**	**0.9680**
	$P_B^{(3)}$	***0.9497***	***0.9178***	***0.9631***	0.9024	0.7784
$n = 3000$	b	420	300	150	75	30
	R_{Bn}	0.0462	0.0439	0.0341	0.0295	0.0286
	$P_B^{(1)}$	0.8674	0.8239	0.6394	0.5549	0.5374
	$P_B^{(2)}$	0.5791	0.6109	0.7924	0.9256	**0.9967**
	$P_B^{(3)}$	***0.9857***	***0.9439***	***0.9680***	***0.9662***	0.9417
$n = 6000$	b	840	600	300	150	60
	R_{Bn}	0.0483	0.0480	0.0383	0.0313	0.0283
	$P_B^{(1)}$	0.9074	0.9003	0.7197	0.5876	0.5318
	$P_B^{(2)}$	0.5523	0.5572	0.6993	0.8622	**0.9725**
	$P_B^{(3)}$	***0.9914***	***0.9520***	***0.9618***	***0.9657***	0.9630
$n = 10000$	b	1400	1000	500	250	100
	R_{Bn}	0.0513	0.0495	0.0410	0.0323	0.0281
	$P_B^{(1)}$	0.9624	0.9287	0.7692	0.6060	0.5272
	$P_B^{(2)}$	0.5205	0.5397	0.6529	0.8321	0.9679
	$P_B^{(3)}$	***0.9829***	***0.9731***	***0.9822***	***0.9961***	***0.9869***

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0069

in the near unit-root region. The expressions derived for the EV predictor for $\beta = 1$ and $0 < \beta < 1$ also lead to finite sample approximations of $n(\mathrm{MSPE}_A - \sigma^2)$ that can achieve a similar goal. The results established here and in Ing and Yang (2014) can be unified as follows:

$$\begin{aligned}
&\mathrm{MSPE}_A - \sigma^2 \\
&= n^{-2/\alpha}\Gamma\Big(\frac{\alpha+2}{\alpha}\Big)\Big(\frac{\alpha}{c}\Big)^{\frac{2}{\alpha}}\Big\{\Big[\frac{\sigma^2 b}{2\mu^2 n^\beta} + \frac{1}{n^2(b/n^\beta)^2}\Big]I_{\{0<\beta<1,0<b<\infty\}} \\
&+\Big(\frac{1}{M'_{\alpha,b}}\Big)^{2/\alpha}L_3(b)I_{\{\beta=1,0<b<\infty\}} + \Big(\frac{1}{M_\alpha(1-b)}\Big)^{2/\alpha}\frac{\sigma^2}{b(2-b)}I_{\{\beta=0,0<b\le 1\}} + \frac{(\alpha+1)^{2/\alpha}}{4}I_{\{\beta=1,b=0\}}\Big\} \\
&+\frac{\sigma^2}{n} + o(\max\{n^{-1}, n^{-\beta-\frac{2}{\alpha}}, n^{2\beta-2-\frac{2}{\alpha}}\}),
\end{aligned} \tag{4.1}$$

$$\begin{aligned}
&\mathrm{MSPE}_B - \sigma^2 \\
&= n^{-1}\left\{\Lambda_1(\beta,b)I_{\{0<\beta\le 1,0<b<\infty\}} + 2\sigma^2 I_{\{\beta=0,0<b\le 1\}} + 4\sigma^2 I_{\{\beta=1,b=0\}}\right\} + o(n^{-1}).
\end{aligned} \tag{4.2}$$

Equations (4.1) and (4.2) provide a more comprehensive perspective on the performance of the EV and LS predictors and may facilitate broader applications.

5 Supplementary Materials

Section S1 contains omitted proofs of the theorems in Sections 2 and 3.1, whereas Section S2 provides practical guidelines for choosing finite sample approximations from those derived in the near unit-root and the general near unit-root models.

Acknowledgments

We would like to thank the Editor, an associate editor and two anonymous reviewers for their insightful and constructive comments, which greatly improve the presentation of this paper. This research was supported in part by grants from HKSAR-RGC-GRF Nos 14300514 and 14325216; HKSAR-RGC-CRF:CityU8/CRG/12G; and the Theme-based Research Scheme of HKSAR-RGC-TBS T32-101/15-R (Chan); by the Science Vanguard Research Program of the Ministry of Science and Technology, Taiwan (Ing); by the National Natural Science Foundation of China (Grant No. 11371318); Zhejiang Provincial Natural Science Foundation of China (Grant No. R16A010001); and the Fundamental Research Funds for the Central Universities, Ministry of Education, China (Zhang).

References

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Ser. B* **63**, 167–241.

Bell, C. B. and Smith, E. P. (1986). Inference for non-negative autoregressive schemes. *Communications in Statistics: Theory and Methods* **15**, 2267–2293.

Chan, N. H. (2006). Inference for time series and stochastic processes. *Statistica Sinica* **16**, 683–696.

Chan, N. H. (2009). Time series with roots on or near the unit circle. In Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, T. (eds.), *Handbook of Financial Time Series*, 695–707. Springer-Verlag, New York.

Chan, N. H. and Ing, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Annals of Statistics* **39**, 1526-1550.

Chan, N. H. and Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* **15**, 1050–1063.

Datta, D. and McCormick, W. P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association* **90**, 1289–1300.

Davis, R. A. and McCormick, W. P. (1989). Estimation for first-order autoregressive processes with positive or bounded innovations. *Stochastic Processes and Their Applications* **31**, 237–250.

Findley, D. F. and Wei, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *Journal of Multivariate Analysis* **83**, 415V-450.

Finkenstädt, B. and Rootzén, H. (eds.) (2004). *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall, New York.

Gaver, D. P. and Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability* **12**, 727–745.

Hsiao, W.-C., Huang, H.-Y. and Ing, C.-K. (2017). Interval estimation for a first-order positive autoregressive process. *Journal of Time Series Analysis*, to appear.

Ing, C.-K. and Yang, C.-Y. (2014). Predictor selection for positive autoregressive processes. *Journal of the American Statistical Association* **109**, 243–253.

Lawrance, A. J. and Lewis, P. A. W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with Discussion). *Journal of the Royal Statistical Society, Ser. B* **47**, 165–202.

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society, Ser. B* **56**, 653–671.

Nielsen, B. and Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *Journal of Time Series Analysis* **24**, 337–344.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547.

Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* **136**, 115–130.

Sarlak, N. (2008). Annual streamflow modelling with asymmetric distribution function. *Hydrological Processes* **22**, 3403–3409.

Smith, R. L. (1994). Nonregular regression. *Biometrika* **81**, 173–243.

Wei, C.-Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with application to time series. *Annals of Statistics* **15**, 1667–1682.

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

Whittle, P. (1963). *Prediction and Regulation: by Linear Least-Square Methods.* Van-Nostrand, Princeton, New Jersey.

Yu, S.-H., Lin, C.-C. and Cheng, H.-W. (2012). A note on mean squared prediction error under the unit root model with deterministic trend. *Journal of Time Series Analysis* **33**, 276–286.